Antenatal care in the European Union: Equal chances for all new citizens of the Community

by

Annette Bernloehr

Submitted for the degree of Doctor of Philosophy

European Institute of Health and Medical Sciences
University of Surrey

2007

© Annette Bernloehr 2007
Abstract

Background

Antenatal care can contribute towards promoting the public health aims of the European Union (EU) and provides a catalyst that enables the process of growing together as a community of nations. Moreover, patient mobility and its implications for patient safety become increasingly important throughout the EU. Differences in the approaches to antenatal care need to be known in order to avoid duplications or gaps in care for women seeking care in a country other than their home country.

Aims

The overall aim of this study was to make cross-border antenatal care safer and to contribute to the evidence base to enable the best possible starting conditions for the Community’s future citizens. To meet this aim, the best available sources on antenatal care within all member states were critically analysed, examining evidence-based and expert recommendations. The theoretical background for making decisions on antenatal care was developed.

Moreover, a comprehensive review of the content of national guidelines on antenatal care was required to find out whether a common minimum guideline would be beneficial, and what this guideline might contain. A model was needed to be developed for integrating existing guidelines to a common minimum guideline to complement national health policies.

Study design

The study used a mixed methods approach, which consisted of a survey conducted across the EU and an extensive critical review of the state of the art regarding guidelines and antenatal care. In addition, a critical in-depth appraisal of the national guidelines from England and Wales and Germany was conducted, using the instrument of the AGREE collaboration for the appraisal of guidelines for research & evaluation as well as a critical comparison of the individual recommendations of the two national guidelines.
In the survey, the Ministries of Health and equivalent bodies, as well as the societies of obstetricians and midwives were asked to complete a structured questionnaire on the content of national guidelines for antenatal care. Descriptive analyses identified which and how many states recommend a test and to how many people this applied. The tests which were recommended by more than 50% of the states and applied to more than 50% of the inhabitants of the EU were compared to the measures supported by scientific evidence. Finally the correlation between the Gross National Product (GNP) of a state and the number of tests recommended was investigated.

Results

After sending 155 questionnaires and 61 reminders via surface and e-mail, as well as distributing another three questionnaires at a conference, answers were obtained from all 25 member states. 20 of them have a national guideline. A total of 47 tests were reported.

From the review of the literature and other sources, a model for decision-making based on double majorities, i.e. recommended by more than 50% of member states and applying to more than 50% of the Union’s inhabitants, was derived. The in-depth appraisal of the two national guidelines and the subsequent comparison of their recommendations gave evidence that the recommendations made by national guidelines can have high quality, irrespective of what an assessment with a tool to assess guideline quality suggests. Despite their similarly good recommendations, areas to improve the credibility of the guidelines from England and Wales and Germany were identified. Moreover, also weaknesses of the AGREE-instrument were identified, followed by suggestions for alleviating them.

Using the model of double majorities on the findings from the survey, it was found that 23 of the reported tests were recommended for routine care by more than 50% of the countries and apply to more than 50% of inhabitants. All but four of these tests were also supported by scientific evidence, which supported the validity of this model for integrating national guidelines
to a common minimum guideline on antenatal care. States with a GNP below EU-average recommend more tests than the others.

Conclusion

This study presents in detail what the national guidelines of the member states of the EU recommend for antenatal care and how they relate to current scientific evidence. As the findings from the survey were seen as important for enhancing safety in cross-border antenatal care, they were published during the course of the study (Bernloehr et al 2005; 2007).

In addition, the findings from the study demonstrate for the first time that extracting the measures from national guidelines that are recommended by the majority of states and apply to the majority of inhabitants of the EU leads to the development of a guideline compatible with scientific evidence. On the basis of all parts of the study, a common minimum guideline for antenatal care in the EU was established and recommended as it was found that such a guideline is useful and possible under the legislation of the EU.
Contents

Acknowledgements ............................................................................................................ ix
Glossary of terms ............................................................................................................. x

Chapter 1 - Introduction to the study ................................................................. 1

1.1 Background ................................................................................................................. 1
  1.1.1 Free movement of persons, services, capital and goods ......................... 3
  1.1.2 Implications of crossing borders ................................................................. 6
1.2 Aims and objectives ..................................................................................................... 9
1.3 Study design ................................................................................................................ 10
1.4 Summary of the chapters of the thesis ................................................................. 12

Chapter 2 - Review of the literature ................................................................. 13

2.1 Introduction ................................................................................................................... 13
2.2 Selection and use of the literature .............................................................................. 13
  2.2.1 Antenatal care in the member states of the EU ......................................... 13
  2.2.2 Implications, advantages and disadvantages of guidelines ....................... 23
2.3 Antenatal care in the member states of the European Union ............................ 24
  2.3.1 Basic documents ......................................................................................... 24
  2.3.2 Antenatal care in the EU ............................................................................ 31
  2.3.3 Timing and number of antenatal visits ....................................................... 37
  2.3.4 Tests recommended for antenatal care ....................................................... 39
  2.3.5 Guidelines on antenatal care in the member states of the EU ............... 42
  2.3.6 Randomised controlled trials comparing different schedules of care ...... 60
  2.3.7 The WHO new model of antenatal care .................................................... 69
  2.3.8 Conclusions to the review on antenatal care in the member states ........ 76
2.4 Implications, advantages and disadvantages of guidelines ............................... 77
  2.4.1 Ethical implications of guidelines ................................................................ 83
  2.4.2 Ethical challenges when setting up guidelines on antenatal care ............. 85

Chapter 3 - Material and methodology ............................................................... 91

3.1 Introduction ................................................................................................................... 91
3.2 Selection and use of material for the concept analysis .......................................... 92

Part 1 – The theoretical basis of guidelines
3.3 Guidelines .................................................................................................................... 96
  3.3.1 The quality of guidelines .......................................................................... 96
  3.3.2 Evaluation of recommendations ............................................................... 97
  3.3.3 Instruments for the systematic appraisal of guidelines ......................... 101
  3.3.4 Appraisal of guidelines for research and evaluation in Europe ............ 105
3.4 Theoretical framework: Evidence-based policy .................................................. 111
  3.4.1 Evidence-based health care and evidence-based policy ......................... 112
### Table of Contents

3.4.2 The nature and the hierarchy of evidence ................................................... 117  
3.4.3 Expert opinion ............................................................................................. 126  
3.4.4 Health policy ............................................................................................... 129  
3.4.5 Factors with the potential to modify decisions ............................................ 133  
3.4.6 How to weight the evidence to set up guidelines? ...................................... 136  
3.5 Implications of the analysis of concepts and of the literature .......................... 148  

**Part 2 – Health care and decision-making in the EU**  
3.6 Health care and decision-making at European level ........................................... 151  
3.7 European integration and public health in the EU ............................................. 156  
  3.7.1 Health care in the Treaties of Maastricht and Amsterdam ........................ 156  
  3.7.2 Framework programmes on community action for research and development.................................................. 159  
  3.7.3 Organisational conflicts, task overlaps and ambiguities of mandates .... 161  
  3.7.4 A new strategy to overcome the ambiguity of mandates ............................ 165  
  3.7.5 General trends, reproductive health and patient mobility ............................ 167  
3.8 Perceived legitimacy, health policy and harmonisation ..................................... 172  
3.9 Decision-making in the Council of Europe ...................................................... 174  
  3.9.1 Decision-making according to the Treaty of Nice ......................................... 175  
  3.9.2 Decision-making according to the Constitution for Europe .......................... 176  
3.10 Implications of the reviews for the study ............................................................ 178  

**Part 3 – The research plan**  
3.11 Ethical considerations ....................................................................................... 180  
3.12 The survey ....................................................................................................... 187  
  3.12.1 Designing the questionnaire ..................................................................... 189  
  3.12.2 Pilot study and properties of the questionnaire ....................................... 190  
  3.12.3 Intra-rater reliability .................................................................................. 192  
  3.12.4 Inter-rater reliability .................................................................................. 195  
  3.12.5 Distribution of the questionnaire ............................................................... 196  
  3.12.6 Analysis of the survey data ....................................................................... 198  
3.13 In-depth appraisal of two national guidelines .................................................. 202  

**Chapter 4 - Results** ............................................................................................. 208  
4.1 Results from the survey ..................................................................................... 208  
  4.1.1 Response rates ......................................................................................... 208  
  4.1.2 Multiple responses ................................................................................... 209  
  4.1.3 Sources of national guidelines ................................................................. 211  
  4.1.4 Tests recommended for antenatal care .................................................... 212  
  4.1.5 National recommendations and published evidence ................................ 218  
  4.1.6 The link between the gross national product and the number of tests ..... 219  
  4.1.7 Frequency of tests .................................................................................... 221  
4.2 Results from the in-depth appraisal of two national guidelines .......................... 221  
  4.2.1 Appraisal of the methodological quality of the two national guidelines .... 221  
  4.2.2 Comparison of the recommendations of the two national guidelines ....... 250
Chapter 5 - Discussion and conclusion to the study

5.1 Introduction
5.2 The link between the methodological quality of guidelines and the quality of their recommendations
5.3 Limitations of the study and suggestions for future research
   5.3.1 The categories 'evidence-based guidelines' and 'expert-opinion-based guidelines' are not appropriate at a national level
   5.3.2 The link between the methodological quality of guidelines and their recommendations
   5.3.3 Guidelines are only an instrument for improving antenatal care
5.4 Contributions made by this study
   5.4.1 Contributions to knowledge and understanding
   5.4.2 Implications for practice
   5.4.3 Contributions made to further developing methodology
   5.4.4 Contributions made to theory and theoretical understanding
5.5 Conclusion

References

List of tables

2.1 Coverage of antenatal care in the EU-15 states
2.2 Timing and schedule of antenatal visits in the EU-15 states
2.3a Antenatal screening tests recommended in the EU-15 states - Physical examinations
2.3b Antenatal screening tests recommended in the EU-15 state - technical tests
2.3c Antenatal screening tests recommended in the EU-15 states - laboratory tests
2.4 Antenatal screening tests recommended in Denmark - Physical examinations
2.5a Antenatal screening tests recommended in Sweden - Physical examinations
2.5b Antenatal screening tests recommended in Sweden - Technical tests
2.5c Antenatal screening tests recommended in Sweden - Laboratory tests
2.6a Antenatal screening tests recommended by the WHO basic model - Physical examinations
2.6b Antenatal screening tests recommended by the WHO basic model - Technical tests
2.6c Antenatal screening tests recommended by the WHO basic model - Laboratory tests
2.7 Levels of evidence according to NICE
2.8 Grading system for recommendations according to NICE
2.9 Levels of evidence according to the Oxford Centre for Evidence-based Medicine: Therapy/Prevention, Aetiology/Harm
2.10 Levels of evidence according to the Oxford Centre for Evidence-based Medicine: Diagnosis
2.11 Levels of evidence according to the Oxford Centre for Evidence-based Medicine: Economic and decision analyses
2.12 Levels of guidelines according to the method of consensus finding
4.1 Completed questionnaires
4.2 National recommendations for individual tests. Number of countries .............. 213
4.3 National recommendations for individual tests. Number of inhabitants .......... 214
4.4 Recommended number of individual tests per pregnancy (by more than 50% of countries applying to more than 50% of inhabitants) ........................................ 216
4.5 Recommended number of individual tests per pregnancy (by less than 50% of countries and/or applying to less than 50% of inhabitants) ...................... 217
4.6 Correlation between Gross National Product (GNP) per capita in Euro and intensity of care ................................................................. 219
4.7 Correlation between average Gross National Product (GNP) and intensity of care ....................................................................................... 220
4.8 Scope and purpose of the guideline from England/Wales ......................... 224
4.9 Scope and purpose of the German guideline ............................................ 225
4.10 Stakeholder involvement of the guideline from England/Wales ..................... 227
4.11 Stakeholder involvement of the German guideline ..................................... 228
4.12 Rigour of development of the guideline from England/Wales ....................... 231
4.13 Rigour of development of the German guideline ........................................ 232
4.14 Clarity and presentation of the guideline from England/Wales ...................... 235
4.15 Clarity and presentation of the German guideline ........................................ 236
4.16 Applicability of the guideline from England/Wales .................................... 238
4.17 Applicability of the German guideline ..................................................... 239
4.18 Editorial independence of the guideline from England/Wales ....................... 241
4.19 Editorial independence of the German guideline .......................................... 241
4.20 In-depth appraisal of the guideline from England/Wales by all four raters ...... 243
4.21 In-depth appraisal of the German guideline by all four raters ....................... 244
4.22 Individual appraisers' and standardised domain scores of the German guideline 248
4.23 Similarities and differences in the recommendations for routine antenatal care of the guidelines from England/Wales (E/W) and Germany (GER) ............ 252
5.1 Physical tests recommended for the EU-guideline ....................................... 275
5.2 Technical tests recommended for the EU-guideline ...................................... 276
5.3 Laboratory tests recommended for the EU-guideline .................................... 276

List of figures

2.1: Important contributing factors to policy making ........................................ 134

Appendices ........................................................................................................ 304

1 Databases used for the review of the literature and of the state of the art
2 Survey questionnaire, covering letter and consent form
3 Publications
Acknowledgements

My special thanks go to the supervisors who accompanied the development of the thesis throughout its different stages: Prof. Dr. Pam Smith, Prof. Dr. Huda Huijer-Abu Saad, Dr. Vasso Vydelingum and Prof. Dr. Bernhard Güntert.

The contribution of the respondents to the survey and the work of the appraisers of the guidelines are gratefully acknowledged, as are the various intellectual and personal challenges provided by others.

I would also like to thank the Robert Bosch Stiftung, Stuttgart, the B. Braun-Stiftung, Melsungen and the St. Vincenz-Hospital, Paderborn for financially contributing to the realisation of the thesis.
Glossary of terms

Alpha-fetoproteine: a protein produced in the foetus, which can be used as a marker for Down's Syndrome

Cardio-tocography: a record of the foetal heartbeat and the uterine contractions

Chlamydia trachomatis: coccoid rickettsia causing venereal infections

Cytomegalovirus: a Herpes virus

Foetal fibronectin: a glycoprotein, which can indicate an increased risk for preterm birth

Fundal height: the distance between the highest part of the uterus and the symphysis pubis, the umbilicus, or the rib

Gestational diabetes: maternal carbohydrate intolerance diagnosed during pregnancy, which results in hyperglycaemia

Gestational week: week of pregnancy after the first day of the last menstrual period

Glucose tolerance test: blood tests of glucose levels after controlled oral intake of glucose

Gonorrhoea: a venereal disease, caused by the bacterium Neisseria gonorrhoeae

Group B streptococci: Streptococcus agalactiae

Haemoglobinopathy: a blood disease characterized by abnormal haemoglobins

Hepatitis B: a viral hepatitis, transmitted by ingestion of contaminated bodily fluids

HIV: Human Immunodeficiency Virus, the virus causing the acquired immune deficiency syndrome (AIDS)

Listeria monocytogenes: a Gram positive bacterium, causing Listeriosis

Lues: a venereal infection caused by Treponema pallidum spirochete, which can be transmitted through the placenta

Multigravida: a woman who is pregnant for at least the second time

Papanicolaou smear: cervical cytology to test for malignant cells

Pelvimetry: measurement of the bony pelvis

Pre-eclampsia: a multisystem disorder, with the leading symptoms of new hypertension and proteinuria manifesting after 20 weeks of gestation

Primigravida: a woman who is pregnant for the first time

Puerperium: the period from childbirth to about six weeks post partum

Rubella: a viral disease, potentially damaging to the foetus

Toxoplasmosis: an infection with the parasite Toxoplasma gondii

Triple test: the assessment of three serological markers to estimate the risk for Down's Syndrome
Chapter 1 - Introduction to the study

1.1 Background

First and foremost, practical experience as the midwife in charge of the labour ward in a major hospital in Western Germany triggered the study. Each year about 1700 deliveries take place there. About 140 children are born to British women, most of them accompanying their husbands who are based in Germany with the British Army. Some of the women start antenatal care in the UK, some of them in Germany. Those starting in Germany are cared for according to a scheme, which was agreed between the hospital and the medical centre of the British Forces Germany. Although care is now well established and shared between the hospital, the British midwives and the medical centre, it took several years before this system was operating to the satisfaction of all parties. In the early years of co-operation, several tests were performed on admission to the labour ward instead of being performed at specific times during pregnancy. This was because these measures were part of the German guidelines on antenatal care (Bundesausschuss der Ärzte und Krankenkassen 2003), but were not within the protocol used by British health professionals, or were scheduled at different times during pregnancy (National Collaborating Centre for Women’s and Children’s Health 2003).

Although many of these differences were resolved after several years of co-operation, other specific traits remained. As an example, at the British medical centre, blood for hepatitis B screening is taken at the first antenatal appointment, as it is recommended in the guideline for England and Wales (National Collaborating Centre for Women’s and Children’s Health 2003: 35). On admission to the German hospital, this test is repeated, as it is specified in the German guidelines that the screening should take place after the 32nd gestational week (Bundesausschuss der Ärzte und Krankenkassen 2003). Due to a lack of communication between the two parties, such issues initially led to dissatisfaction amongst the British and the German health professionals, as each of the parties felt mistrusted by their colleagues. Moreover, both parties perceived the different courses of action as non-adherence to the
guideline, rather than realising that 'the guideline' was in fact two similar guidelines, but with sufficient differences as to cause misunderstandings. These misunderstandings amongst the health professionals also led to mistrust and dissatisfaction in the pregnant women, although clinical outcomes did not differ with regard to maternal and infant morbidity or mortality.

Although the degree of feeling 'not well cared for' and the feeling of insecurity were never assessed formally, the problem was evident from complaints to the management of the German hospital, as well as to the medical centre of the British Forces. In the first years of cooperation, frequent meetings of the professionals had to be held. Within these, the dissatisfaction of carers from both sides as well as the perceived dissatisfaction of the women was the most important issues on the agenda. Another sign for feeling insecure was that the British women asked the midwives of the British Forces to accompany them to the labour ward, although the British midwives were not allowed to practice in the setting. They were brought into the situation as advocates. Only after assessing the differences of the two guidelines in depth and after setting up instructions of how to manage the differences, the British women stopped bringing their midwives to the German labour ward. Also the complaints arising from discrepancies between the two systems diminished. As there is a constant turnover of British soldiers and their wives, as well as a regular turnover of British staff, this effect can not be ascribed to a mechanism of simply getting used to the German system. However, the situation was and is still worse when the women started antenatal care within the UK and changed countries during pregnancy. Due to the differences in the national guidelines, which are not known to the average practitioner, such patients still cause distress amongst the health professionals and leave the pregnant women feeling insecure. The fact that cross-border health care has implications, which need to be assessed, is also evident from activities of the European Commission. This is demonstrated in the following section.
1.1.1 Free movement of persons, services, capital and goods

One of the major challenges of the European integration process is the right of free movement of persons, services, capital and goods (Paton et al. 2002: 4). On April 28th 2001, the European Court of Justice has ruled that patients are allowed to seek ambulatory medical treatment in other EU countries. Since then, national institutions as well as health insurance companies are obliged to pay for this extra-territorial ambulatory treatment (Busse 2002: 240, Wismar et al. 2002: 23). In May 2003, the European Court of Justice again simplified the mechanisms for this off-territorial treatment. From then on, patients do not even need to inform their health insurances before seeking routine treatment in a country other than their home country (Deutsches Ärzteblatt 2003). Latest case-law of the European Court indicates that this right has been extended to hospital care (Commission of the European Communities 2004, The European Commission’s Health and Consumer Protection Directorate General 2006).

However, the resulting situation was well described by the former Commissioner for Health and Consumer Protection, David Byrne. He stated that according to EU law, patients have the right to go for treatment in other member states, but that this is not easy in practice. His speech at a conference on free movement and cross-border co-operation in Europe, held on 20 June 2003 in Luxembourg, highlights the need for transparency (Byrne 2003). In this, Byrne claimed that in order to being able to access health care in another member state information is required about the quality, availability and appropriateness of the treatments available there.

At present, there are not yet numbers available, as to how many people seek medical treatment in a country other than the one from which they originate (Commission of the European Communities 2004). In addition, it is difficult to estimate, how many women change country completely during pregnancy, or immediately before becoming pregnant. Moreover, an increasing number of people are just seeking medical treatment in another member state, but not changing countries totally. The European Commission acknowledged this problem by starting a series of meetings and publications for a “high level reflection process on patient mobility”, which was launched by two Commissioners in December 2003 (Commission of the
European Communities 2004, European Commission (n.d.) a, b + f, South East Partners 2003). At present, the minutes of an important meeting are available, which has taken place on November 21 2003 in Luxembourg (Fahy 2003). More detailed data were announced for 2004, but were not included in the follow-up document of this high level reflection process (Commission of the European Communities 2004). Instead, the European Commission recently adopted a policy to not only tolerate cross-border health care, but to actively support it (Commission of the European Communities 2005, Kyprianou 2005a). Cross-border care and an increased co-operation at European level are now regarded as a means to improve the quality and accessibility of health care by sharing information and spare capacity.

Although classified as urgent, no clarification as to how to apply the rulings of the Court of Justice to health services is available to date (European Communities 2006). For this, the European Commission launched in September 2006 a public consultation regarding EU action on health services, with which comments on how to ensure legal certainty regarding cross-border health care under Community law is sought, and on how to support co-operation between the health systems of the member states (The European Commission 2006).

However, the need for transparency and information is not only evident for patients, but also for health professionals wanting to practise in another member state. The practical experience at the German hospital only reinforced the insights gained as a postgraduate student in the UK. In this course, midwives from across the EU studied together. Especially the non-UK midwives reported difficulties in familiarising with a system, in which they were not trained. This is of special interest, as the European Court of Justice argued that cross-border care does not pose a threat to human health, as a similar standard of health care can be expected in all member states of the European Union (Wismar et al. 2002: 25). This opinion was based on the assumption that the mutual recognition of diplomas and established minimum training requirements for health care personnel go some way to guarantee this similar standard. However, the practical experience from the German hospital as well as the discussions amongst the midwives from across the EU gave rise to major concerns with regard to this statement. On the one hand, there are sufficient differences in the national guidelines to
produce gaps or an unnecessary, expensive and possibly harmful multiplication of tests for pregnant women. On the other hand, it was found that health care workers practising in a country other than that in which they were trained need clear guidelines to integrate themselves fully and safely. A major problem in this respect was a perception that it was nearly impossible to gain an overview of what was the current state of the art in the other EU states. This was reinforced by a policy update on public health, in which the Health Ministers of the member states, representatives of Europe’s health care sector, the European Parliament and the European Commission recommend activities to meet the information requirements of patients, professionals and policy-makers (South East Partners 2003).

From the above situation in the German hospital, which triggered the study, it was decided to use antenatal care as an example to explore the issues arising from cross-border health care in greater depth. This idea was reinforced by the statement of the Council of Europe, that antenatal care is a classic example for health problems, which are treated similarly across the EU regarding its organisation and content (Europarat 2001: 25). However, whilst on the one hand it needs to be demonstrated in detail, whether the organisation of care is similar, on the other, it needs to be specified, what the similar content of antenatal care is. It was therefore expected that antenatal care offers an excellent field for analysis on the European level. Moreover, it was also found that antenatal care is one of the most commonly used preventive measures and has therefore public health implications on a broad basis (Europarat 2001: 24ff). As a consequence, antenatal care has also the potential to be a rewarding field to study. After some initial reading on the subject, it was found that with two or three exceptions at least the member states of before May 2004 have some kind of guideline for antenatal care (Blondel et al. 1985). However, a study including 13 European countries also found that there are 13 different organisational systems for providing antenatal care (Hemminki & Blondel 2001). Both findings demonstrate that it was justified to expect to find sufficient material for study.

However, already the broad reading on antenatal care in the context of the EU stimulated the thought that it is worth exploring the potential for a common minimum guideline. This essentially means to address the basic conflict between independence and unification in the
member states of the EU. However, providing information about the existing guidelines for antenatal care in normal pregnancies is only a first step to ensuring safety in cross-border health care. A common guideline on antenatal care in the EU might help to highlight areas of good practice, as well as to identify areas for potential improvement. As EU health policy provides the basis and framework for a potential common minimum guideline, the cornerstones of the current health policy in the EU needed clarification. This is explored in a separate section of the thesis. At this stage, the implications of crossing borders are addressed as they provide a clearer picture of the background situation in the European Union.

1.1.2 Implications of crossing borders

From the obligation to pay for medical services outside the home country of the patient, several problems arise. Even within themselves, the member states’ health care systems are subject to conflicting pressures. From practitioners, managers, politicians as well as from the public, there is a rising demand for optimum, effective, efficient, evidence-based and affordable care, with different weighting of these factors, according to the perspective taken (Byrne 2004, Commission of the European Communities 2004, Harrison et al. 2002c). On the one hand, the proportion of the gross domestic product devoted to health care spending has doubled over the last three decades and is rising steadily. This development was found to be mainly due to demographic factors, the cost of new medical technologies and citizens’ increased expectations from health care (Commission of the European Communities 1998). On the other hand, although the member states of the EU all have a high standard of living and have invested a lot into their more or less well developed health systems, they are still confronted with high expectations from the public, and finally, they have limited resources to meet these needs and expectations (European Health Management Association 2000).

Even within the member states themselves, the health systems face an explosion of costs. By setting up practice guidelines that consider the quality of care as well as the cost-effectiveness of measures, the national key stakeholders try to limit expenses. The stakeholders try to
achieve a rationalisation of care in order to avoid the rationing of care to special patient groups (Güntert 1998 + 1999). However, the rational allocation of resources becomes a nearly insurmountable task when the former restrictions to patient mobility are abolished, as the former limits to mobility helped the effective planning as well as the control of costs (The European Commission (n.d.) c). As this barrier falls, a new strategy needs to be developed.

However, before the financial implications of cross-border health care are assessed, another fundamental prior condition must be fulfilled. A similar standard of health care must be achieved within the member states in order to avoid gaps or an unnecessary, expensive and possibly harmful duplication of care. With regard to this quality aspect, the European Court of Justice argued that cross-border care does not pose a threat to human health, as a similar standard of health care can be expected in all member states of the EU (Wismar et al. 2002: 25). However, as discussed before, this was based on the assumption that the mutual recognition of diplomas and established minimum training requirements for health care personnel go some way to guarantee this similar standard. This means that health care workers are allowed to practice in all member states, given that certain training criteria are fulfilled. However, health professionals practising in a country other than that in which they were trained need guidelines to integrate themselves safely into the respective national practice. Another problem of this general assumption is that it was made before ten new member states joined the Union in May 2004. It is possible that the former members have achieved similar standards over the years, but that this assimilation process starts just now in the new member states. However, it might also be that the picture is less homogenous even for the member states which joined the EU earlier.

In order to support or refute this rather general assumption of the European Court of Justice, it seems to be important to explore the actual basis and content of care in the member states. It might not be the case that a similar standard of health care can be expected on the basis of mutually recognised professional training. Philippe Busquin, the European Commissioner for Research stated that there is a need for a coherent, policy-oriented research effort at the EU level in the field of public health (Busquin 2002: v). The European Commission recommends
bringing together information about how certain issues are addressed in the member states. Moreover, common elements across the EU should be identified (The European Commission 2004). It demands a European strategy to ensure that citizens can seek care in other member states, and a European co-operation that helps to meet the challenges, the health systems face. The improvement of information and knowledge about health systems in order to enable the identification of best practice is highlighted.

Bringing the issues of costs and quality of care together, another problem of allowing or even stimulating cross-border health care becomes evident. Cross-border care and an increased co-operation at European level are now regarded as a means for improving on quality, as well as accessibility of health care by sharing information and spare capacity. However, as a by-product, this consumer-driven introduction of market forces changes the health systems. Although the member states still have the right to determine what health benefits their citizens are entitled to, and how this should be financed (Kyprianou 2005b), consumers now have the right and the chance to select the care they consider to be best. By consumers selecting what care they want to receive, member states might be forced to introduce measures which they do not think are beneficial, but which are offered by a neighbouring country.

When this mechanism works for a while, antenatal care will become very similar throughout the European Union, but will consist of measures which should not be offered to the whole population on a routine basis. In addition to that, care will not only become problematic with regard to its quality, but also more expensive. As these are serious issues, a counter-acting mechanism needs to be developed.
1.2 Aims and objectives

In this study, the best available sources with regard to antenatal care within all member states of the European Union will be critically analysed.

Aims of the study

The overall aim of the study is to make cross-border antenatal care safer by establishing the current state of the art within the EU, and to contribute to the evidence base for the best possible starting conditions for the Union’s future citizens.

Subsidiary aims

The subsidiary aim is to develop the theoretical understanding on antenatal care further, which may improve the acceptance of widely recommended methods throughout the EU, and to contribute to making policy-makers as well as practitioners aware of outdated and possibly harmful interventions. The distribution of limited resources will be optimised by highlighting interventions, which are proven to be effective.

The objectives are to

establish what guidelines on antenatal care exist in the different member states

examine the implications of different sorts of evidence on guideline development and clinical practice

develop the theoretical background for making health policy on antenatal care in the European Union

develop a model, which can be useful for making policy decisions in other disciplines in the health sector

recommend a common minimum guideline on antenatal care for the member states of the European Union, if this is appropriate.
1.3 Study design

Phase 1

The objective of critically appraising the relation between evidence-based guidelines and guidelines based on expert opinion for the defined public health problem 'antenatal care' was tackled by analysing the paradigms underlying and the methods for setting up evidence-based guidelines and guidelines based on expert opinion. The critical discussion of the nature of evidence was part of this, highlighting differences and similarities between scientifically generated evidence and evidence referred to by experts.

A critical appraisal of the implications that different methods for setting up guidelines have on guideline quality was achieved by a critical in-depth assessment of the methodological quality of the guidelines from England and Wales (National Collaborating Centre for Women's and Children's Health 2003) and Germany (Bundesausschuss der Ärzte und Krankenkassen 2003), using the instrument developed by a group of experts funded by the European Commission (The AGREE Collaboration 2001).

Phase 2

In this phase of the study it was planned to find out, whether the basic principle for decision making on the basis of 'double majorities' (majority of states and majority of citizens) is suitable for making decisions in the health sector, when regulations based on different paradigms already exist. In order to test the hypothesis that a guideline based on 'double majorities' of all national guidelines on antenatal care in the EU contains the same recommendations as a guideline based on the principles of evidence-based medicine, the following steps were taken:

Developing a questionnaire and conducting a survey on the content of national guidelines on antenatal care in all member states of the EU.

Extracting all tests which are recommended by at least 50% of the member states and which apply to at least 50% of the inhabitants of the EU (double majorities).
All measures recommended by double majorities were compared to the measures recommended by the most comprehensive evidence-based guideline on antenatal care to date (National Collaborating Centre for Women's and Children's Health 2003).

Phase 3

In order to develop a framework according to which the existing national guidelines on antenatal care could be used to efficiently set up a common minimum guideline, without compromising its quality, the horizon had to be broadened by exploring whether a common minimum guideline based on the findings of Phase 2 would mean insurmountable hardships for the less wealthy member states of the EU. For this, the number of measures currently recommended by the member states with a gross national product below average was compared to the number of measures recommended by the potential minimum standard based on the findings of Phase 2 of the study.

Phase 4

After critically reviewing and drawing together the findings of Phases 1 to 3 of the study, a common minimum guideline on antenatal care for the EU was provided. To achieve this, the potential strengths and weaknesses of such a guideline were critically analysed, and the guideline was located in the current health policy of the EU. On the basis of this, a new method for integrating national guidelines into a true European approach was suggested.
1.4 Summary of the chapters of the thesis

To delineate the research problem and to demonstrate how the problem is tackled, Chapter 1 of the thesis provides the reasons for and the background information to the study. In addition, the aims and objectives are set out together with the study design. The chapter ends with this brief summary of what to expect from the following chapters of the thesis.

Chapter 2 is occupied with the initial review of the literature. The first review explores the current knowledge about guidelines on antenatal care in the EU. This includes a search for national guidelines, as well as for trials comparing different schedules of care.

In Chapter 3 on material and methodology it is argued that the concepts underlying guidelines need to be critically analysed before the research plan can be devised. Hence, in this chapter the critical analysis of the theoretical aspects of the thesis can be found. In order to demonstrate the rigour of the reviews, the selection and the use of the literature is described in detail along the development of the theoretical discourse. To ease the delineation of the underlying concepts, Chapter 3 is divided into three parts. Part 1 is focused on the theoretical basis of guidelines, while part 2 addresses health care and decision-making in the EU. Part 2 ends with an analysis of the implications of the reviews for the study, before the detailed research plan is devised in the third part of this chapter.

Chapter 4 details the results of the survey, reporting the response rates and the sources of national guidelines as well as their actual content. Following this, the findings from the survey are compared to the current evidence-base. In the second part of Chapter 4, the results from the in-depth appraisal of the two national guidelines are presented.

After the detailed description of the results in the previous chapter, the findings of all phases of the study are discussed and synthesised in Chapter 5. Special care was taken to relate the findings to what was previously known and to demonstrate the implications of the study, while acknowledging its potential limitations. At the end of the chapter, it is demonstrated what new knowledge is gained from the study, and what might be researched in the future.
Chapter 2 - Review of the literature

2.1 Introduction

To achieve the aims and objectives of the study, a broad review of the literature was necessary. The following chapter provides a systematic review of the literature on antenatal care in the EU, which lays the basis for the study. In particular, it was necessary to establish, what guidelines existed in the member states on antenatal care. Special care was taken to describe the databases and the methods used for the identification of relevant publications.

2.2 Selection and use of the literature

2.2.1 Antenatal care in the member states of the EU

It is the aim of this review to identify and synthesize the relevant evidence from the published literature in order to provide a detailed picture of the current recommendations for antenatal care in the countries of the EU. In order to describe the situation exactly, a systematic review of the literature was conducted. The operational definition of systematic review is the systematic and exhaustive search for published material according to predefined criteria, which are presented in detail within this chapter. It relates not to the systematic synthesis of the results of the identified material, resulting in a meta-analysis of the findings. The analysis and presentation of the review's findings is critical and comparative, though descriptive, as it aims at providing a detailed picture of the situation. The search for relevant literature started with a search for national guidelines on antenatal care from the member states of the EU. Unfortunately, national guidelines seem to be published in a manner inaccessible to an international community. As a consequence, the search was extended to publications on the following:

- studies comparing antenatal care between the member states of the EU
- national guidelines on antenatal care from the member states of the EU
- timing, number and scheduling of antenatal visits in the member states
- randomised controlled trials testing new models for antenatal care

Databases

In order to demonstrate the detailed track record of how the literature was identified, all accessed databases are listed in Appendix 1 to the thesis. In that register, the databases can be found together with the abbreviations used in the text, as well as the source, which provided the gateway to the databases. To ease reading, in the following text only the names or the reference numbers of the databases are mentioned.

As basis for the literature review on antenatal care, the database Medical Literature On-Line [MEDLINE] was used (Appendix 1, Reference 11). MEDLINE was chosen as primary source, as it is the most commonly used source when searching for medical publications, including reproductive biology. With about 4,800 journals, it covers the widest range of national as well as international medical journals (National Library of Medicine 2005, Polit & Hungler 1999: 88). However, antenatal care is not only provided by obstetricians and other physicians, but also by midwives. Therefore, the specific databases “Midwives Information & Resource Service” [MIDIRS], and the “Cumulative Index to Nursing and Allied Health Literature” [CINAHL] were used to complement the searches (Appendix 1, References 12 + 6). The latter contains excerpts from more than 950 nursing journals, in which articles by and for midwives might also be published (Polit & Hungler 1999: 85).

Initially, the search was performed from 1993 to date, which marked the past decade at the time of searching, but it was subsequently enlarged to go as far back as to when the electronic databases were available. In particular very few studies were found, in which antenatal care between member states of the EU was compared. It was hoped to identify some additional studies on this specific topic by going further back in time in the literature. This was done in order to gain more information about the differences and similarities of antenatal care in the
member states, but also to get more information about the methods, which other authors used to compare care schemes. To detect any relevant material, MEDLINE was searched from 1966 to date, the MIDIRS database was used in the most recent version, and CINAHL was searched from 1982 to present. As it was possible with that approach to pick up early and possibly outdated publications, it was decided that on the basis of those, comments could be made about the development of antenatal care in Europe. To prevent false conclusions, it is indicated that the current state of the art might be different, when old references are used.

A problem when searching for national guidelines is that they are not necessarily published in journals, which are included in international databases. However, if a national journal is of high profile, there is a high probability of it being referenced in MEDLINE, and that at least an English abstract is available. Nevertheless, even major guidelines published in the journals of the national professional bodies might not be covered by the aforementioned search strategy. Even worse is the situation, when national guidelines are issued by the government of a country, and are published in legal or administrative journals, rather than in medical ones. Due to this, the search for national guidelines on antenatal care from the member states of the EU was extended to other sources. First of all, the website of the International Confederation of Midwives was searched in order not to miss potential guidelines issued by professional organisations of midwives (Appendix 1, Reference 10). This international non-governmental organisation unites 85 national midwives' associations from over 75 countries. In addition to that, the database of the World Health Organization on maternal and infant health was searched (Appendix 1, Reference 20). Unfortunately, this search did not bring up information about antenatal care in the individual member states of the EU. In order to gain at least access to the national guidelines from the two member states, which will be analysed in depth, the registers for evidence-based clinical guidelines of the German and the British medical societies and the societies of obstetricians were searched. Why these two guidelines were selected for the in-depth appraisal is explained in the chapter on material and methodology.

To cover guidelines from the UK, the guidelines section of the Royal College of Obstetricians and Gynaecologists' homepage was searched (Appendix 1, Reference 16). Moreover, the
clinical guidelines section of the National Institute for Health and Clinical Excellence [NICE] was searched (Appendix 1, Reference 14). This was perceived as an important source, as NICE is the independent organisation responsible for providing national guidance on the promotion of good health and the prevention and treatment of ill health within the National Health Service for England and Wales. However, the NICE guidelines are usually developed under the auspices and with funding of the National Institute for Health and Clinical Excellence, but in one of seven National Collaborating Centres. These centres set up the guidelines commissioned by the Department of Health for England and the Welsh Assembly Government. The most relevant centre for this study is the National Collaborating Centre for Women’s and Children’s Health. However, the guidelines of this centre can also be accessed via the homepage of NICE (Appendix 1, Reference 14). In addition to the databases targeted at guidelines for England and Wales, for Scotland the guidelines section of the Scottish Intercollegiate Guidelines Network [SIGN] was used (Appendix 1, Reference 17). SIGN was established by the medical Royal Colleges to develop evidence-based national guidelines for National Health Service Scotland. It provides excellent information on guideline development and guideline quality, but does not issue a guideline on antenatal care.

For Germany, the guidelines database of the “Ärztliches Zentrum für Qualität in der Medizin” [ÄZQ] was used, the central German institution for quality assurance in medicine (Appendix 1, Reference 1). The members of this institution are recruited from the German Medical Association and the National Association of Statutory Health Insurance Physicians. From this starting point on, the links to the German as well as the international databases on guidelines were followed. This was an efficient way, as under these internet addresses a search engine is operated, by which international guidelines can be located according to medical specialties and specific topics. No additional sources were accessed, as the investigator knew from her professional background that there is only one official guideline available from Germany.

For searching publications about the timing, number and scheduling of antenatal visits in the member states, as well as for randomised controlled trials testing new models for care, an additional source for locating literature was used. The databases for the additional searches
were the Cochrane Database of Systematic Reviews and the Database of Abstracts of Reviews of Effectiveness of the Cochrane Collaboration (Appendix 1, Reference 4). The Cochrane Collaboration is a well known source for rigorous up-to-date systematic reviews about the effects of health care interventions (The Cochrane Collaboration 2004). However, it is important to note that the systematic reviews of the Cochrane Collaboration select studies with a randomised controlled trial design only. The argument for this was that there is “general acceptance that this study design will lead to the most reliable estimates of effects” (The Cochrane Collaboration 2004). Although this can be accepted when investigating into the effects of interventions, using such data only would not have sufficed for the aims of this thesis. For locating studies comparing antenatal care between countries, or to find national guidelines on antenatal care from the member states of the EU, the Cochrane databases would not have led to suitable results, as it references no such studies.

Search strategies: languages, search terms and additional strategies

Although the official language of the searched databases is English, it is possible to identify publications in other languages when using MEDLINE. In such cases, only the abstracts are in English. As English is the main publication language within the scientific community, it is likely to identify comparisons between member states of the EU, or publications on new models of care in English. However, it is very likely that national guidelines on antenatal care are not published in English, but in the national languages. Although this problem could not be overcome completely, it was decided to include at least articles in German and French into the review. German was chosen as it is the native language of the reviewer, and her French is sufficiently good to enable an understanding of publications related to the subject of the study.

Before entering the databases, the keywords for the searches were defined. For this, the Medical Subjects Heading thesaurus / MeSH tree of the US National Library of Medicine was used, which is the hierarchically organised system used for indexing articles (Appendix 1, Reference 19). Two approaches were used for identifying the best keywords. To start with, the MeSH tree was searched from top down, starting from 'Health Care' down to 'guidelines'. In a
second step, the terms 'antenatal' and 'prenatal' were used as a starting point, from which on
the search was broadened to identify more keywords by the respective cross-references.

In order to identify studies, which provided a direct comparison of antenatal care between
member states of the EU, the following search terms were used: 'antenatal' and 'antenatal
care', 'prenatal' and 'prenatal care', 'antenatal visits', 'prenatal visits', 'antenatal screening',
'antenatal tests', 'prenatal testing', 'pregnancy', 'national', 'guidelines', 'regulations', 'clinical
guidelines' and 'pregnancy care'. Also a combination of search terms was used. In addition to
that, all terms were also truncated, e.g as antenatal test*, in order to bring up studies
referenced under antenatal tests, as well as antenatal testing. After this, the names of the pre-
May 2004 member states of the EU were cross-matched with the results of the searches
mentioned before, e.g. 'German*' and 'antenatal' as well as 'German*' and 'antenatal care'.
Also 'Europe' was used to cross-match the searches. To identify studies about the timing,
number and scheduling of antenatal visits, as well as to locate national guidelines on antenatal
care from the member states of the EU, the same search terms were used. Also trials testing
new models for antenatal care should be covered by this search.

As search terms for the Cochrane databases, the following keywords were used in their full, as
well as in truncated form: 'antenatal', 'prenatal', 'pregnancy' and 'care schemes'. Moreover, all
tests recommended for antenatal care in the member states of the EU as identified in the
previously mentioned literature search were used as search terms, e.g. 'hepatitis', 'rubella',
and 'ultrasound', etc. However, the titles of the systematic reviews were additionally hand-
searched for relevant studies. As Cochrane reviews represent a very high standard in
systematic reviews, this effort was judged as being justified. Using this strategy, it was to be
avoided to miss any relevant review for reasons of using the wrong key words.

After the keyword searches in the electronic databases, additional searches were conducted
with different strategies. The databases MEDLINE, MIDIRS and CINAHL were searched again,
using the names of the authors of the most significant papers for a search on authors' name.
As examples, Elina Hemminki and Mika Gissler were identified as key publishers for Finland,
Marion Hall for the UK, and Béatrice Blondel for France. These authors were for example involved in several relevant publications from their respective countries, and they were cited also in international publications.

In addition to the above mentioned searches, several journals were hand searched from 1999 to date. This was done to identify possible additional material not covered by the database searches. The hand-searched Journals were MIDIRS Midwifery Digest, Der Frauenarzt, HebammenForum, and Deutsches Ärzteblatt, where most medical guidelines were expected. As a last step, the references of all publications identified by the methods mentioned above were searched. By this, any seminal papers that had been missed by the other search strategies where then included in the evaluation of the current state of the art.

**Methods of the review: selection criteria and data analysis**

All trials identified by the methods described in the search strategy were scrutinized by the review author for their properties with regard to the studied topics. Before the selection of the literature commenced, the concept of antenatal care was operationally defined on the basis of the literature (National Collaborating Centre for Women’s and Children’s Health 2003):

Only literature that focused on the baseline clinical care of healthy women with uncomplicated singleton pregnancies was considered.

Moreover, only routine antenatal care up to the estimated date of delivery was included in the review. Every piece of literature reporting on a relevant aspect of care was reviewed, regardless of its methodological quality. However, as it is a critical review of the literature, problems regarding the quality of publications are clearly addressed.

As it was aimed at providing a detailed picture of the current recommendations for antenatal care in the member states of the European Union, it was necessary to synthesise the relevant evidence. In order to describe the situation exactly, narrative analysis was the overall method of choice for analysing and demonstrating the results of the comprehensive descriptive literature review. The findings of the review are presented under a series of subheadings, with
a commentary relating to the respective literature (Forbes & Griffiths 2002), including the
critical appraisal of the literature as well as their relevance to the subject.

The search for studies comparing antenatal care between the member states of the EU
brought up five publications (Blondel et al. 1985, Hemminki & Blondel 2001, Heringa & Huisjes
included the member states of the EU either completely, or exclusively. Moreover, all of them
are either focused on the general organisation, or on the provision of care, rather than on its
content. Acknowledging these limitations, the studies are critically analysed and presented in
the review. However, as no publication could be identified, which compared all national
guidelines of the member states of the EU, the search strategy was broadened by searching
for individual national guidelines.

Despite an extensive search for national guidelines on antenatal care from the member states,
only limited information can be presented. When searching for literature in English, German or
French, the national guidelines of most member states are very likely to be missed. They are
usually not published in the journals accessible to foreign researchers, and they are likely to
use titles and keywords in the national languages only. Therefore, the search strategy was
again broadened.

As only the complete national guidelines for Germany and England and Wales were accessible
to the reviewer, indications were sought of what might be contained in the other national
guidelines through the analysis of general papers on antenatal care in the member states.
Additional information was sought from publications on certain aspects of care, which
originated from the EU-15 States. An example for such a search is the combination of
‘hepatitis’ and ‘Italy’ as search terms. The search on this topic was complemented by the
findings from the Cochrane databases, especially with the results from the publication on

However, picking information from studies only marginally related to the original search topic is
an extensive task. In addition to that, it has to be kept in mind that the extracted information is
only a by-product of these publications, which makes it less reliable. As a consequence, it was decided to limit the review to the EU-15 states. In addition to limiting the review to the EU-15 member states, no attempt was made to access grey literature, or to consult national experts on the subject. It was decided that the gap in knowledge was sufficiently large to justify the structured collection of reliable and comparable data.

In order to prepare the extracted information from the literature on national guidelines for comparison, a structured way of presenting the data was required. However, no tool for doing so was found in either of the publications identified during the literature review. As a consequence, a tool needed to be developed to disentangle what is recommended with regard to the number and distribution of visits. Moreover, the tool also needed the ability to demonstrate, which tests are recommended and at what time during pregnancy. For this, a comprehensive table was developed, in which all required information could be recorded. As this tool was also used for the survey, it needed to be developed and piloted carefully. How this was done, and what the properties of the tool are, is described in detail in the methodology section of the survey. The entire survey tool is included as Appendix 2 to the thesis.

After it was found that the search on studies comparing antenatal care between the member states of the EU did not bring up a single satisfying study or report, it was decided that additional evidence was needed about the effectiveness of different models for antenatal care. In the introduction to the chapter on antenatal care in the European Union, the uncertainties around the clinical effectiveness of antenatal care are debated (World Health Organization 2003). Also the methodological problems in attributing outcomes identified in research to any one particular aspect of care are addressed there (Department of Health 2005). On the basis of this, it was decided to conduct an additional search on randomised controlled trials comparing different schedules of care. By this search, also a systematic review and meta-analysis of randomised controlled trials of routine antenatal care was identified, which was performed on behalf of the World Health Organization (Carroli et al. 2001). With regard to the quality of the literature review here, it was reassuring to find that the underlying studies of the WHO review were also detected by the primary search strategy used to retrieve the literature.
for the thesis. This small triangulation of methods indicates the validity of the search strategy.

One important aspect when using the study of the World Health Organization [WHO] as core reference is that the literature search of the WHO was completed by December 2000. As the final thesis assesses the usefulness of a common minimum guideline on antenatal care for the member states of the EU, it was considered crucial not to stop this search for literature before handing in the thesis. Hence, care was taken not to miss any relevant study published after the completion of the systematic review. Although special care was taken to locate studies that provided contradictory evidence to what was found by the WHO, no such study was found. The question of whether this is due to the fact that studies with positive results are more likely to be published, or whether the evidence is unequivocal can not be answered finally. However, the most interesting and important finding of this search was that the WHO developed what was called "The WHO new model of antenatal care" (World Health Organization 2003). As this WHO model for antenatal care is of utmost importance for the thesis, it merits its own section in which it is subject to an in-depth analysis.

Limitations

The classical definition of a systematic review as being used for locating, appraising and synthesising evidence from scientific studies in order to provide informative empirical answers to scientific research questions is not fulfilled completely by the literature review demonstrated above (NHS Centre for Reviews and Dissemination 2001). Although it used open, rather than empirical questions, it was nevertheless conducted systematically, aimed at being comprehensive, and tried to cover pre-defined fields around the research problem. This approach was considered useful for providing a detailed picture of the situation, setting the scene for the study by identifying gaps in and problems with the published literature to date.

Despite these problems with locating and accessing the content of the relevant literature as previously discussed, a strategy was developed to overcome these problems at least partially. An attempt was made to minimise selection bias by the comprehensive search strategy, and the literature that has been identified is generally judged as representative for the current state
of the art in the field. However, the track record of this literature review suggests a significantly large gap in the published literature, which justifies a separate study. To support this, the main results and the conclusions to the literature review are described in the section on antenatal care in the European Union. Although it could be criticised that the adopted search strategy leads to an incomplete picture of the situation, the huge differences with regard to the amount of detected information suggests that the search strategy was efficient for detecting the relevant literature, if this was available.

2.2.2 Implications, advantages and disadvantages of guidelines

On completion of this thesis, it should be possible to answer the question as to whether a common minimum guideline on antenatal in the EU is useful, and what its possible advantages and disadvantages would be. To achieve this, a search for already existing national guidelines does not suffice. Hence, the review was extended to the literature on guidelines and especially guideline development. In this part of the review, it was aimed at providing a comprehensive thematic analysis of the implications, advantages and disadvantages of guidelines, which are important in the context of antenatal care. It was planned to identify all relevant arguments in favour and against guidelines, but not to identify all publications concerned with them. Searching and reading therefore proceeded until no new aspects relevant to the study were found, which is equivalent to a saturation of the data.

According to the above criteria, the search commenced with a broad reading on evidence-based medicine, starting from a widely used textbook from Germany called "Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis" (Kunz et al. 2000). In this, the background of evidence-based medicine is addressed, as are the techniques and the practical application of it. Moreover, the book provides a comprehensive collection of internet resources on the subject. However, an important finding when searching for relevant publications was that the literature on guidelines is developing fast. Although there are excellent textbooks to lay the basis for an understanding of the subject, it was found that there is a lively discussion in
scientific as well as professional journals. In addition to the classical print media, a lot of the argument takes place on the internet. Due to this, several internet sources were used as a basis for this review. At the beginning, the thesaurus system of the US National Library of Medicine was searched for the relevant medical subject headings (Appendix 1, Reference 19). The identified MeSH tree was the following:

Health Care

  Health care quality, access, and evaluation

       Quality assurance, health care

              Guidelines

                   Practice guidelines

       Quality of health care

              Health care evaluation mechanisms

                   Guideline adherence

In addition to the search with search terms, a wealth of national as well as international websites was identified on the development and appraisal of guidelines for clinical practice (Appendix 1, References 1, 3, 8, 13, 14, 16, 17 + 20). They provide search engines and links to each other, leading step by step into the depth of concepts, as well as to the original papers. As the references for the original literature are cited alongside the evidence used in the findings of the review, only the organisations and their entrance sites are listed in Appendix 1 to the thesis, which were used as gateways to the topic.

2.3 Antenatal care in the member states of the European Union

2.3.1 Basic documents

It is the aim of this literature review to identify and synthesise relevant evidence from the published literature to provide a detailed picture of the official guidelines regulating antenatal care in the countries of the European Union. The findings are critically analysed and
synthesised. In the following, the focus is put on the screening tests recommended as well as the timing and scheduling of them. In addition to that, the results of randomised controlled trials testing new models for antenatal care are critically evaluated. Before investigating the details of its organisation and components, antenatal care is defined and explored for its properties.

The WHO defines antenatal care as the complex of interventions that a pregnant woman receives from organized health care services (World Health Organization 2003). Again according to the same authority, the purpose of antenatal care is to prevent or identify and treat conditions that may threaten the health of the foetus/newborn and/or the mother, and to help a woman approach pregnancy and birth as positive experiences. To a large extent antenatal care can contribute greatly to this purpose and can in particular help provide a good start for the newborn child. The operational definition for antenatal care used within this thesis is not as comprehensive as that of the World Health Organization. Although all the components named by the WHO are important and well established aspects of complete and good antenatal care, they all have their own body of literature, and their own methodologies used to research them. Preparation for childbirth, antenatal classes, and maternal expectations from care are only some of them, which will not be explored within this thesis.

The operational definition of antenatal care

In this study, antenatal care is used synonymously to the routine diagnostic care for normal pregnancies. The exact operational definition used is that antenatal care is the baseline clinical care of all pregnancies of a healthy woman with an uncomplicated singleton pregnancy (National Collaborating Centre for Women's and Children's Health 2003).

Moreover, only routine antenatal care up to the estimated date of delivery is considered. Care beyond this point can be defined as the treatment of an alteration from the normal course of events, and the line between normal and abnormal can be drawn at different points after the estimated date.

For this study, it was decided that antenatal care is studied on the population / public health
level, rather than at the level of the individual pregnant woman. Public health is operationally defined as measures and strategies that are occupied with ensuring and improving the health of populations. Treating individual patients is not at the centre of interest (Department of Health, Chief Medical Officer 2003). As a consequence, the focus is put on national guidelines.

Clinical properties of antenatal care

Another important aspect to discuss with regard to antenatal care is its effectiveness and cost-effectiveness. Although antenatal care is supposedly always provided with the best intentions in mind, the dimension of its positive properties can be and are debated. There are measures in antenatal care, which are very likely to improve outcomes, such as identifying and treating severe anaemia. For other measures, the proof of effectiveness and cost-effectiveness is not that easy. This is especially the case, as measures in antenatal care are normally not taken in isolation, but are provided as part of a package consisting of several tests and interventions.

Also the WHO (World Health Organization 2003) acknowledges the uncertainties around the clinical effectiveness of antenatal care, especially in relation to its costs. The WHO stated clearly that given the limited resources of health care and the wide range of services provided as part of antenatal care, such questions must be dealt with and highlights that care should be “appropriate, cost-effective and based on the needs of the specific pregnant woman”.

As the National Service Framework for Children, Young People and Maternity Services (Department of Health 2005) states under section 11:

“..., designing research to take account of the many factors (such as multi-disciplinary working, women’s views and clinical outcomes) that contribute to the delivery of high quality care for women is challenging and it can be difficult to attribute outcomes identified in research to any one particular aspect of care.”

According to this, later in this chapter, also studies will be analysed in which complete models of care are compared. The issue of effectiveness is critically followed throughout the entire thesis, but it is specifically discussed in the sections on evidence-based care, evidence-based
policy and the section on guidelines. Although it is not at the centre of interest in this thesis, another important aspect around antenatal care is its relative expensiveness. Costs must be viewed especially critical with regard to the clinical effectiveness of antenatal care. Unfortunately, this is beyond the borders of what this study can achieve.

The coverage of antenatal care

To get an impression about the coverage of antenatal care in the EU, data from the former 15 member states are presented. From these data it becomes obvious that at least the EU-15 States share the idea that antenatal care should be provided to all pregnant women. As can be seen from Table 2.1, all of the EU-15 states provide a system for antenatal care, and the uptake is generally high. Table 2.1 shows that in 12 of 15 countries at least 90% of pregnant women have contact with the health care system. Unfortunately, no national data for Greece, Ireland and the UK were available to the WHO (World Health Organization 1997).

Table 2.1: Coverage of antenatal care in the EU-15 states (World Health Organization 1997)

<table>
<thead>
<tr>
<th>Country</th>
<th>Uptake of antenatal care in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>100</td>
</tr>
<tr>
<td>Belgium</td>
<td>90</td>
</tr>
<tr>
<td>Denmark</td>
<td>100</td>
</tr>
<tr>
<td>Finland</td>
<td>100</td>
</tr>
<tr>
<td>France</td>
<td>99</td>
</tr>
<tr>
<td>Germany East</td>
<td>100</td>
</tr>
<tr>
<td>Germany West</td>
<td>98</td>
</tr>
<tr>
<td>Greece</td>
<td>no data</td>
</tr>
<tr>
<td>Ireland</td>
<td>no data</td>
</tr>
<tr>
<td>Italy</td>
<td>100</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>98</td>
</tr>
<tr>
<td>Portugal</td>
<td>96</td>
</tr>
<tr>
<td>Spain</td>
<td>96</td>
</tr>
<tr>
<td>Sweden</td>
<td>100</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>95</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>no data</td>
</tr>
</tbody>
</table>

Although according to the WHO (World Health Organization 1997) a coverage of antenatal
care is achieved in over 97% of all pregnancies, it has to be noted that all pregnant women having contact to a health professional at least once during pregnancy were counted. This is definitely not the quality of antenatal care, this study aims at. Unfortunately the number of contacts needed to be counted as having had antenatal care is not known for the EU-funded PERISTAT project. In this study, data were extracted from official statistics of the years 1997 to 2001 (Wildman et al. 2003). However, as the uptake of antenatal care was not at the focus of the publication, only a few facts are mentioned. In this publication it was stated that 2.5% of women in Greece did not receive antenatal care. Moreover, it was found that less than 1% of women of all other EU-15 States were without any contact with the health care system during pregnancy. This would mean that in Ireland and the UK less than 1% of women did not receive antenatal care, or that incomplete data were provided. However, from personal contact with British midwives and women, the first option is more likely to be the case. Hence, the nearly full coverage of antenatal care again reinforces the idea that antenatal care is a major public health issue for the EU. Although not yet complete, it is likely that over the coming years such data are made available for the states that joined on 1 May 2004 with the efforts for compatible or common statistical systems within the EU. Up to now, such data are not available. Only the World Health Organization (2003) acknowledges possible problems with access to antenatal care services in some countries. The authors from the WHO state that in Western Europe, full coverage for antenatal care appears to exist. However, in Eastern Europe, the countries which have made a transition from the Soviet model of health care to social insurance were found to provide full coverage. For the 'other countries', the authors report anecdotal evidence which indicates problems of access due to incomplete coverage, geographic problems etc. In contrast to that, there might be particular other reasons why there were no data available for Greece, Ireland and the UK. It could be possible that these are the member states which do not have national guidelines on antenatal care, and as a consequence of this no national recording system for local services. Later in this literature review, an answer should be possible whether no data means no service provision, or whether there might be other reasons for the non availability of data to the WHO in 1997.
Drawing all the above mentioned facts together, it can be concluded in line with the WHO that

- every pregnant woman should have full access to antenatal care.

- excessive, unneeded and unproven interventions are often provided to women with normal pregnancies (World Health Organization 2003).

All these factors make antenatal care to an important issue in public health. As the WHO (World Health Organization 2003) stated clearly:

"Any health care programme that sincerely wishes to improve the health of its population must pay serious attention to the health of the pregnant woman and her fetus."

Public health aspects

Antenatal care is directed at the health of women, but also aims at ensuring the best possible starting conditions for the future citizens of the EU. It contains aspects of health protection, as it screens for harmful maternal conditions caused by pregnancy, and tries to treat them immediately. Antenatal care has also major aspects of health promotion, as it aims at improving intrauterine conditions for the foetus. Possible examples for this are ensuring a good maternal diet, and ensuring good oxygenation by controlling or improving maternal haemoglobin levels. However, it has to be noted that antenatal care can be studied on the basis of the individual pregnant woman, or at the population and public health levels.

Health policy is influenced by several internal factors, such as political systems, culture, religion, and ethnicity. Moreover, there are exogenous factors that might influence health policy, such as threats from war or terrorism, and the prevalence and marketing strategies of international or multinational companies. These factors have the potential to influence health policy (Walt 1994: 3ff). However, one of the common traits of the European Union is that the member states share most of these factors and values. As is exhaustively discussed in the chapter on general trends, reproductive health and patient mobility, the EU takes an active stance in public health in the years to come. The 6th framework programme for research and
development for the years 2002 to 2006 has set three major public health themes on the agenda (Commission of the European Communities 2000, CORDIS 2002, European Commission Research Directorate General 2001). Exactly the same aims were set out in an extended version in the Union’s public health programme for the years 2003 to 2008 (Commission of the European Communities 1998, European Parliament and the Council of Europe 2002, Watson 2001). These aims are to

- improve health information and knowledge,
- respond rapidly to health threats, and to
- address health determinants.

With its abilities to address health determinants, which means preventing ill health before it develops, antenatal care fits perfectly into the public health context of the EU (Byrne 2004). However, as is demonstrated and criticized in the sections on EU health policy, it is observable that the specifications from sides of the EU relate mainly to major illnesses of older people, rather than taking a more general stance (Commission of the European Communities 2005: 7, CORDIS 2002, Kyprianou 2005b). According to the above mentioned EU documents and programmes, premature deaths due to major illnesses of adulthood should be reduced by addressing some of the underlying causes of major illnesses, such as lifestyle behaviours, socio-economic circumstances and the environment. However, the important contribution antenatal care has to make in this respect is neglected. By considering a healthy lifestyle and a healthy diet throughout pregnancy, as well as by treating unfavourable conditions immediately, best starting conditions are ensured for the future citizens. As research suggests, conditions during the intra-uterine period can programme an organism for the rest of its life (Barker 1995, Barker et al. 1993, Paneth et al. 1995, Eriksson 2005). As this is especially evident for cardiovascular diseases – one of the major health problems in Europe – antenatal care fits perfectly into the public health strategy of the EU (Commission of the European Communities 2005: 44).

Up to now, Children's health is not mentioned in the documents on public health in the EU at all, although it would be of utmost importance to investigate into this aspect of public health as
well. Children are citizens in their own right, but not yet able to act as self-advocates, particularly at the population level (Rigby et al. 2003). Here, the EU could make a unique contribution towards the health of its citizens, as children’s health determines the health of the future population (Rigby et al. 2003). As children suffer particularly from unfavourable conditions in their (social) environment, ensuring best starting conditions for all future citizens would be also an important act with regard to the equality of citizens of the Union.

An initial critical analysis of the literature has revealed six major themes, namely, antenatal care in the EU, timing and number of ante-natal visits, tests recommended for antenatal care, guidelines on antenatal care in the member states, comparisons of different schedules of care and the WHO model of care. These will be discussed in turn.

2.3.2 Antenatal care in the EU

Studies comparing antenatal care in the EU

All in all, five papers were found that provide a comparison of antenatal care in the countries of the European Union (Blondel et al. 1985, Hemminki & Blondel 2001, Heringa & Huisjes 1988, Langer et al. 1999, World Health Organization 1987). However, they are either focused on the general organisation, or on the provision of care, but not on the guidelines according to which care should be provided. Although these aspects form the context of antenatal care, guidelines that specify what care should be provided represent the core decisions about what is seen as optimum care in the member states. They therefore justify a separate investigation.

Langer and colleagues reported on antenatal care practice in nine settings within eight countries of the EU (Langer et al. 1999). Included were Belgium, the United Kingdom, Luxembourg, The Netherlands, Spain, France, Germany and Italy. In this study, the policy of individual departments was assessed, rather than the policy of the countries themselves. Moreover, it does not become clear, why not all member states of the EU were included, or how the departments were chosen. In addition, with the chosen questionnaire survey there
might be more information about what care was intended to be provided in the settings, rather than what care does in fact take place there. There were neither additional observations, nor did the authors perform any counter-checks with data from antenatal records. In spite of the mentioned methodological limitations, the study demonstrates that there are significant differences in the routine use of clinical examinations and investigations performed during the first visit in pregnancy between as well as within countries.

Already in 1985, Blondel and colleagues highlighted the scarcity of details about medical practice in European countries, and concluded that there are not sufficient and no reliable data on recommendations on antenatal care in EC countries (Blondel et al. 1985). Also the differences and similarities of care between the countries were demonstrated. Unfortunately, the situation has not changed much since then. The rest of the study with the Perinatal Study Group of the WHO Regional Office for Europe was mainly occupied with organisational characteristics of antenatal care in 13 European countries. Again, not all member states of the EU were included. Reference is made to the situation in Belgium, Denmark, Finland, France, Luxembourg, The Netherlands and Sweden. As the data stem from 1981, Germany was still divided in the Federal Republic of Germany and the German Democratic Republic. Moreover, England and Wales were analysed separately from Scotland. Northern Ireland was not included at all. This makes the use of findings more difficult. In general, their findings were that most countries demand a minimum number of visits necessary for the women to be eligible for prenatal allowances. It was also found that the required number of visits varies in the EU countries from five in Luxembourg to 14 in Finland. In addition, some organisational features of antenatal care were mentioned. It was found that the Scandinavian countries and The Netherlands give an important role to midwives and to a lesser degree to general practitioners. In other countries, mainly obstetricians were found to be responsible for providing antenatal care. As examples, Belgium, Germany and Luxembourg are mentioned. For France and the UK it was found that obstetricians frequently carry out visits, but rarely provide the entire care.

In a later study of Hemminki and Blondel, 13 European countries were included (Hemminki & Blondel 2001). Unfortunately, not exclusively and not all member states of the European Union
were included, and there is no reason provided for this. Data are available for Finland, Sweden, Germany, Portugal and Denmark. For these and the other countries studied, the authors identified 13 different organisational systems for providing antenatal care. The only common features were that in all countries care is provided either free, or at low cost for the women. However, the two most important findings of the study were that only six of the studied countries have a dominant system according to which antenatal care is provided. All other countries have parallel systems in operation. However, although different professionals might provide antenatal care in different organisational systems, they also found that most health authorities or other national bodies provide guidelines and other recommendations for care provision. The existence of guidelines was also reported by other authors, e.g. with Blondel and co-workers citing recommendations or regulations from Finland, Luxembourg, The Netherlands and Sweden (Blondel et al. 1985). Backe & Buhaug reported that guidelines exist for all Nordic countries (Backe & Buhaug 1994). However, as guidelines are often published in the national languages only, they do not lend themselves to international comparisons.

Definitely the most comprehensive study is that of the World Health Organization's study group on perinatal care in the WHO European region (World Health Organization 1987). Between 1981 and 1982, the group conducted a postal survey on routine antenatal care in 31 of the 33 countries of the European region. Unfortunately, the WHO does not provide the names of the countries, to which any findings apply. It is for example just stated that 21 of the 33 countries recommend a specific examination or a certain pattern of care, but there is no hint, which countries might be meant. As this applies to the entire publication, only general information can be extracted. Moreover, much information is provided about who is doing the examinations, how many visits are planned, and where the visit is taking place. Less information was collected about the actual content of care.

The most significant findings were that in 21 of the 33 countries examined the number of antenatal visits is legally specified. 18 countries reported that they have official guidelines on routine examinations that have to be conducted during pregnancy. 12 of them have guidelines issued by the state, 6 countries have recommendations issued by major universities (World
Health Organization 1987: 19ff). However, even after their major survey, the authors concluded that very little information is available about the actual content of antenatal care. This still needs to be brought together, as to how this will be done as part of this thesis.

As background information it is therefore important to know that about 50 years ago, the countries of the European region started to formalise and regulate the provision of antenatal care. Most of them defined the time span between the individual visits, the examinations that have to be carried out routinely, which have to be carried out by a medical doctor or midwife who was accredited by the government. Today, each country was found to have a legally required or recommended scheme for all pregnant women (World Health Organization 1987: 83). However, all different schemes cannot be based on sound scientific evidence. It is evident that those antenatal care schemes including many visits require immense sums of monetary and personnel resources. Although the positive effect of these intensive care schemes is not sufficiently proven, they tend to persist. As this is difficult to change, the WHO recommends putting the focus on the content of care, rather than on the number of visits.

As expected, also in the WHO study (World Health Organization 1987), major differences are reported with regard to the recommended examinations. Blood group testing is recommended for all pregnant women by all but one country. This country recommends blood group testing only on a selective basis. In contrast to that, screening tests for lues are common in 5 countries for all pregnant women, in 19 countries only on a selective basis. Differences in recommendations are even larger for ultrasound examinations. Three countries provide it routinely, 19 on a selective basis. Unfortunately, also here the names of the countries are not given so that no conclusions can be drawn about the effectiveness of the care schemes.

Similar to the other studies mentioned, the publication by Heringa and Huisjes provides information about the provision of antenatal care in the individual countries of the European Union, but not on possibly existing official guidelines (Heringa & Huisjes 1988). These authors did a very detailed study on what care was provided in 1985 in the university hospitals of the countries of the EU. A maximum of ten university hospitals were approached per country for
this EU-funded project. Finally 67 obstetricians from nine member states answered the
questions on 30 different screening procedures in their hospitals. Unfortunately, in 1985 the
EU did not have the current number of member states, so that sixteen of today's members are
not included. These countries are Luxembourg, Sweden, Portugal, Spain, Finland, Austria, and
all states which joined the Union on 1 May 2004.

Variations in antenatal care practice

Some of the most interesting findings are described in the following discussion, as they provide
a very good overview about what can be and is tested routinely during pregnancy. For
example, vaginal examinations were carried out by all studied units at the first visit. However,
in Belgium, France, Germany and Italy, the majority of clinics performed this examination at
each visit. This is supported by the findings of other studies (Bréart 1995, Kristensen et al.
1995). In the other countries, this is not the case. In 1985, all clinicians checked the mother's
blood pressure at each visit. Since then, other patterns of blood pressure measurement have
been tested (Hall et al. 1985). Whether they found their way into official recommendations, or
into practice, is not clear. The same applies to routine weighing of the mothers. Nearly all
clinicians reported that the women are weighed at each visit. Also this practice was debated
over the past years. According to the study of Heringa and Huisjes, 63% of the responding
clinicians routinely perform a cervical smear during pregnancy (Heringa & Huisjes 1988).
Another 33% recommend this practice. In this case it would be interesting to know whether this
practice grounds on any perceived benefit, or whether this follows official guidelines.

In all university hospitals included, the maternal blood group was typed routinely. In 64 out of
67 departments, typing of the Rhesus factor (D) is also included. The three departments
missing did not respond to the question, which could have happened by mistake. The
haemoglobin level is tested by 90% of the respondents at least once during pregnancy. Most
of them screen more frequently. An additional 19% screen for haemoglobinopathies. In the
countries with the highest prevalence of such conditions, i.e. Italy and Greece, all clinicians
recommended screening.
Serological screening for rubella is either routinely performed or recommended by 91% of the clinicians. This finding is not surprising. However, a rather unexpected finding was that 97% of the clinics routinely screen for lues. In one country the test is repeated several times during pregnancy. In contrast to that, toxoplasmosis screening is not widespread. Only 34% of the clinicians test for this condition, mainly in Italy, Belgium and France.

In addition to the routinely, or at least frequently performed tests, there are several tests not performed on a routine basis. Screening for cytomegalovirus is disregarded by 63%, and for listeria monocytogenes by 75% of the clinics. Also for group B β-haemolytic streptococci no routine screening is performed in 61% of the clinics. For hepatitis B virus, in 1985 only 36% of the clinicians screened routinely. Another 19% recommended it at least. It would be interesting to know how these numbers changed during the past two decades. Today, numbers might be much higher, as the subject gained importance. One example is e.g. that during that time routine vaccination against hepatitis B was introduced in Germany.

Relatively simple is the test for bacteriuria, which is routinely performed by 54% of the clinicians. Another 21% recommend it. 66% of the clinicians reported that they performed some kind of glucose tolerance test to screen for gestational diabetes, i.e. impaired glucose tolerance during pregnancy. It might be interesting to know how things changed until today.

The importance of information about ultrasound screening is difficult to assess. In this field, much has happened over time since the investigation took place. It can be estimated that numbers are much higher today than they were in 1985. However, already then 82% of the clinicians carried out a routine ultrasound examination for several purposes, and most of them did so more than once during pregnancy. In addition to these ultrasound examinations, 25% of the clinicians in the studied countries carried out a cardio-tocography during pregnancy. In Belgium, Germany and Italy the majority of clinicians did this. It is not clear whether this practice has increased or decreased over the past decade.

Foetal movement counts on a routine basis were conducted by 64%, the assessment of placental functions by hormones was reported by 22%, and screening tests for coagulation
disorders by 9%. 28% of the clinicians routinely screened for maternal serum alpha-fetoproteïne.

All in all there was no study identified which compared or demonstrated official guidelines on antenatal care in the countries of the European Union. For this, a broad search of the published literature on antenatal care in the countries of the EU was conducted. From these studies, as a by-product some information could be extracted about the guidelines on antenatal care of individual countries. The results of this search are presented in the following.

2.3.3 Timing and number of antenatal visits

In Table 2.2, all available information about the timing and scheduling of antenatal visits in the EU-15 states has been compiled. When different schedules for first and subsequent pregnancies were available, the scheme for primigravidae was used. Unfortunately, only for 6 of the EU-15 states there were data available. From the data shown, a certain pattern of scheduled visits becomes evident. In most countries, women meet a health professional every four weeks until the 28th gestational week. From week 32 on the latest, the women attend every two weeks. Weekly visits are scheduled from weeks 36 or 37 on in Denmark, Finland, The Netherlands, and Sweden. However, over the past years efforts have been made to reduce the number of routine visits, especially for multigravidae. This is already evident in the schedule for Denmark and that of the United Kingdom (National Collaborating Centre for Women’s and Children’s Health 2003). In section 2.3.6 on randomised controlled trials comparing different schedules of care this will be discussed in detail. In general, the comment that the intervals between antenatal visits as introduced in 1929 by the United Kingdom Ministry of Health had been chosen arbitrarily and have little scientific basis might be true, if a reduction of visits can be achieved without limiting the benefits of care itself (Liu et al. 1992). Nevertheless, several countries have adopted and kept this scheme until today.
Table 2.2: Timing and schedule of antenatal visits in the EU-15 states

<table>
<thead>
<tr>
<th>State</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
<th>38</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>DK</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIN</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>NL</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>S</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>UK</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3.4 Tests recommended for antenatal care

In order to demonstrate what tests are recommended for routine antenatal care throughout the European Union, all tests recommended in one or the other country of the Union have been compiled in Tables 2.3a to 2.3c. From the information given in these tables it becomes clear that no pattern could be identified of tests that are recommended throughout all countries of the European Union. However, from the detailed descriptions of the countries in the following section it will become clear that there is only rudimentary information about official recommendations. Only for the countries for which a guideline was available, sufficient information is presented, e.g. for Denmark, Germany, Sweden and the UK. Due to the scarcity and unreliability of information, a comparison of the tests used in the different countries is not valid and should be postponed until more information is available, i.e. from the survey.
Table 2.3a: Antenatal screening tests recommended in the EU-15 states - physical examinations

<table>
<thead>
<tr>
<th>Test</th>
<th>A</th>
<th>B</th>
<th>DK</th>
<th>FIN</th>
<th>F</th>
<th>D</th>
<th>GR</th>
<th>IRL</th>
<th>I</th>
<th>L</th>
<th>NL</th>
<th>P</th>
<th>E</th>
<th>S</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Foetal position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Full physical examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Fundal height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Maternal weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Vaginal examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Table 2.3b: Antenatal screening tests recommended in the EU-15 states - technical tests

<table>
<thead>
<tr>
<th>Test</th>
<th>A</th>
<th>B</th>
<th>DK</th>
<th>FIN</th>
<th>F</th>
<th>D</th>
<th>GR</th>
<th>IRL</th>
<th>I</th>
<th>L</th>
<th>NL</th>
<th>P</th>
<th>E</th>
<th>S</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrasound</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Foetal heart rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
Table 2.3c: Antenatal screening tests recommended in the EU-15 states - laboratory tests

<table>
<thead>
<tr>
<th>Test</th>
<th>A</th>
<th>B</th>
<th>DK</th>
<th>FIN</th>
<th>F</th>
<th>D</th>
<th>GR</th>
<th>IRL</th>
<th>I</th>
<th>L</th>
<th>NL</th>
<th>P</th>
<th>E</th>
<th>S</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-fetoproteine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Blood group</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Chlamydia (cervix)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Coagulation disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Foetal movement counting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Beta-haemolytic streptococci</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Hepatitis B surface antigen</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Hum. Immunodeficiency Virus</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Lues</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Papanicolaou smear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Placental functions (hormonal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Rhesus antibody screening</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Rhesus factor</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Rubella titer</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Triple test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Urinalysis</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
2.3.5 Guidelines on antenatal care in the member states of the EU

The description of the current guidelines on antenatal care in the individual countries follows a certain scheme. The questions are whether there is a national guideline on antenatal care in the respective country and who has issued this guideline and is therefore responsible for its content and update. In addition to that, any information about the intervals at which the guideline is revised and on how legally binding the guideline is was collected.

After this, it is attempted to disentangle what care scheme is recommended for uncomplicated pregnancies with regard to the number and distribution of visits. In this context, it is also demonstrated which tests or interventions are scheduled, and at what time during pregnancy. For this, a comprehensive table is used, in which the individual tests are recorded. This table has been developed during the review process.

Austria

For Austria, no document was found on the regulations or practice of antenatal care. However, a common recording system for results of antenatal screening procedures exists, which is called “Mutter-Kind-Pass”.

Belgium

Belgium is one of the countries for which nearly no data are available in the published literature. However, it was stated by one author that in 1986 no official regulations by the Belgian health system existed, which prescribed or recommended a special scheme or content of antenatal care (Wollast et al. 1986). Another author stated that there was no specific legislation dealing with prenatal care in Belgium (Humblet et al. 1989).

Although a study was identified with the title “Organization of prenatal care in Belgium”, no comments can be made about the content of care. In this paper, a compilation of literature is cited, which is mainly occupied with the question of who is providing antenatal care, rather than what this care consists of. Moreover, it does not become clear for which reasons the cited studies have been included or what aim is followed with this paper (Humblet et al. 1989).
Although these methodological problems leave many questions unanswered, some conclusions can be drawn about prenatal care in Belgium.

General practitioners and obstetricians/gynaecologists mainly provide antenatal care in Belgium. Interesting is that there are major differences in care between the Flemish and the French speaking communities. Humblet and colleagues (1989) state that in the French speaking communities nearly all pregnancies are supervised by obstetricians, whereas in Flemish-speaking communities, only 60% of the pregnancies are supervised by obstetricians.

With regard to special screening procedures, only two policies were mentioned. In Belgium, no standard prevention programme for the vertical transmission of hepatitis B exists (De Groote et al. 1997). However, most gynaecologists were reported to screen for hepatitis B virus during pregnancy, but the type of serology tested and the timing of the screening varied widely. Moreover, the data stem from a preliminary survey amongst 32 Antwerp gynaecologists only.

The second test on which data are available is HIV screening. In 1990, no official guideline regulating HIV screening in pregnancy existed for Belgium. By a postal questionnaire, Belgian gynaecologists have been asked for their screening policy [815 contacted, responses: 54.7%]. 91% of the respondents offer HIV screening to pregnant women. 49.1% of them to all pregnant women, 41.9% to those with risky behaviour. Unfortunately, there is no mention of what ‘risk behaviour’ is and how it is identified (Denayer et al. 1990).

**Denmark**

In Denmark, the Danish National Board of Health issues guidelines on antenatal care, and the 16 Danish counties are responsible for the local application of these guidelines. Within them, reference is made to content and scheduling of visits (Kristensen et al. 1995, The Ministry of Health 1997). Timing and scheduling of the visits has been demonstrated in Table 2.2. The guidelines are called "Directives on Pregnancy Hygiene and Maternity Care", but are published in Danish. The part that could be extracted is compiled in Table 2.4.
Table 2.4: Antenatal screening tests recommended in Denmark - physical examinations

<table>
<thead>
<tr>
<th>Test</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
<th>38</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full physical examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundal height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The tables on technical tests and laboratory tests were omitted, as nearly no information was provided about such tests. With regard to these, it was only found that urinalysis should take place about seven times during pregnancy. However, this does not provide a complete picture.

The Danish association of midwives reported in a personal letter to the author that the National Board of Health has issued new guidelines on maternity care in summer 1998. Also these are not available in English, but they build the basis of the facts cited under the agenda on the basis of the personal communication with the Danish association of midwives (Den Almindelige Danske Jordemoderforening 1998). One publication might provide valuable insights, as it is directed on "Preventive health examinations of pregnant women in Denmark. Structure and organisation". However, this is also published in Danish, and the MEDLINE abstract was truncated at 250 words (Andersen et al. 1989).

Antenatal care in Denmark is shared between general practitioners, midwifery centres and hospital departments, and is generally free of costs for the pregnant woman (Den Almindelige Danske Jordemoderforening 1998, Kristensen et al. 1995). Seven antenatal examinations are to be carried out by a midwife. Three visits are scheduled with a general practitioner, and one visit takes place at 16 to 18 gestational weeks at a specialist hospital department (Kristensen et al. 1995). In contrast to that, Backe and Nakling (1993 citing Kamper-Jørgensen et al. 1986 who published in Nordic languages only) as well as Blondel and colleagues (1985) state that a second examination by a specialist in obstetrics is mandatory in the last trimester.

Some hints towards what care is recommended by the Danish National Board of Health can be gained from a study comparing practice to the official recommendations (Kristensen et al. 1995). For this cross-sectional, nationwide study, a randomly selected group of general practitioners, midwives and hospital doctors received questionnaires on the content of their antenatal care. With 75 to 89% response rates in the different professional groups and large samples of 958 GPs and 678 midwives, some valid conclusions could be drawn about actual practice. The authors concluded that nearly all GPs carried out procedures with unproven value, which were additionally not recommended in the national guidelines.
It is interesting that the authors judged screening for rubella antibodies as insufficient, although this procedure was reported as not recommended in the guidelines. Also cervical smears were frequently carried out although not recommended. Problematically, there are no data available for any of the tests as to whether they were performed at the recommended time during pregnancy, nor whether there were any comments on the optimal timing. Some additionally recommended measures were mentioned, but there was no hint as to when they should be performed. These tests were a check for oedema, clinical pelvimetry, fundal height measurement, urine culture and a test for lues. The only tests recommended to be carried out regularly at each visit are urine analysis and the measurement of blood pressure and maternal weight. In addition, at two visits after the booking visit, a vaginal examination is recommended. These examinations are scheduled during visits to the general practitioner.

With regard to ultrasound scanning, the Danish Board of Health has its own policy. The board did not recommend any routine screening since 1986 (Kristensen et al. 1995, Stoll et al. 2001). In 1995, the issue of routine ultrasound scanning was under consideration as a part of a general revision of the Danish antenatal care programme. However, despite this obvious recommendation, ultrasound scanning found also in Denmark its way into routine practice. Already in 1995, the 49 departments responsible for hospital antenatal care in Denmark offered routinely to 51.4% of pregnant women an ultrasound scan in weeks 10 to 20. 34 of the 49 departments wanted to continue or start offering ultrasound screening (Jørgensen 1998).

Finland

In Finland, antenatal care is based on individual free of charge outpatient centres, which are part of primary municipal health care since the 1940s. Private health services, subsidised by the national sickness insurance since 1964, are responsible for outpatient care. In addition to these, public health services are available. The distribution of public services is under the responsibility of the local authorities, but planning and allocation of resources are directed from the central government by directives and state subsidies (Hemminki 1983).
By the end of the 1950s, the use of maternity centres became common practice and covered 99.9% of all antenatal care in 1978 (Hemminki 1983, Ministry of Health and Social Affairs 1984 [Finnish] cited by Hemminki & Gissler 1993). In these maternity centres, primary physicians supervise care. In practice, mainly public health nurses or midwives provide it. On average, women had 3.5 visits to a medical doctor, and 12.9 visits to a midwife. Traditionally, midwives are responsible for antenatal care and normally there was continuity of carers throughout pregnancy. If the physician is involved in the care, it is in most of the cases a general practitioner who is employed by the state. Only major towns might have a specialist obstetrician available for antenatal care. If specialist care is required, women are sent to hospital outpatient departments. However, it has to be noted that these data refer to the time span between 1950 and 1980. Major changes will most likely have occurred since then (Hemminki 1983, Hemminki & Gissler 1993). The organisation of the Finnish prenatal care system is described as very uniform. According to the nationwide Medical Birth Registry, over 80% of antenatal visits took place in these special outpatient maternity centres between 1987 and 1990 (Hemminki & Gissler 1993). However, in ‘recent’ years, care is reported to shift from the traditional maternity centres outside the hospitals that are run by municipalities to hospital outpatient clinics (Hemminki et al. 1990).

In 1993, an expert group gave out new recommendations for antenatal care in Finland. Unfortunately it is not clear, what experts were represented in this group and how binding and accepted these recommendations are. Problematically, these recommendations are published in Finnish only. From the scarce data available in English it becomes evident that it is recommended that primigravidae attend early in pregnancy, but that healthy women expecting a subsequent child could postpone their first visit up to the 16th gestational week. However, it is not clear how these women should be sure about the gestational week in which they are in (Lumme et al. 1993 cited by Gissler & Hemminki 1994). In contrast to these later recommendations, in 1987, early attendance at maternity centres was encouraged “for full utilization of the potential benefits” (Hemminki & Gissler 1993).
According to the previously cited recommendations, in 1987, Finnish women started antenatal care early in pregnancy. 22% of the women attended before the 8th gestational week, 79% had their first appointment before the 12th gestational week, and only 4% came in the 16th gestational week or later. As a consequence of this early attendance, but also due to a schedule with relatively frequent visits, Finnish women clearly had the most antenatal visits in western European countries between 1987 and 1990 with a mean of 15.2 visits.

Two laws are concerned with the organisation of prenatal care in Finland. The law of municipal midwives originating from 1944 and the law on maternity benefits from 1949. The latter specifies that maternity benefits are paid to women, if early contact with health care providers was established (Hemminki et al. 1990, citing a ministerial paper [Finnish]).

In general, the following schedule of visits was officially recommended in 1984, but was still identical at time of publication of the cited study in 1993: in gestational weeks 6 to 32, one visit in every four weeks. From weeks 33 to 36, one visit every two weeks and in weeks 37 to 40, one visit a week was recommended. After the 40th week, two visits a week were scheduled (Ministry of Health and Social Affairs 1984 [Finnish], cited by Hemminki & Gissler 1993). This recommended schedule, which is also represented in Table 2.2, led to good compliance from sides of the women. In 1987, only 6% of primigravidae and 9% of multigravidae had fewer antenatal visits than the recommendations suggest (Gissler & Hemminki 1994).

Unfortunately, no data on the content of antenatal care are available for Finland. The only fact mentioned in the published literature was that in 1987 there was no national recommendation on any ultrasound scans during pregnancy (Hemminki et al. 1990).

France

Although no official guideline for antenatal care in France was found, a law is reported that prescribes four antenatal examinations during pregnancy. These examinations should take place before the end of third month, during the sixth month, during first fortnight of the eighth month and in the first fortnight of ninth month (Garcia & Saurel-Cubizolles 1983). The first three
visits were also required to be eligible for antenatal allowances in 1983. In 1993 this number was changed to seven required visits (Blondel et al. 1993). In 1983 it was found that less than 5% of the women fail to attend the required visits. For these four visits, some antenatal screening procedures are required and others are recommended. Unfortunately, no evidence could be found about which procedures these are.

From 1983 on, there was a trend towards more frequent visits in France in contrast to the trend towards a reduction of visits in Britain with an average of about seven visits in France and still 12-13 visits in Britain (Garcia & Saurel-Cubizolles 1983). It was common practice in the 1980s to commence with a very early initial visit and to return for one visit per month to an obstetrician or midwife (Rumeau-Rouquette et al. 1984 [French] cited by Blondel et al. 1993).

Germany

In Germany, the Federal Joint Committee has issued guidelines on antenatal and intrapartum care which are based on social law and insurance laws and regulations (Bundesausschuss der Ärzte und Krankenkassen 2003). These guidelines, which are called ‘Mutterschaftsrichtlinien’, are published in the official organ of the German government.

In the preface of the regulations, their purpose is described. It is to ensure that care by physicians is based on the commonly accepted state of the art in medicine and is sufficient, suitable and cost-effective throughout pregnancy, delivery and the puerperium. It is highlighted that physicians should decide on care within this framework and that physicians, insurances and midwives should work together in order to communicate the importance of medical care during this time. At the end of the guidelines it is highlighted that a midwife can also carry out some examinations, such as blood pressure measurement, urine examinations, control of haemoglobin levels and fundal height measurement. However, in 1998, less than 50% of the German midwives provided antenatal care (Mead & Ashton 1996 + 1998: 42).

The German guideline recommends the early commencement of antenatal care, that is, as soon as pregnancy is diagnosed. Similar to other guidelines, the German guideline specifies
what examinations are recommended and at what time during pregnancy. At the first visit, the
detailed history should be taken, including the pregnant woman’s own, family- and
reproductive history. In addition to that, her work- and social circumstances should be
addressed. The history is followed by a general physical examination and a gynaecological
examination. If during history taking a genetic risk becomes evident, the pregnant woman has
to be informed about the opportunity of genetic counselling or screening. In addition to the
regulations for normal pregnancies, an exact definition of risk pregnancies is included,
combined with a schedule for further measures to be taken when a risk has been detected.
However, as the German guideline will be appraised in depth in Chapter 4 of the thesis, its
content is not further discussed here.

Greece

Hemminki and Blondel (2001) report from their survey, that Greece was one of the three
countries that had no national or regional guideline. From the article of Lekea-Karanika and
colleagues (1991), some conclusions about antenatal care in Greece can be drawn, although
the article itself concentrates on risk factors for preterm delivery. Their study takes into account
a retrospective analysis of 10,859 singleton births, which took place during 30 consecutive
days in 1983 in entire Greece. In the study group it was found that an obstetrician with or
without a midwife cared for 99% of the pregnant women. 95% of the pregnancies were cared
for by an obstetrician alone and 1% by a midwife only.

It was also reported that the blood pressure is taken at some point during pregnancy, although
there is no hint as to how often or when this is done. The same applies to tests for proteinuria.
In 67% of the cases in their studied group, the haematocrit levels were reported. Tests for
blood group were made, and the Rhesus factor was known in 94.8% of cases. More
information could not be obtained about antenatal care in Greece by this extensive review.

Ireland

Data on antenatal care in Ireland is more than scarce. The only information that could be
detected by the literature review was a study concerned with hepatitis B testing. Routine antenatal screening for hepatitis B carriage was found to be cost-effective in a sample of 16,222 pregnant women in Ireland. The measure was well accepted by the women, with 99.98% of women agreeing on being tested. The authors of the study recommend routine antenatal screening as a standard of care in Ireland (Healey et al. 2001).

Italy

Italy is one of the three EU-15 countries for which no hints about official recommendations or guidelines on antenatal care were found (Hemminki & Blondel 2001). In nine of the EU-15 states, midwives are generally permitted to provide antenatal care for women experiencing a normal pregnancy. Again, Italy is one of the exceptions that do not allow midwives to provide antenatal care (Mead & Ashton 1998: 41).

Only two measures of antenatal care are mentioned in the literature with respect to Italy. Routine cervical examination for the identification of risk for preterm delivery is reported to be conducted routinely (Bréart 1995). Although this is reported to be common practice, it might also be that it is not an official policy. The second measure discussed is universal HIV-testing. An expert group of the reference centre for HIV and pregnancy in Milan recommend universal testing for their population (D'Ubaldo et al. 1999), but they acknowledge that the uptake and the practice of offering the test are not always appropriate. The authors conclude that specific guidelines should be issued in order to implement and uniform universal HIV testing during pregnancy. Who should be the issuing body of the guidelines is not specified.

Luxembourg

No official guidelines were found for Luxembourg. However, some facts could be extracted from other published literature. Langer and colleagues report that in Luxembourg mainly obstetricians are responsible for the provision of antenatal care (Langer et al. 1999). Surprisingly, in a major study on the competences of midwives in European countries it was found that more than 50% of midwives in France and Luxembourg undertook ultrasound as
part of normal antenatal care procedures. Whether they do this under the responsibility of an obstetrician was not mentioned. In addition to that it was stated that in Luxembourg 7% of pregnancies were first diagnosed by midwives (Mead & Ashton 1996 + 1998: 42). In general terms, midwives are permitted to provide antenatal care for women experiencing a normal pregnancy (Mead & Ashton 1996 + 1998: 41).

The Netherlands

For The Netherlands, no official guideline was found. However, there is a major study available, which examined antenatal care practice between 1995 and 1996 (Wildschut et al. 1999). The authors collected data on frequency and content of standard care in a nationwide structured survey by mailed questionnaires among a sample of specialist obstetricians and midwives who were asked to report the standard tests routinely used for antenatal care in their own setting. With return rates of 80% and 71% for 132 and 394 accessed obstetricians and midwives, the study provides valuable insights into the actual situation.

The most common procedures reported were the assessment of maternal blood pressure, which was recorded at nearly all visits, and the maternal weight. This was assessed by 81% of the obstetricians and 94% of the midwives at each visit. Urinary tests for protein were also reported frequently. Screening for hepatitis B antigen was reported by 96% of the obstetricians and 100% of the midwives, and screening for lues in 95% and 99% respectively. Screening for the rubella titre was reported less frequently (66% of obstetricians, 59% of midwives). Vaginal examinations were performed infrequently.

In The Netherlands, neither obstetricians, nor midwives routinely administer an oral glucose tolerance test to screen for gestational diabetes. Midwives assess haemoglobin levels three times during pregnancy, obstetricians only twice. Interesting is that despite the large proportion of antenatal care provided by midwives, the duration of the gestation is frequently estimated by ultrasound. Surprisingly, also 33% of the midwives reported the use of ultrasound, as well as 69% of the obstetricians. An additional ultrasound scan to screen for anomalies at 18 to 20 weeks is offered by 30% of the obstetricians and 44% of the midwives. To a much lesser
extent, ultrasound is used for detecting foetal growth restriction. However, another study found that from 1996 to 1998 no routine ultrasound screening was performed (Stoll et al. 2001).

In The Netherlands, usually community midwives provide care for low risk women. Shared care is not common, as for women at risk obstetricians provide care. For both professions, a routine pattern for the distribution of visits is evident. In the first trimester, monthly visits are common practice. From 24 to 28 weeks on, every two to three weeks a visit is scheduled. From 36 weeks on, weekly assessments are planned. There were no different schedules for first or subsequent pregnancies. Interesting is that 75% of obstetricians and 94% of midwives did not consider it necessary to change the traditional frequency of antenatal visits.

In another study, the policy of antenatal screening for hepatitis B was assessed (Grosheide et al. 1995b). The authors reported a nationwide increase in test rates from 1989 to 1992 from 46% to 84%. In detail, they found test rates of more than 95% in rural areas as well as in large city hospitals. Most tests were performed at about 14 gestational weeks, but 10% of the screenings took place at delivery (Grosheide et al. 1995a).

Portugal

About antenatal care in Portugal no information but the following was found. In Portugal, midwives are not generally allowed to provide antenatal care (Mead & Ashton 1996 + 1998: 41). In addition, the WHO published that in 1995 91% of women who gave birth had four or more antenatal visits (World Health Organization 1997). Obviously, Portuguese societies or scientists tend not to publish in English, French or German journals.

Spain

In Spain, there are recommendations of SEGO, the Spanish Gynaecology and Obstetrics Society, on different quality criteria of the procedures used in antenatal care. Unfortunately, they are published in Spanish only, so that only the English abstract of the study mentioning these guidelines is cited (Goberna I Tricas et al. 1996 [Spanish]).

Antenatal care in Spain is provided to 99% by gynaecologists and to 32% by midwives. 22% of
pregnancies were cared for by GPs and 1% by “other specialists”. These were the results of an observational study in two counties of central Catalonia for which 171 women were interviewed after they had given birth in 1994. Other authors found that less than 50% of the midwives in Spain provided antenatal care (Mead & Ashton 1996 + 1998: 42).

Sweden

Although no official guidelines could be found for Sweden, there are hints that regulations exist (Åberg & Lindmark 1992, Blondel et al. 1985, Lindmark & Cnattingius 1991). However, these guidelines are not necessarily legally binding, as Åberg & Lindmark stated in 1992: “There is no mandatory national recommendation about the number of visits or the content of the care.” Their comment is that regional variations do exist with regard to the number of visits as well as to the use of tests. In each county, an obstetrician is appointed who is responsible for good and equal quality of antenatal care. This obstetrician is also required to set up guidelines on how to handle prenatal complications in the county. However, in 1992 the accepted routine programme of antenatal care followed the recommendations of a professional working group of the Swedish Association of Obstetricians and Gynaecologists from 1989 (Åberg & Lindmark 1992). These recommendations are shown in Tables 2.5a to c.
Table 2.5a: Antenatal screening tests recommended in Sweden - physical examinations

<table>
<thead>
<tr>
<th>Test</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>●</td>
</tr>
<tr>
<td>Fetal position</td>
<td>●</td>
</tr>
<tr>
<td>Full physical examination</td>
<td>●</td>
</tr>
<tr>
<td>Fundal height</td>
<td>●</td>
</tr>
<tr>
<td>Maternal weight</td>
<td>●</td>
</tr>
<tr>
<td>Vaginal examination</td>
<td>●</td>
</tr>
</tbody>
</table>

Table 2.5b: Antenatal screening tests recommended in Sweden - technical tests

<table>
<thead>
<tr>
<th>Test</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>●</td>
</tr>
<tr>
<td>Fetal heart rate</td>
<td>●</td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td>●</td>
</tr>
</tbody>
</table>

*optional*
Table 2.5c: Antenatal screening tests recommended in Sweden - laboratory tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Test</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-fetoproteine</td>
<td></td>
<td>6  8  10  12  14  16  18  20  22  24  26  28  30  32  34  36  38  40</td>
</tr>
<tr>
<td>Blood group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia (cervix)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coagulation disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal movement counting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-haemolytic streptococci</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis B surface antigen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Immunodeficiency Virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papanicolaou smear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placental functions (hormonal)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus antibody screening</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubella titer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triple test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

■ For rh-negative women
Prenatal care in Sweden is mainly provided by midwives who organise extra visits to a physician, if this is required (Lindmark & Cnattingius 1991). In contrast to that, another author claimed that one examination by a specialist obstetrician is mandatory in the last trimester (Kamper-Jørgensen et al. 1986 cited by Backe & Nakling 1993). Again differently described is the situation by another author (Westin 1980). He found that a healthy woman has more than 15 antenatal visits. Two visits to an obstetrician, and the rest to a qualified midwife who refers to an obstetrician or hospital, if necessary. However, although the author claims to describe the Swedish system of antenatal care, it does not become clear, whether he refers to entire Sweden, or to the hospital he works at. In 1992, other authors claimed that Swedish antenatal care is provided by midwives in close collaboration with hospitals. They report that the doctor is consulted on one occasion during normal pregnancy. Most of them are said to be obstetricians, but in remote areas also general practitioners participate in care (Åberg & Lindmark 1992). Finally it was reported that there is a traditional model of care which was recommended in 1981, consisting of 14 visits to a midwife and two visits to an obstetrician, and a revised model of antenatal care, recommended by an expert committee in 1989, consisting of 11 to 12 visits to a midwife and one to two visits to an obstetrician (Lindmark 1992).

As expected, also in Sweden theory is different from practice. With 13 visits, Swedish women do have the highest number of visits per pregnancy on average compared to the United Kingdom and all Nordic countries. Moreover, compliance to the routine program is extremely high according to a national survey in 1990 (Lindmark & Cnattingius 1991, Lindmark 1992). The authors found in a national survey that all women have had antenatal care. 90% have had 9 visits; the mean number of visits was 13.2.

In general, the history of Swedish antenatal care is better documented than that of most other European countries. Organised antenatal care was introduced in the late 1920s due to the idea that pregnant women living under deprived conditions would especially benefit from attention during pregnancy. At this time, the focus was put on counselling in hygiene and nutrition. In 1937, Parliament accepted a proposition by the National Medical Board, which recommended
a certain scheme for prenatal care. Initially, three visits were planned with the doctor: the booking visit, and at 24 and 36 gestational weeks. In addition, monthly visits were scheduled with the midwife until 24 weeks. From then on, biweekly visits were recommended until 36 weeks, thereafter, weekly visits were recommended (Lindmark 1992, citing an official source [Swedish] called SOU). According to this scheme, at every visit the maternal weight was measured, blood pressure and haemoglobin taken and a urine test for glucose and protein were performed. In the second half of the pregnancy also the foetal heart rate was assessed.

In 1955, a statute on maternity health care was issued that prescribed blood group testing and lues screening for all women. In 1969, a revised version was issued. From then on, cervical cytology and pulmonary x-ray (!) were recommended additionally. Moreover, urine tests for bacteria were recommended. The timing and frequency are not mentioned. In 1981, a new programme was approved by the Commission for Maternal and Child Health Care that clearly states additional aims of antenatal care, e.g. preparation for parenthood. Unfortunately, both statutes from 1955 and 1969 and the new programme from 1981 were published in Swedish only (cited by Lindmark 1992). In contrast to the previous recommendations, in 1981 no recommendations on the exact content and procedures of the visits were made. In contrast to that, other authors cite the recommendations of a professional working group of the Swedish Association of Obstetricians and Gynaecologists from 1989 (Åberg & Lindmark 1992), which were shown in Tables 2.5a to c.

**United Kingdom**

From the UK, it was found that antenatal care was first systematically introduced by the Ministry of Health in 1929 (Liu et al. 1992). The traditional schedule parallels the current one of several other countries, such as Finland, Germany, The Netherlands and parts of Sweden. Within this schedule, antenatal visits were planned monthly until 28 weeks, every two weeks until 36 weeks and every week until delivery (Audit Commission of the National Perinatal Epidemiology Unit 1998: 25, Jewell et al. 2000). Despite the existence of the traditional schedule, there is evidence that up to at least 1992 there seemed to be no singular official
guideline on antenatal care in operation (Liu et al. 1992). Instead, a system was prevalent, in which individual health care trusts decided on a care scheme (Audit Commission of the National Perinatal Epidemiology Unit 1998: 23). However, in 1998, the National Perinatal Epidemiology Unit reported that the government had recommended reviewing the provision of antenatal care (Audit Commission of the National Perinatal Epidemiology Unit 1998).

In October 2003, the envisaged guideline on routine antenatal care for the healthy pregnant woman was issued (National Collaborating Centre for Women's and Children's Health 2003). However, it is important to note that the guideline was commissioned by the National Health Service National Institute for Clinical Excellence on behalf of the Department of Health and the Welsh Assembly Government (NHS National Institute for Clinical Excellence 2002a + b). This means that although the guideline is a milestone in the re-organisation of antenatal care in the UK, it was set up for the NHS in England and Wales. For Scotland no guideline was found.

After the critical assessment of the new guideline it was concluded that it is based on the most extensive critical literature review on the subject to date. All in all, 631 pieces of original publications and meta-analyses were analysed by a multi-professional group of nine specialists and two consumer representatives. The reviewers analysed in a structured manner the literature published to date and clearly specified the methods used for identifying, selecting, analysing and integrating the evidence. Also economic data about cost-effectiveness were introduced, where available and appropriate. However, as not for all measures there is sufficient evidence to come to a unambiguous conclusion, recommendations were classified from grades A to D, depending on the robustness of the underlying evidence. In addition to these grades, also the label "good practice point" was introduced for recommendations based on the view of the guideline development group. It was therefore found that this report provides excellent transparency so that a skilled critical reader is able to follow the argument, access the underlying literature and draw his or her own conclusions.

However, the guideline is more comprehensive than needed for the thesis. In addition to the recommendations about which screening tests should be performed at what time during
pregnancy, also reference is made to the provision of information, antenatal education and lifestyle considerations. Also the issues of where and by whom care should be provided are addressed. Although these are necessary components of a comprehensive guideline, these aspects will not be covered by the thesis. As the guideline on routine antenatal care for the healthy pregnant woman will be discussed in greater detail in all following chapters of the thesis, a compilation of recommended tests is not provided here.

2.3.6 Randomised controlled trials comparing different schedules of care

In recent years, attempts have been made to evaluate new, mainly reduced, schedules of care by randomised controlled trials. Although ethical aspects have formerly made it nearly impossible to alter established and trusted care schemes, such trials recently came to be approved for several reasons. The new care schemes are developed by multidisciplinary expert panels, which also include consumer representatives. This makes changes acceptable to professionals as well as to the recipients of care. Secondly, the care packages do not alter the number of visits much, and most of the usual examinations are carried out also within the reduced care schemes. Moreover, all women of the study groups are offered the opportunity to arrange additional visits, if they feel a necessity for this, or if complications arise. Finally, the new care schemes are more and more backed up by scientific analyses of individual measures and their clinical effectiveness. Hence, it has become easier to conduct scientific trials with the exclusion of ineffective measures than some years ago, when these measures were still regarded as effective. The special traits of the trials will be presented in detail in the following, but their results are discussed and compared together at the end of this section.

To start with, one systematic review of randomised controlled trials/RCTs of routine antenatal care could be identified, which was performed on behalf of the World Health Organization (Carroli et al. 2001). The reason for conducting the systematic review was a perceived lack of strong evidence on the effectiveness of the content, frequency, and timing of visits in standard antenatal care programmes. The main hypothesis was that a model with a lower number of
antenatal visits would be as effective as the standard local antenatal care model in terms of clinical outcomes, perceived satisfaction, and costs. All RCTs were considered, which compared a model of a lower number of visits with the standard model in the respective setting. Finally seven randomised controlled trials were included, involving 57,418 women.

The conclusions to the systematic review were that there were no differences between the new and the standard models of care with regard to clinical outcomes (Carroli et al. 2001). This was statistically proven by a meta-analysis for all chosen outcome indicators, which were pre-eclampsia, urinary tract infection, postpartum anaemia, maternal mortality, low birthweight, and perinatal mortality. However, the authors state also that the sample sizes were not large enough to measure differences in maternal mortality accurately, as the event is rare.

Flaws in the systematic review

Although the statistical analysis is well founded, the systematic review itself has some flaws. First of all, the reason for conducting it was a perceived lack of strong evidence on the effectiveness of the content, frequency, and timing of visits in standard antenatal care programmes. Unfortunately, the study does not contribute information on most of these aspects. Despite the acknowledgement of whether care was "goal-oriented" or not, the details of the individual models, which were assessed, are not explained in greater detail. This applies at least to the published material. Individual measures, such as laboratory tests to screen for different conditions are not mentioned. Also no comment was made about the timing of visits. All in all, only the frequency, i.e. the absolute number of antenatal appointments, is mentioned. Therefore, the review is not able to answer the questions placed by the trigger of the study.

Another problem is that the compared interventions were the provision of a lower number of antenatal visits, and a standard programme for antenatal care. All RCTs were considered, which compared a model of a lower number of visits with the/a standard model. Despite this common trait, a comparison or combination of results is questionable, as the final number of visits in the reduced schedules varies between 4 and 12 in studies 2+7 and study 4. This
means that the number of visits in the reduced schedule of one RCT equals the standard model of care in another study. In one case, the number of visits in the reduced scheme even exceeds the number of scheduled visits for routine care in Europe. It is therefore not surprising that clinical outcomes are similarly good, as has been described in study 4.

It has to be noted also that the variation in absolute numbers of visits in the reduced care schemes themselves is sufficiently large, as to make a combination of results again difficult. In addition to that, the respective reduction in visits varies between approximately 18% and 37% in studies 1 and 5. This demonstrates that the true reduction rate in visits varies much greater than the commonly achieved reduction of 3 visits suggests. Despite these methodological problems, the authors combine the results of all studies without weighting the differences in settings, sample sizes, reduction in the number of visits, or the original models of care.

Other questions apply to the outcome indicators pre-eclampsia, urinary tract infection, postpartum anaemia, maternal mortality, low birthweight, and perinatal mortality. The authors claim that they had selected a priori outcome indicators, “for which antenatal care should have an effect”. No further comments were made on how the authors came to this judgement. It would have been beneficial if they justified their indicators against scientific evidence, or had other explanations for choosing them. Unfortunately, neither is the case. Moreover, it can be also asked whether these indicators make sense in every setting. Postpartum anaemia might not be a relevant indicator in wealthy settings with only two children per woman, but a very relevant indicator in poorer settings, or those with a higher birth rate.

When drawing all above mentioned facts together, it is concluded that a reduction in the number of antenatal visits seems to be possible without compromising clinical outcomes. It should be nevertheless kept in mind that the measures included in the care packages were not made explicit for either one of the models in the systematic review of the WHO. In addition to that, three of the underlying studies took place in countries not comparable to the EU. Therefore it was decided to analyse those underlying studies of the WHO systematic review in detail, which are likely to be relevant for antenatal care in the EU.
The underlying studies of the WHO systematic review

After analysing the studies included in the WHO systematic review in depth, three were chosen for closer analysis (McDuffie et al. 1996, Sikorski et al. 1996, Villar et al. 2001). With study number 7 one study was dropped, as no reduction of visits was achieved at all, but a goal-oriented approach was introduced in antenatal care. In addition to that, the study was conducted in a rural area of Zimbabwe, and one of the outcome indicators was the use of rural health centres for delivery. It is unlikely, that this study with a total number of antenatal appointments of four in the study- as well as in the control population contributes to solve the problems currently encountered in the European Union. The same reason led to the decision not to analyse study 2, which was also conducted in Zimbabwe. Another study was not used for closer analysis due to the high risk of bias. Randomisation problems made the results questionable (Binstock & Wolde-Tsadik 1995). A fourth study was not analysed in greater detail, as it took place in what the authors called a "free-standing birthing center in southern California" (Walker & Koniak-Griffin 1997). The findings of this study are included in the WHO systematic review, and do have their due place there. However, it was found that the study population within an alternative birth setting in the generally conservative medical system of the United States was highly selective. Therefore it was decided not to use this study in greater detail. The three remaining studies from the WHO systematic review (Carrol et al. 2001) are critically analysed and described in the following.

McDuffie and co-workers conducted from 1992 to 1994 a study in Colorado, including 2.764 women with diagnosed low-risk singleton pregnancies (McDuffie et al. 1996). They used a multi-disciplinary expert panel to develop a schedule with a reduced total number of antenatal visits. The schedule defined the exact content and the duration of the visits. After an initial risk assessment, the women were randomly allocated to the study group with 9 scheduled visits, and the control group, following the traditional care scheme with 14 visits. However, the tests performed seemed more or less identical in both groups.

In the new scheme, at the fist visit, a "routine laboratory blood analysis" was carried out,
presumably consisting of at least blood group, Rhesus factor and haemoglobin testing. In
addition to that, a Papanicolaou smear, a culture for gonorrhoea and testing for chlamydia
were performed. At 15 through 18 weeks of gestation, a test for maternal serum alpha-
fetoprotein was offered. Routinely, screening for gestational diabetes was performed by a 1-
hour glucose tolerance test. The haematocrit was assessed between 24 and 28 gestational
weeks, and antibody screening was performed at 28 weeks, if the mother was Rhesus
negative. At each return visit, blood pressure and weight were established, as well as the
foetal heart rate and the fundal height. Usually, a urine test for glucose and protein was also
performed (McDuffie et al. 1996). In this context it is interesting to remark that this nearly
parallels antenatal care for normal pregnancies in Germany.

In contrast to that, the group of Sikorski used care schemes with a more significant change in
the number of visits (Sikorski et al. 1996). Their control group followed the traditional British
schedule with 13 visits. The women of the study group with a first pregnancy attended 7 times,
women with subsequent pregnancies 6 times. The new distribution of visits for primigravidae
was as follows: booking, then visits at 24, 28, 32, 36, 38 and 40 gestational weeks. For
subsequent pregnancies, care commenced at 26 gestational weeks after the booking visit, and
the visit at 28 weeks was left out. Unfortunately, the content of care was not mentioned at all.

When analysing the findings of Sikorski and co-workers, it is definitely noteworthy that a
reduction in the number of visits does not compromise the clinical outcomes of pregnancy
(Sikorski et al. 1996). This finding was found to be reinforced by other studies. However, the
main conclusion of the authors themselves was that a reduced schedule of antenatal visits
may lead to reduced psychosocial effectiveness and dissatisfaction with the frequency of visits.
Unfortunately, this study provides insights into the effects of a new care scheme in a local area
in southeast London, rather than in an entire country or region. Nevertheless, with 2 794
women taking part in the study, at first sight the sample size appears to be large enough to
draw some valid conclusions. However, in a letter to the editor of the British Medical Journal, in
which the study was published, severe critique is uttered (O’Connell 1996). The writer of the
letter found that the study population was recruited from 3 inner city locations in London. 30% of the participants of each group were drawn from 8 or more ethnic minorities. Non-participant rate was 26.3%, and amongst those who agreed to take part the non-response rate to questionnaires on satisfaction ranged from 30% to 37%. The sample is therefore not representative for a wider population. In addition to that, the real differences in the findings were small, although statistically significant. As the authors of the study described themselves, a major problem in reducing the number of antenatal visits is the lack of acceptability to the pregnant women themselves. Irrespective of the clinical effectiveness, there seems to be a strong emotional binding to the traditional care scheme.

The third study, a large multi-centre randomised controlled trial, was conducted under the agenda of the World Health Organization (WHO) in Argentina, Cuba, Saudi Arabia and Thailand and included 24,678 women in 53 clinics (Villar et al. 2001). Most important is that this study put the focus on a new model of care which highlighted actions known to be effective in improving maternal or perinatal outcomes, and not only on a reduction in the number of visits. The new schedule of care was compared to the respective care schemes recommended by the governments, which were based mainly on the traditional western model of care with about 12 visits. In contrast to the studies mentioned before, this WHO-trial used outcome measures, which are not that common to be used to evaluate care in more industrialised countries. As typical and universal indicators, low birthweight and pre-eclampsia/eclampsia were introduced. In addition to those, severe postpartum anaemia and treated urinary tract infections were used, indicators that might be more suitable for countries with partially poorer living conditions, less access to antibiotics and food supplements, or different nutritional habits. This idea is reinforced when the authors mention that multiple dipsticks were given to the clinics where urine culture was not possible, and that free folic acid and iron supplements for all women at clinics were provided where these have previously not been available.

Most interesting in this study, however, is that an assessment for quality of care and an economic evaluation were carried out. In contrast to other authors, Villar and colleagues
(2001) did not randomly allocate women in one site to study or control, but randomised entire clinics. This must lead to several specific consequences with regard to the interpretation of findings, as confounding factors, such as living conditions, regionally different education and training levels, or the motivation of personnel might have had an impact on outcomes.

The results of the studies mentioned were as follows. McDuffie and colleagues found no significant differences between the study group with a reduced number of visits and the control group (McDuffie et al. 1996). However, the content and schedule of the new model of care was described exactly, but the deviations from the old model were not mentioned. The average total number of visits did not deviate much from the usual care scheme, with 10.3 +/- 2.8 visits compared to 12.9 +/- 2.8 visits. Again it has to be remarked that 10.3 visits on average are rather similar to the traditional schedule of care in most European countries. For example, 10 visits are recommended for normal pregnancies in Denmark and Germany, 12 visits are recommended in The Netherlands, Sweden, England, Scotland and Wales (Blondel et al. 1985). A WHO study group found also, that ten countries of the European region have legally binding recommendations of 10 to 12 antenatal visits for normal pregnancies (World Health Organization 1987: 19). However, these are officially recommended numbers, which do not necessarily reflect practice; hence they have to be interpreted with caution. Moreover, little is known about the actual content of care – information that might be more interesting than solely the number and distribution of visits.

As McDuffie and colleagues, Sikorski and co-workers found also no differences in the clinical effectiveness of their new model and the traditional model of care, although their study group had only 8.6 visits on average +/- 2.77 (McDuffie et al. 1996, Sikorski et al. 1996. No significant differences were found regarding the incidence and potential complications related to pregnancy induced hypertensive disorders. Also no significant differences were reported for maternal or perinatal morbidity. However, these authors reported poorer psychosocial outcomes.

Irrespective of the difficulties with the interpretation of findings as mentioned above, both
authors concluded that they maintained equivalent perinatal outcomes with significantly fewer visits (McDuffie et al. 1996, Sikorski et al. 1996). However, it can be debated whether a statistical power to detect a 2.5% increase in preterm delivery rates is sensitive enough, as the retrospective analysis revealed that the sample size had 80% power to detect a 2% absolute increase in the rates of preterm birth and low birthweight (McDuffie et al. 1996). Already in monetary terms, it might well be true that an increase in preterm delivery rates by 2% is much more expensive than 2.7 visits more for all pregnant women in a certain population, not to speak of other ‘costs’ involved, such as grief and other personal and societal ‘prices’ to pay. Unfortunately, with such sample sizes and settings, only rises in negative outcomes can be measured. A decrease in unwanted side-effects of antenatal care, or positive effects of the new care schemes, can not be detected. For measuring a decrease here, much larger samples would be required. As a consequence, studies on reducing numbers of antenatal visits are not planned to improve care in the sense of improving clinical outcomes, but to reduce numbers of visits and thus limit costs without compromising clinical outcomes.

In contrast to the trials conducted in the USA and Britain (McDuffie et al. 1996, Sikorski et al. 1996), the women in Cuba, Thailand, Saudi Arabia and Argentina had a median of five visits only (Villar et al. 2001). However, their control groups had only a median of eight visits, which is even less than the study groups in the trials conducted in the USA and Britain. In the WHO trial, the authors found in both care schemes similar rates of low birthweight, post partum anaemia and urinary tract infections. However, rates for pre-eclampsia and eclampsia were slightly higher in the study group with fewer visits. This trend persisted also after correction for confounding factors. The results for other clinical outcomes were again similar in both groups.

With regard to the costs involved, it has to be mentioned that calculations were available for Thailand and Cuba only. For these countries, the costs involved were calculated per pregnancy, including costs for neonatal specialist care, if required. In general terms, costs for the standard model of antenatal care were higher than for the new model of care. However, the detailed analysis of costs and their distribution goes beyond the scope of this thesis. As a
final thought it might be worth to consider that in the trials in the western countries, only the number of visits was reduced, but not the tests. Therefore, the reduction of costs might apply to the costs of health personnel, rather than to those for test material, laboratory equipment, etc. Relating this back to the member states of the EU, this makes huge differences in the potential to save money, as the costs for personnel vary greatly. The financial benefits will be much greater in countries where labour is expensive. The poorer countries, where labour is cheap, might not benefit as much as hoped for.

Consequences from the literature comparing different schedules of care

Taking all mentioned facts into consideration, it can be concluded that however promising results might be, the antenatal care scheme recommended by Villar and colleagues is not transferable to Europe (Villar et al. 2001). Pregnant women as well as health professionals would hardly be prepared not to have any ultrasound examination for example. By the similar good clinical results that could be achieved by each of the tested new models of care in their local context, but the somehow limited satisfaction of women with care, it becomes clear that antenatal care has to be tailored to the cultural and traditional needs of the populations it is planned to serve. These findings demonstrate that current regulations for antenatal care have to be carefully analysed, before a common European approach could be developed, which leads to similarly good, or even improved clinical and psycho-social results.

From the studies mentioned above, it becomes also evident that there is a worrying lack of documentation and communication between countries. An example of this is the fact, that researchers of one country assess a ‘new’ model of care, although this kind of care seems to be already standard practice in another country. One example for this is the care scheme used by McDuffie and colleagues, which is similar to standard practice in Germany (McDuffie et al. 1996). What makes things complicated is the fact that there is not enough description of the details of care for normal pregnancies, their general use and the timing of the respective interventions. This need for a detailed description and the need to avoid a multiplication of research efforts again highlight the need for an assessment of current guidelines used in the
countries of the European Union. Despite these needs for further investigation, the WHO developed a care model for the European region on the basis of its randomised controlled trial, which has been described above (Villar et al. 2001). This new model of antenatal care will be analysed in the following section.

2.3.7 The WHO new basic model of antenatal care

The world health organization issues a manual on essential antenatal, perinatal and postpartum care, which was especially designed to be used in the European region (World Health Organization (n.d.)). Although it was regarded as highly important to compare the findings of the study here to the recommendations of the WHO, this will not be possible. Despite major efforts and multiple enquiries through the secretaries of the European Institute of Health and Medical Sciences, it was impossible to obtain a copy of this manual. The minimum facts about the manual, which are published on the internet, are that it is available in English and Russian for governments, governmental agencies or other organisations. In addition, it is stated that there are still wide differences in care and outcomes between the countries in the European region, and that the tool of the WHO can help national governments to alleviate this. However, although it would have been most important to have the WHO manual in hands, the fact that it is targeted at health care providers at peripheral levels, which were specified as health centres, health posts and dispensaries, suggests that the focus of the WHO is different from this study. This study here uses national guidelines, approaching the problem from top down, rather than from the bottom-up perspective. Despite this limitation for future comparison, the search for the WHO manual was continued. Although it was finally not possible to obtain the manual directly, the Department of Reproductive Health and Research of the WHO published a handbook for the implementation of the new model on the internet (Department of Reproductive Health and Research (n.d.)). From this, the basic model for antenatal care in the European region could be extracted.

When assessing the basic model on essential antenatal, perinatal and postpartum care closer,
which was according to the WHO designed for the European region, it was a surprise to discover that it is the same model as the one that had been tested in the randomised controlled trial by Villar and colleagues on behalf of the WHO (Villar et al. 2001). As discussed in the previous section, the working group of the WHO concluded after the trial that the new basic model for antenatal care "provides the benefits of more complicated models while tending to save money" (Villar et al. 2001, World Health Organization 2003). However, after this randomised controlled trial, the WHO obviously went on to publish the tested model as a new basic model for antenatal care that should be introduced in common practice, also in the European region. As this thesis aims to explore the options for a common minimum guideline for antenatal care in the EU, it became necessary to assess the WHO model for its properties for the member states of the European Union, rather than for the entire European region.

Properties of the WHO new model of antenatal care

To start with, it is important to note that the working group of the WHO stated itself that many of the underlying studies assessing models for antenatal care have been carried out in different settings and countries, and that therefore generalising from the results may be difficult (World Health Organization 2003). This is also true for the study of the WHO working group itself, especially as the new model was tested in no single European state, but included Cuba, Thailand, Saudi Arabia and Argentina (Villar et al. 2001).

The new WHO basic model for antenatal care claims to be based only on measures which have been scientifically proven to be effective (World Health Organization 2003). As the model is strictly limited to these measures, the authors recommend that it should only be applied to women who have no evidence of pregnancy-related complications, medical conditions, or major health-related risk factors. When trying to use the WHO guideline for this study, it has to be noted that as a major difference the WHO model applies to screening, therapeutic interventions and the education of women. This study includes screening tests only. However, both guidelines are targeted at healthy pregnant women (World Health Organization 2002: 8).
As the model of the WHO strictly stays within the borders of what has been scientifically proven to be effective, only four antenatal visits are recommended. However, the authors also warn from the possible side-effects of a radical reduction in contact with the health care system during pregnancy. Perceived social support by means of antenatal appointments might be an important component of antenatal care in its own right. A reduction of contact might therefore have unwanted side-effects. In addition, also the huge gap between the initial visit at probably six or eight gestational weeks without any further contact up to the 26th gestational week, might lead to dissatisfaction and fears. Although this does not put the idea forward that the WHO basic model for antenatal care should be transferred without major adjustments to the EU, it demonstrates how far care can be reduced without compromising clinical outcomes.

A problematic aspect of the WHO basic model is that before a woman is eligible for the basic model of care, an assessment is made as to exclude all factors which would justify a more sophisticated care scheme (Department of Reproductive Health and Research (n.d.), World Health Organization 2003). For this initial grouping, a checklist is available. This list consists of 18 questions about the obstetric history, the current pregnancy and the medical history of the pregnant woman. The questions are kept simple and should be answered by yes or no only. According to the authors, the questionnaire aims at identifying those approximately 25% of women, who require more than basic care. Unfortunately, there is no hint in which sample these 25% of women are to be detected. It would be more than helpful to know whether this applies to populations in poor countries with their specific risks, or especially to women living under relatively affluent circumstances in the western world. It would be crucial to answer this question before transferring the WHO basic model to the member states of the EU. It would be necessary to test the checklist for initial grouping in western populations, as these are likely to carry different risks from the populations of Cuba, Thailand, Saudi Arabia and Argentina.

Although appealing in its nature, it is difficult to think of replacing the established models of antenatal care in an EU member state, by that of the WHO. In order to illustrate the scarcity of tests and appointments, the full model of basic care is demonstrated in Tables 2.6a to c.
Table 2.6a: Antenatal screening tests recommended by the WHO basic model - physical examinations

<table>
<thead>
<tr>
<th>Test</th>
<th>1st visit</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
<th>38</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full physical examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundal height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body Mass Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>only women with low weight at first visit</td>
<td></td>
</tr>
<tr>
<td>Vaginal examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6b: Antenatal screening tests recommended by the WHO basic model - technical tests

<table>
<thead>
<tr>
<th>Test</th>
<th>1st visit</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
<th>38</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrasound</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal heart rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.6c: Antenatal screening tests recommended by the WHO basic model - laboratory tests

<table>
<thead>
<tr>
<th>Test</th>
<th>1st visit</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-fetoprotein or Triple</td>
<td></td>
<td>6  8  10  12  14  16  18  20  22  24  26  28  30  32  34  36  38  40</td>
</tr>
<tr>
<td>Blood group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia (cervix)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coagulation disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal movement counting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-haemolytic streptococci</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis B surface antigen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Immunodeficiency Virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papanicolaou smear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placental functions (hormonal)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibody screening</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubella titer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis / Bacteriuria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis / Proteinuria</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>only 1st pregnancy or with previous history of pre-eclampsia</em></td>
</tr>
</tbody>
</table>
One of the first problems is the severe reduction in contacts with the health care system and the tests performed. It seems unlikely that a woman who had routine antenatal care in Germany in a first pregnancy (Bundesausschuss der Ärzte und Krankenkassen 2003) could accept the WHO model for subsequent ones. As an example, it might be difficult to explain why no ultrasound examination is performed at all, when she was told in her first pregnancy that at least three such examinations are necessary. The same applies to several other examinations. This problem is reinforced by the findings of the WHO systematic review of randomised controlled trials of routine antenatal care (Carroli et al. 2001). This study found that although clinical outcomes are not compromised by reduced numbers of visits and tests, some dissatisfaction with care occurred, particularly among women in more developed countries.

Another important aspect when thinking about the introduction of the WHO basic model is that clinicians from the European countries might feel disturbed when they have to change their practice back to more manual medicine. Many of them might not be used to perform and rely on clinical examinations to test for anaemia, rather than taking blood to check the haemoglobin level. The same is likely to occur when clinicians are requested to screen for symptomatic sexually transmitted diseases, instead of performing laboratory tests for lues, for example. Also the idea of immunising women in the developed countries during pregnancy against tetanus has a strange appeal. Moreover, the routine supplementation with iron and folic acid as recommended by the basic model of the WHO is often regarded critically by pregnant women themselves, and by a significant proportion of health professionals. At this point the latest it must be severely doubted that the basic model for antenatal care of the WHO was either developed for, or is transferable to the member states of the EU.

This doubt is further aggravated by the weighting of issues within the WHO basic model. Three components are postulated as essential. The first one is the screening for health and socio-economic conditions likely to increase the possibility of adverse outcomes. The second is the provision of therapeutic interventions known to be beneficial, and the third component is to educate the pregnant woman about planning for safe birth and how to deal with emergencies
during pregnancy. Although all these components are covered by routine antenatal care, e.g. in Germany (Bundesausschuss der Ärzte und Krankenkassen 2003), they do not have the same weighting as in the WHO model. In the traditional western models, screening for socio-economic conditions plays a minor role. Planning for safe birth and educating for emergencies plays also a minor role, as the broad and cheap availability of hospital care is self explaining and well established in most of the member states.

Drawing together all the facts discussed above, it can be concluded in line with the WHO itself that the recommendations for a basic care scheme are a goal for a minimum level of care for all women worldwide (Department of Reproductive Health and Research (n.d.)), rather than as a common guideline for antenatal care in the European Union. Although due to the exclusive testing in non-European countries doubts are still prevalent, the WHO model might be suitable for some of the less developed countries of the European region, with probably less firmly established schedules for antenatal care. However, it is concluded that the WHO model is not appropriate to be transferred to the member states of the EU. It seems to be more promising to build a minimum consensus out of the existing national guidelines, rather than defining a new basic model, which does not acknowledge the current situation in the member states of the EU. This notion is finally supported by the WHO, which despite the recommendation of the very basic model also for the European region, a reduction of the number of antenatal visits in existing care schemes is not recommend.

In contrast to recommend leaving the number of antenatal visits unaltered, the Antenatal Care Trial Research Group of the WHO (World Health Organization 2002) concluded also that in developed countries, each activity included in standard antenatal care should be scrutinized or tested for evidence of its effectiveness before being retained in the standard model. However, if this was systematically applied, this would lead directly to the basic WHO model. According to the authors, only measures with scientifically proven effectiveness are retained in the basic model for antenatal care. As a consequence, a common minimum guideline seems to be a suitable step forward, when it is based on a practicable compromise of scientifically proven
evidence of effectiveness, but also on current care schemes in the member states of the EU. The necessity of such a step is, despite all critique, documented well by the most important statements of the WHO on antenatal care (World Health Organization 2002). Firstly, the main issue is the need to evaluate the content of antenatal care and to determine it in a balanced manner so that the individual receives appropriate care. Secondly, it has to be considered that in some countries medico-legal pressures, or other influences, make it difficult to cut unneeded services. Finding a common basis for antenatal care in the EU requires action of policy-makers, rather than that of the individual providers, or individual member states. Only this could be a true step forward.

2.3.8 Conclusions to the review on antenatal care in the member states

One of the main objectives set out at the beginning of the thesis was to conduct a systematic review of the literature of the best available sources with regard to antenatal care within all member states of the European Union. Through the subsequent critical analysis of the current knowledge in the field, gaps and weaknesses in the literature were identified. This applies to knowledge about the content of official guidelines as well as to actual care practice.

Although some studies were identified that compared at least antenatal care practice from several countries, in the best case, eight of the former 15 member states were included. Hence, information about seventeen other members of today's union is missing. On the one hand, this is due to the time when the studies were conducted, on the other hand, either not all countries responded, or it was not reported why some countries were included and others not.

Moreover, it was found that nearly all of the 15 member states of before May 2004 have some kind of official guideline, but only for two states these guidelines were accessible. For all other countries, information was extracted as a by-product of studies occupied with antenatal care in the widest sense. This made the literature review extensive, but left the results unsatisfactory. What became clearly evident was that the obvious lack of knowledge and documentation has
detrimental effects on subsequent studies and care. Much effort could be saved, if the current situation was analysed properly and if the results were made easily accessible. Based on the findings of the systematic review it was concluded that there is a sufficiently large gap in the literature as to justify an investigation into the subject.

Analysing the currently available literature on randomised controlled trials comparing different schedules of care, and a systematic review of such trials, it was found that a reduction in the number of antenatal visits is clinically safe. There are no more detrimental clinical outcomes with regard to pre-eclampsia, urinary tract infection, postpartum anaemia, maternal mortality, and low birthweight. Unfortunately, some dissatisfaction among pregnant women was found with a reduced schedule of care. However, lower costs were achievable with such a schedule.

During the review on antenatal care it was also found that the World Health Organization established a new basic model for antenatal care, claiming that it would also be useful for the countries of the European region. As this model is suggested by one of the most influential organisations in the international health sector, this new basic model was analysed with regard to its properties for EU member states. Despite the credibility of the recommending organisation it was found that although the model is also recommended for the countries of the European region, it should not be transferred to the member states of the EU without major adoptions. However, by this analysis, the idea that a common minimum guideline would be beneficial also for the member states of the EU was reinforced.

2.4 Implications, advantages and disadvantages of guidelines

An operational definition of guidelines and their aims

A wealth of national as well as international publications from electronic as well as print media was identified, which are concerned with guidelines for clinical practice. In one of these publications (Harrison et al. 2002b), guidelines were defined as being
"essentially algorithmic formulations that guide their users to courses of (diagnostic or therapeutic) action, dependent upon stated prior conditions, though they do not necessarily claim to determine clinical action completely."

However, an easier definition (Field & Lohr cited by European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 41) is that they are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances."

Both definitions try to capture the nature of guidelines by highlighting the fact that guidelines have the aim to guide or recommend a certain course of action, rather than to prescribe it. However, for the study here, 'guidelines' needed to be operationally defined and set into the context of antenatal care in the EU. First of all, a precisely defined prior condition is required, to which the guideline relates. For this study, this is uncomplicated singleton pregnancies in healthy women. The required action is the baseline clinical care for them (National Collaborating Centre for Women’s and Children’s Health 2003). Baseline clinical care is additionally limited to diagnostic tests only. Preventive interventions, such as the administration of anti-D immunoglobulin for Rhesus negative women, are excluded. In addition, guidelines are limited to national guidelines. Clinical guidelines from individual trusts, hospitals, or other settings are not considered. For the thesis, the focus is put onto the entire population of pregnant women, thus conceptualising antenatal care as public health issue. Budgetary constraints and the need to allocate limited resources require decisions on the population level when setting up guidelines. However, the fact that care needs to be adapted to the individual pregnant woman is not neglected. The entire population refers to the sum of all pregnant women of the EU, or to the pregnant women of the individual member states, where indicated. On the basis of this operational definition, the review on the implications, advantages and disadvantages of guidelines commenced.

In addition to the definition of guidelines, a concise description of their aims was found. This
stated that it is the aim of guidelines to assure and to improve continually a high level of health care for the public, to avoid outdated and possibly harmful interventions, also leading to unnecessary costs. In addition, guidelines intend to stimulate evidence-based and cost-effective care, respecting individual patient’s needs and wishes. Another aim is to keep the public informed about the latest state of the art in the health field (Bundesarztekammer & Kassenärztliche Bundesvereinigung 1997). Although this is an excellent description of what guidelines can achieve, it has nevertheless to be doubted, that all guidelines can fulfil this ambitious aims. The continuous inclusion of newly acquired evidence can improve care. However, it is a major task to constantly review and include upcoming evidence.

Moreover, there might be also conflicts of interest, which need to be outweighed against each other. As an example, it might be that a new diagnostic test is effective for detecting a condition. However, if the condition is rare, cost-effectiveness will not be achieved. Yet another problem of the above mentioned aims might be that there are tests, which are statistically found not to be effective according to the protocol of a randomised controlled trial. Nevertheless, the same tests might be well known and perceived as reassuring by pregnant women, e.g. from previous pregnancies. Excluding such measures from a guideline might have negative effects outside the realm of clinical effectiveness. From this argument it becomes clear that the process of setting up guidelines is not a linear one. It always needs a person to weigh the evidence, and to finally decide. This will be explored in the section on evidence-based health care. Up to this discussion, a preliminary and less ambitious synthesis of the definitions and aims will be used. Until then, guidelines are defined as a straightforward way to integrate scientific evidence into practice.

Potential advantages and disadvantages of guidelines

When exploring the disadvantages of guidelines, a common critique is that the individual situation of a patient might not be considered, i.e. that the attempt to standardise care ignores the heterogeneity of patients and the complexity of medical decisions (Woolf cited by European Commission Directorate General for Employment, Industrial Relations and Social
Affairs 1999: 45). This argument might lead to the idea that guidelines should be regarded as a suitable instrument to decide about the care for entire populations, but not for the care of individuals. However, the same critics could argue that a guideline for a population, e.g. pregnant women in the United Kingdom, does not consider the special circumstances of a group of vegans living in a small town in Yorkshire. Guidelines have clear advantages on local, as well as on population level. Avoiding the use of guidelines would be a rushed measure, eradicating the advantages of guidelines together with their limiting factors. It might be better to acknowledge the requesting character of guidelines with regard to adopting the content to the individual situation of a pregnant woman, a state, or even a confederation of states, such as the EU. Guidelines enable individuals to tailor care to specific needs. In the sense of the word, they guide decisions, rather than prescribing a certain way of action. Guidelines should only be used in combination with the clinical expertise of the health care practitioner and the individual preferences of the recipient of care. Hence, using guidelines does not mean to follow a strict schedule for all patients with the same diagnosis, but to provide clear high-quality guidelines that can and should be adapted to the individual situation of patients as well as to the specific situation of countries, when a national guideline is concerned (Europarat 2001: 18, Sackett et al. 1997). Evidence-based guidelines therefore lay only the basis for individual care decisions.

However, another facet of the same point of critique is that the practitioner will be limited in his or her clinical judgement (Harrison et al. 2002b). In contrast to that, it could be also argued that a guideline of high quality, which is regularly updated, helps the practitioner to stay up to date in a challenging environment. The enthusiasm of health professionals to permanently update their knowledge should probably not be overestimated.

A clear advantage of guidelines could be that they are a way to demonstrate the scientific basis of the work of health professionals. By issuing well founded guidelines, a profession gains trustworthiness and enhances its status. However, this is likely to apply more to professions with a lower status (Berg 1997 cited by Harrison et al. 2002b). Groups with a higher professional status, e.g. medical doctors, were found to have a tendency to being bored
by clinical guidelines. They tend to have less positive attitudes on guidelines than e.g. nurses (Harrison et al. 2002b). However, the professionals with a higher status might have other reasons for interest in guidelines. For them, guidelines can be an instrument to delegate less popular or routine work to other professions, mainly some way down the hierarchical ladder. This was demonstrated for general practitioners and practice nurses in England (Harrison et al. 2002b). Transferred to antenatal care, the same mechanisms might apply. In countries with obstetricians as the lead professionals for antenatal care, midwives might be occupied with routine tasks, while the obstetrician has the competence to explain the results to the woman. This can end up with a midwife taking bloods and blood pressure as an auxiliary to the obstetrician, with the latter being the one in charge of care.

However, also professionals from groups with a higher status have negative attitudes towards guidelines. This group was found to fear that professional elites use guidelines as an instrument to diminish the professional independence and status of other members of their own profession (Harrison et al. 2002b). Yet another common suspicion is that the intent of guidelines is a managerial one which diminishes the independence and role of the health profession as such through bureaucratic rules (Harrison et al. 2002b).

In contrast to the negative argument from diverse standpoints, a positive view is that guidelines can be a tool to protect practitioners from medico-legal claims (Harrison et al. 2002b). Adhering to a guideline places a practitioner on the safe side, especially if the guideline was issued by a person or organisation of high profile. However, others again fear the opposite, with practitioners being assailable, if they do not follow the "arbitrary standards proclaimed by guidelines" (Hyams et al. 1996 cited by European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 46).

This debate offers valuable insights into the possible 'dark sides' of guidelines. They have the potential to being used as an instrument to exert power, forcing individual practitioners or other professions into directions, which they would have perceived as unfavourable otherwise. It might be that there are conflicts of interest between stakeholders. An individual pregnant
woman is likely to have other priorities as have the professions occupied with antenatal care. Also the interests of policy makers, the users of guidelines and the recipients of care are unlikely to be congruent in every aspect. This highlights the need to critically assess the quality of a guideline, together with a clear identification of the issuing body and its aims. A critical mind can not be overestimated. Even guidelines of superb quality, issued by well respected authorities should not be obeyed without critical thinking.

Besides the aspects already discussed, there are other factors limiting the positive potential of guidelines. One of these is the fact that they are only one instrument to enhance the quality of health care. Considering this, their relative expensiveness is clearly counterproductive for the broad use of guidelines. The estimated costs to set up a clinical guideline are 200,000 US Dollars in Germany (114,000 GBP), and about 900,000 US Dollars (513,000 GBP) in “other countries” (Merten 2006). Unfortunately, these countries were not specified, and the sources of the information are not quoted. Despite this, the reliability of the information is enhanced by the fact that the information about German guidelines stems from one of the members of the German office of the institution for quality assurance in medicine. These sums become plausible when considering that the development of a good guideline takes several months (Europarat 2001: 25). However, although important and interesting, this thesis will neither cover the monetary aspects of guidelines, nor will it investigate into the dissemination, implementation, or the uptake of them (Ollenschläger et al. (n.d.)). All these aspects are relevant after a guideline of high quality has been set up, but represent a field for research on their own. Trying to include them would have led too far from the focus of the study. Moreover, e.g. the monetary consequences of a guideline are assessed better by health economists. Other disciplines are suited better to assess whether the potential savings on the basis of a guideline outweighs its costs, in case the guideline is adhered to.

In addition to the monetary and political implications guidelines have, there are also ethical aspects to consider in the context of setting up and issuing guidelines on antenatal care. These are addressed in the following section, together with the ethical implications of a
potential common minimum guideline for the European Union.

2.4.1 Ethical implications of guidelines

Much of the advantages and disadvantages of guidelines has been already discussed in the previous section. Despite this, the main conclusions will be summarised and assessed for their ethical implications in the following. First and foremost it was stated that it is the aim of guidelines to assure and to improve continually a high level of health care for the public, to avoid outdated and possibly harmful interventions, also leading to unnecessary costs. In addition, guidelines intend to stimulate evidence-based and cost-effective care, respecting individual patient's needs and wishes (Bundesarztekammer & Kassenärztliche Bundesvereinigung 1997). However, it was also concluded that this is what guidelines can achieve in the best case, but that it can be doubted that all guidelines fulfil this. It is therefore of utmost importance to ensure the highest possible quality to ensure that a guideline is ethical.

A common point of critique was the idea that the individual situation of a patient might not be considered, i.e. that the attempt to standardise care potentially ignores the heterogeneity of patients and the complexity of medical decisions (Woolf cited by European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 45). Although this fear of guidelines making care formal and not responsive to the individual's needs, this is not necessarily the case. In the sense of the word, they guide decisions, rather than prescribing a certain way of action. Another facet of the same critique is that the practitioner will be limited in clinical judgement (Harrison et al. 2002b). However, the opposite can apply, if a guideline of high quality helps the practitioner to stay up to date in a challenging environment, as they readily present the evidences of the medical part of decisions. It can be concluded that guidelines recommend a pathway, on the basis of which choices should still made by clinicians and pregnant women themselves.

Despite their ability to put the focus on effective measures of antenatal care, guidelines might
also have the negative effect of focusing care on medical aspects. Moreover, when following a pre-defined scheme of care, routine and a false feeling of security can become evident. When this occurs, important personal and psycho-social issues can easily be overlooked, which means that valuable components of antenatal care are lost. In contrast to that, it can be argued that guidelines save resources by freeing up time of practitioners for other tasks, such as exploring the pregnant woman's preferences and her non-medical needs.

Another important issue is the fear that professional elites could use guidelines as an instrument to diminish the professional independence and status of other members of their own profession (Harrison et al. 2002b). If guidelines are issued by institutions other than the professional organisations themselves, such as by the Ministry of Health, or a group that has several members from funding bodies, there might be fears that the intent of guidelines is a managerial one which diminishes the independence and role of the health profession as such through bureaucratic rules (Harrison et al. 2002b). Guidelines therefore need not necessarily be ethical, as they have the potential to being used as an instrument to exert power, forcing individual practitioners, other professions, or members of the own profession into directions, which they would not have pursued otherwise.

Despite the above argument of ethical implications of guidelines of high or questionable quality and the possible unethical use of guidelines to reach other aims than ensuring the best possible care, the most important positive ethical effect of guidelines should not be overlooked. This is the fact that guidelines have the potential to limit or prevent unequal treatment due to gender, native origin or social status. When health care professionals are obliged by guidelines to provide a certain extent of care, deviations from routine practice have to be justified. This means that exceptions can be made when certain circumstances from sides of a pregnant woman suggest this, but not for reasons of withholding effective care from individuals or groups due to the idea that they are not worth a measure. Guidelines therefore help to ensure that a useful amount of care is provided on a broad basis.

From the above argument it can be concluded that guidelines should not be declared as
ethically questionable for principal reasons, nor does it say that whatever is recommended by
guidelines is ethically justified. Guidelines can be instruments for implementing ethical
standards in antenatal care, but this is not necessarily the case. Much depends on the content
of guidelines, which will be critically discussed in the following.

2.4.2 Ethical challenges when setting up guidelines on antenatal care

When dealing with antenatal care on the population level, the crucial question is how limited
resources can be used best. The WHO (World Health Organization 2003) summarised this
problem stating that the central ethical problem in antenatal care relates to access, questioning
whether it would be ethical to refuse the routine provision of interventions with proven benefit.
However, the WHO also questions whether it would be ethical to provide intensive antenatal
care to low-risk women, thus wasting scarce resources that could be used elsewhere.

As this statement poses more than one question, it is necessary to disentangle the underlying
aspects. To highlight the ethical questions arising when setting up guidelines on population or
national level, the key ethical principles of deontology and utilitarianism will provide a
framework for the discussion. To start with, the question will be addressed of whether it is
ethical not to provide an intervention routinely, when benefits for maternal and child health are
clear. Directly related to this is the question of whether it is ethical to provide intensive
antenatal care to low-risk women, wasting scarce resources that could be used elsewhere.

One factor limiting the positive potential of guidelines is that they are only one instrument to
enhance the quality of care, but that they are relatively expensive. The costs to set up a clinical
guideline were estimated as 200 000 US Dollars in Germany (114 000 GBP), and about 900
000 US Dollars (513 000 GBP) in "other countries" (Merten 2006). These sums become
plausible when considering that the development of a good guideline takes several months
(Europarat 2001: 25), and involves a multidisciplinary team of experts and lay representatives
(National Collaborating Centre for Women's and Children's Health 2003). As this means that
money is used for setting up guidelines, rather than for the care itself, it can be argued that this wastes valuable money for managerial activities.

However, it can also be argued that a guideline based on the assessment of the monetary consequences of different courses of action can help to save resources, and to allocate them most effectively. According to utilitarian ethics, this course of action is justified as it helps to bring the highest benefit for the greatest number of people (Montgomerie 2000, Wikipedia 2006b). However, this applies only as long as the costs for setting up a guideline do not outweigh its benefits. Health economists are able to assess whether the costs from introducing a screening test on a routine basis are outweighed by the reduction of treatment costs if the condition goes undetected. It is also possible to calculate the costs for setting up a guideline and to subsequently compare them to the savings from using it.

Conflicting interests

However, the monetary implications of guidelines are relatively easy to calculate. Much more problematic are conflicts of interest, or when non-monetary factors need to be weighed up against each other. Although theoretically possible, the scenario of a routine screening test and the subsequent treatment of the detected condition being cheaper than the treatment of a higher number of the same undiagnosed condition pictures an ideal situation. A less favourable option would be that a new diagnostic test is effective for detecting a condition, but that the condition is rare, and cost-effectiveness will not be achieved. In such cases the question arises of whether the test should be performed although the money could be used for effective tests that screen for more common conditions. But how to decide if the test has a slightly lower predictive value as the one, which screens for the rare condition? How to decide, if for one condition no treatment exists? And how to decide, if no treatment exists, but the termination of an affected pregnancy saves such amounts of money as to enable better screening for many others?

Yet another problem is that there are tests, which statistically are found not to be effective for
detecting a clinical condition. Nevertheless, the same tests might be well known and perceived as reassuring by pregnant women, such as the auscultation of the foetal heart rate (National Collaborating Centre for Women’s and Children’s Health 2003). Excluding such measures from routine action might have negative effects other than clinical effectiveness. This would mean that the test is not effective in the sense of detecting a condition, but in providing reassurance to the women. Should such a test be performed? What, if the test is cheap? What, if the test has the potential to frighten some women by false positive results? The International Confederation of Midwives international code of ethics for midwives (International Confederation of Midwives 1999) suggests that midwives and women should work with policy and funding agencies to define women’s needs for health services. This will ensure that resources are allocated considering priorities, therefore ensuring justice and fairness.

The last question arising from the above statement from the WHO is whether it is ethical to provide intensive antenatal care to low-risk pregnant women, which perhaps causes harm to them (World Health Organization 2003). This means that in addition to the process of optimising effectiveness and efficiency of actions, also risks arising from procedures need to be reduced as far as possible (Seedhouse 1998:209). However, this situation is further complicated by the fact that positive as well as negative effects are likely to be weighted differently by different persons, populations and institutions. Such questions are usually addressed by philosophers, rather than by health professionals.

The position of deontologists

The proponents of deontologist ethics state that praiseworthy goals can never justify immoral actions. In brief, one of the most important statements of deontology is that ends do not justify the means. According to deontologists, decisions should be made solely or primarily by considering one’s duties and the rights of others. It claims that a priori moral obligations exist that do not change merely as a result of a change in circumstances (Wikipedia 2006a). In further explorations of deontological ethics, the distinction between the right and the good was explained. It was stated that whereas utilitarianism argues or assumes that an act is right and
should be carried out, if it maximizes the good, deontological theories assert that an act can maximize the good, but still be wrong. An act should therefore not be carried out if it violates some deontological principle, such as a right or a duty (Rawls cited by Wikipedia 2006a). Deontologists therefore believe that actions can be inherently right or wrong, irrespective of their consequences (Montgomerie 2000).

Applying the principles of deontology to antenatal care would lead to a major reduction in the number of tests performed. Tests with the potential to lead to false positive results are likely to be questionable according to deontological ethics. They have the potential to be beneficial for some pregnant women, but to hurt others. The same applies to tests, which carry a risk of following complications, such as miscarriage after amniocentesis. However, questions remain as long as the relation between the mother and the foetus is not defined. If the foetus is defined as a person in his or her own right, it would be immoral to terminate a pregnancy, when i.e. a genetic disorder was diagnosed. If the foetus is not defined as a person with own human rights, it might be possible that deontologists accept the termination of pregnancy. The law would then be that the mother's rights prevail as long as the foetus is in utero.

Utilitarian philosophers' views

In contrast to deontologists' views, utilitarian philosophers claim that a measure is justified, if more people benefit from an action than are harmed by it. Utilitarianism is a theory of ethics that prescribes the quantitative maximization of good consequences for a population. Although happiness is usually named as the classical wanted outcome of utilitarian ethics, other consequentialists argued that consequences such as justice or equality should also be valued, regardless if they increase happiness or not. According to utilitarianism, ends can justify the means because decisions are judged primarily in terms of their consequences (Wikipedia 2006a + 2006b). In its final consequence, this theory provosts that harm to one individual may be sanctioned if it is for the benefit of a larger group (Coldicott et al. 2003). A concise definition is that utilitarianism is the doctrine that the morally right thing to do is whatever produces the greatest good for the greatest number...” (Montgomerie 2000).
When using utilitarian ethics to re-think antenatal care, negative utilitarianism might also be applicable. In contrast to utilitarian theories that deal with producing the greatest amount of good for the greatest number, negative utilitarianism requires to prevent the greatest amount of harm for the greatest number of people. This was proposed according to the idea that the greatest harms are more consequential than the greatest goods (Wikipedia 2006b).

However, there are several issues that need discussion, before using utilitarianism to make decisions on antenatal care. First of all, this is the statement that utilitarianists judge all actions by their ability to maximise good consequences, and that this always justifies harm to a single individual, if there is a greater gain to other individuals. However, it is clearly stated that this is only true if either the number of those who benefit is sufficiently large to outweigh the loss for the individual, or the gain for a few is larger than the loss for the individual (Wikipedia 2006b). When applying this to antenatal care, the problem becomes evident. How to compare the grievance from the termination of a pregnancy on the basis of a false positive test result against the relief of the termination of a pregnancy with a truly affected foetus? Is the grievance outweighing one relieved mother, or two, or are loss and gain balanced, when one hundred were diagnosed correctly? This example demonstrates well the difficulty of calculating happiness or harm. Even more difficult is the situation, when happiness, or grievance, need to be compared to monetary factors, which then can be subsequently used to do good things. Would it be justified to force a woman to terminate an unhealthy pregnancy, to enable better care for several women with lower risk? Although it is of vital importance that each society finds its own standpoint within such discussions, it has to be kept in mind that other laws and principles should not violated, such as human rights, which are not negotiable.

Summary of the implications, advantages and disadvantages of guidelines

In the above discussion, extreme examples were used, which lay mostly in the realm of prenatal diagnosis. This was done to illustrate the ethical and moral implications of any tests performed during the antenatal period more clearly. However, also the ethical theories presented can have extreme implications, if used consequently. The most radical version of
utilitarianism could, for example, mean the complete cessation of antenatal care, if money would do more good if used for other things, such as free housing for everyone. If deontological ethics were applied consequently, antenatal care would be also reduced to an absolute minimum of measures, which are 100% safe. However, antenatal care for healthy pregnant women is different from such extremes.

In current practice, guidelines on antenatal care seem to represent something that might be called a system immanent utilitarian approach with a moral component. Within the system, as money is set aside specifically for antenatal care. Within this budget, the money is used for those measures that produce the most good. The moral component, however, ensures that ends do not justify means in any case. Care is taken not to harm individuals for the benefit of others. However, what also became clear from the above examples is that values are difficult to assess, but are most influential in antenatal screening.

In addition to the above mentioned 'extrinsic' problems of guidelines, which have to do with questions of whether guidelines as such are a good thing, they might have also problems and limitations in themselves. However, as the evaluation of guidelines is not yet familiar to the average practitioner, it is of utmost importance that guidelines fulfil certain standards, and are labelled according to their properties. Those standards, as well as the currently available instruments for the appraisal of guidelines are explored in the corresponding section on material and methodology. This theoretical discourse on the quality of guidelines will contribute to shed light onto the 'intrinsic' problems of guidelines that need consideration.
3.1 Introduction

After the initial review of the literature it became clear that the objective of Phase 1 of the study, i.e. the critical appraisal of the relation between evidence-based guidelines and guidelines based on expert opinion, can not be reached by a simple review of the literature. In order to achieve the objective, the paradigms underlying and the methods for setting up guidelines needed to be theoretically discerned and analysed.

From the literature it was also found that the quality of guidelines is of overriding importance, but that the concept of guideline quality is difficult to grasp. It was therefore decided that a theoretical discourse on the quality of guidelines will contribute to shed light onto the 'intrinsic' problems of guidelines, and also the nature of evidence will be addressed. To achieve this, a critical analysis of the literature will be conducted with the intent of a concept analysis on the theoretical basis and the quality of guidelines, treating the literature as secondary data. This analysis is demonstrated in Part 1 of the following chapter.

Part 2 of the chapter is focused on the current health policy of the European Union and the strategies of decision-making on European level, as these two themes lay the basis for the research. Based on the concept analyses from Parts 1 and 2 of this chapter, the research plan will be further specified and presented in Part 3 of this chapter.

Finally, the study uses a mixed methods approach, consisting of a survey, an extensive critical review of the literature and of the state of the art, theoretically discerning and analysing the paradigms underlying and the methods for setting up evidence-based guidelines and guidelines based on expert opinion, and a critical in-depth appraisal of two national guidelines. Although an established instrument is used for the appraisal of the guidelines, the method for this part of the study was amended. Despite using simply the recommended number of four appraisers with unspecified qualifications, a panel of experts was selected purposefully in
order to elicit informed judgements about the quality of the two national guidelines on antenatal care from England and Wales and Germany. Special care was taken to relate the interpretations of the experts back to the original guidelines, bringing up their individual professional views. This approach was chosen as the theoretical analysis of the state of the art demonstrated the relative nature of evidence and the need for expert judgement to make good policy decisions.

3.2 Selection and use of material for the concept analysis

It is the aim of this part of the review to describe and analyse the conceptual underpinnings of the research problem. This section is therefore less practice, but more theory-oriented. Concepts are operationally defined as abstractions of observable phenomena (Polit & Hungler 1999: 125). Only by defining and operationalising the underlying concepts, the findings of the study will become meaningful and generalisable, and enable an evaluation of what can be learned from the collective application of the findings. Conceptual clarity will be achieved by a comprehensive quantitative overview of the concepts, their theoretical discussion and the synthesis of the theoretical framework at the end. The relationship between and the trends within the concepts are to be analysed and made explicit. However, it has to be kept in mind that although the findings should be generalisable to some degree, the concepts are analysed in the direct context of the study, which is antenatal care in the EU.

In contrast to problems with a clearly formulated research question, a concept analysis is not approachable by means of a systematic review of the literature in the classical sense. For clinical questions, there are preferred corresponding research designs, such as randomised controlled trials or cohort studies with accepted criteria for assessing their respective validity. With them, informative empirical answers should be given to scientific research questions (NHS Centre for Reviews and Dissemination 2001). However, when addressing complex phenomena, there are less direct links to study designs and their validity (Greenhalgh 2003). In
addition to that, a review of theories and concepts means to explore intellectual ideas of different validity. This was thought to be best assessed by a broad exploration of the concepts related to the study. Therefore, it was decided to conduct a comprehensive critical descriptive review of the state of the art with open questions. Evidence from scientific studies is critically appraised and subsequently synthesised with evidence from other sources, such as publications on methodology, theoretical discourses, conceptual literature, textbooks, newspapers and consultations with experts. Citation tracks are followed up, and a comprehensive picture will evolve of how the concepts and ideas from different authors relate to each other. This methodology was recommended as being acceptable to approach the conceptual literature by a consulted expert for health policy (Harrison 2005).

In order to get an impression of the wider context of the research problem, the review of the state of the art commenced with a broad reading on the subjects identified during the systematic review on antenatal care in the EU. The evolving ideas were followed up and constantly refined through a more and more focused search on the key concepts and theories. This means that again narrative analysis was the method of choice, with the presentation of findings under a series of subheadings and a commentary relating to them (Forbes & Griffiths 2002). Care was taken to consider not only material which reinforced an idea, but to search for contradictions and gaps in the literature, as well as for publications challenging an idea. It was considered to be of utmost importance to keep this critical stance throughout.

With regard to the depth of the review, it was aimed at providing a comprehensive thematic analysis of the key concepts and theories that are important to the study. All relevant concepts should be identified, together with the arguments in favour and against them. The aim was therefore to be comprehensive with regard to the concepts and the issues around them, but not with regard to identifying all publications concerned with them. All aspects that develop a concept further should be included in the final review. However, the focus was put on the saturation with regard to having explored the entire properties of a concept, rather than having read all publications on it. Searching and reading commenced until no new aspects relevant to
the study were found, or when the point was reached at which authors seemed to cite each other. These searches were conducted throughout the entire process of the thesis.

The overall thematic analysis of the key concepts and theories should acknowledge their core facets, the specific strengths and weaknesses in general, as well as in relation to the focus of the study. The outcome of this section takes the form of a theoretical discussion, trying to depict the relationship between, the trends within and the discussions around the identified concepts. Consistencies as well as contradictions should become clear. Possible explanations for the contradictions, such as different conceptualisations are made explicit. At the end of this, the theoretical stance of the study within the current debates should become evident.

An important factor in the review of concepts is the rigour of the method, which is applied. This is demonstrated by providing a detailed track record of the searches. Special care is taken to describe the data sources and the methods used for the identification of publications alongside the concept analysis. From this it should become evident, how the themes for the theoretical discussion emerged during the review process, and how they are used.

However, the application of uniform and rigorous standards of appraisal is also essential to synthesise the available evidence critically and concisely. For this, a pre-defined set of questions was used to guide the review. The questions guiding the critical exploration of the concepts were derived from deductive reasoning, publications of previous research and the conceptual literature. The questions applied to each idea were the following:

- Is the described aspect of relevance to the study?
- Is the (methodological) quality of the material such as to justify its inclusion?
- Is the argument plausible (face validity)?
- Are there other explanations?
- Is it an opinion or a theory?
- Is there more than one author / group of authors reporting the phenomenon?
- Is there literature from other disciplines reinforcing the ideas?
- What is the relation to other literature (does it fit into the picture, or is it a different view)?
- What are the strengths and weaknesses of this concept?
- How do the ideas relate to antenatal care?
- How do the ideas relate to the European Union?
- What are the implications of these ideas on EU-level?
- Is there a need to continue reading?

Themes evolving from the reviews on antenatal care and the implications of guidelines

A variety of themes and issues arose from the review process on antenatal care, which was described in the literature review. During this review, the initial concepts of interest to the study were identified. The first and foremost aspect that needed thorough consideration was guidelines as such, which was partially presented in the literature review. However, during these initial reviews, themes emerged, which needed to be followed up. These were:

- the need to discriminate between the concepts ‘guideline’ and ‘recommendations’,
- the appraisal of guideline quality,
- the question of what ‘evidence’ really is and
- how evidence of different sources and qualities should be handled.

Especially during the reading process on the last two aspects, it was found that the concepts - evidence-based medicine / health care / policy and
- the diverse aspects of making and assessing health policy

play an important role for the thesis and need further exploration. These will be presented in the following section.
3.3 Guidelines

3.3.1 The quality of guidelines

On the basis of the literature review on the implications, advantages and disadvantages of guidelines it was concluded that a search for existing national guidelines on antenatal care does not suffice for the purposes of the study. For the clarification of concepts, the search commenced with a broad reading on the subject, as has been described in the section of selection and use of the literature. Starting from the aforementioned textbooks, the search to determine the current state of the art developed into the national as well as international websites on the development and appraisal of guidelines for clinical practice. They provide search engines and links to each other, leading step by step into the depth of concepts, as well as to the original papers. As the references for the original literature are cited alongside the evidence used in the review, only the organisations and their entrance sites are listed, which were used as gateways to the topic (Appendix 1, References 1, 3, 8, 13, 14, 16, 17 + 20).

By critically assessing the publications from the above internet sources, as well as the original publications cited by the organisations active in the field of quality assurance in health care, the concepts of relevance for the theoretical framework were identified. A theme that was discussed fiercely is the quality of guidelines, and how this can be assessed and assured. It was found that over the past decade, much has been published about the quality and the appraisal of clinical guidelines. In most of the member states of the European Union, institutions were constituted that developed a national approach to assure the quality of guidelines. One example for this is the German "Ärztliches Zentrum für Qualität in der Medizin", in brief ÄZQ, which was planned to be a competence centre for guidelines and patient information in medicine (Rabbata 2003). This centre aims at supporting science-based practical guideline programmes (ÄZQ (n.d.) a). As one of its central achievements, the ÄZQ has set up a formalised process for the critical assessment of guidelines in order to enhance
their transparency, practicality, scientific quality and cost-effectiveness (Ollenschläger 2003, Ollenschläger et al. 1998).

Despite major advantages of formalising the process of assessing the quality of guidelines, also fears have to be acknowledged that governmental control or the interference of health insurances might have too much influence on the treatment of individual patients. There are also sceptical voices, which fear that financial aspects might supersede the interests of patients in general (Richter-Reichhelm & Encke cited by Rabbata 2003). When bias is to be minimised, the assessment of guideline quality can not be solely left to organisations, such as the ÄZQ. At the very least, each practitioner should be able to assess the quality of guidelines using one of the currently available instruments for this task. However, from the initial reading it was found that the instruments to assess the quality of guidelines are addressing the structural quality of guidelines, but not the quality of the recommendations they make. For this, the clarification of the concepts 'guideline', 'recommendation' and the underlying evidence is necessary before the instruments to assess the quality of guidelines are analysed.

3.3.2 Evaluation of recommendations

When appraising the quality of guidelines, it is of utmost importance to distinguish clearly between the quality of a guideline, and the quality of its individual recommendations.

'Guideline' refers in this study to the entity of diagnostic tests recommended for baseline antenatal care of healthy women with uncomplicated singleton pregnancies.

'Recommendation' is used for a final statement about whether a single diagnostic test should be included in the guideline or not.

In the following, schemes for the evaluation and grading of recommendations are analysed. A popular example for a system to classify the quality of recommendations is provided by the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001). Basically, this scheme
consists of two steps. In the first, the studies leading to a recommendation are assessed for their level of evidence as defined in the 'Oxford-Scheme'. This means that the studies are grouped according to their scientific impact following from design and quality of the study in relation to the study type. The ranking gives marks from 1 for high-quality systematic reviews of randomised controlled trials to a mark of 5 for publications of expert opinion without explicit critical appraisal. In a next step, the recommendation itself receives a label from A to D according to the levels of evidence of the studies on which it is based. In this scheme, Class A recommendations are based on Level 1 studies only. Class D is reserved for recommendations based on weak, or no scientific evidence.

As this is a common way of grading recommendations, other examples with similar schemes can be named. For example, the Royal College of Obstetricians and Gynaecologists [RCOG] has set up its own grading system for recommendations, which uses a simpler form than that of the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001, Royal College of Obstetricians and Gynaecologists (n.d.)). Within this grading system, recommendations are labelled only from A to C. According to this system, Grade A recommendations are based on randomised controlled trials only. Grade B indicates that a recommendation is based on robust experimental or observational studies. Finally, recommendations which are “based on more limited evidence but the advice relies on expert opinion and has the endorsement of respected authorities” are graded as C (Royal College of Obstetricians and Gynaecologists (n.d.)). As another example, also a task force of the German Society of Gynaecology and Obstetrics [DGGG] uses a similar scheme, which for example was used to draw up a guideline on the treatment of breast cancer, for example (von Minckwitz et al. 2002).

In contrast to the other instruments demonstrated above, the German task force was more explicit with regard to what consequences should be drawn from the grade of a recommendation. Although it might not be appropriate in every setting to give such explicit advice, this seems to be perceived as helpful in Germany. According to the task force, Grade A recommendations are based on strong evidence, and the intervention or treatment should
therefore be used in a specific situation on a routine basis. Grade B means that there is moderate evidence for the routine use of an intervention or treatment. Grade C labels insufficient evidence, which means that it is not clear whether practitioners should recommend or dismiss the intervention/treatment. Grades D and E are used for measures that should be considered as harmful or useless according to either moderate or strong scientific evidence.

Unfortunately, such easily applicable and straightforward grading systems bear the danger of doing more harm than good. A common way of drawing the wrong conclusions is to overestimate recommendations that are based on randomised controlled trials only. Important information is likely to be left out when no additional evidence than that of randomised controlled trials is considered. In addition to that, such Grade A recommendations bear the danger of being judged as untouchable gold standard, which can lead to forget about integrating new evidence as it is coming up. Such recommendations might then be adhered to for too long without any reassessment. Another major limitation is that findings from qualitative studies are generally never labelled as Level 1, although many aspects of antenatal care can be assessed by such designs only. As an example, the attitudes of women with regard to certain diagnostic tests are likely to be captured best by studies with a qualitative design.

As another disadvantage, such grading systems are not directly applicable to diagnostic tests and screening procedures (Royal College of Obstetricians and Gynaecologists (n.d.)). For these, experimental designs are not appropriate for proving their effectiveness. Instead, sensitivity, reliability, positive- as well as negative predictive values and false positive and false negative rates are appropriate criteria for screening tests. When setting up screening programmes, other important factors have to be considered, such as ethical implications of the test and of the consequences drawn from the test result. Moreover, also the costs, the risks for adverse outcomes and the acceptance in the public play a role.

Taking all of the above mentioned into consideration, it is concluded that it is necessary to determine the scientific basis of the recommendations of a guideline. Due to the central role of the NICE guideline on routine antenatal care for this study (National Collaborating Centre for
Table 2.7: Levels of evidence according to NICE
(National Collaborating Centre for Women’s and Children’s Health 2003)

<table>
<thead>
<tr>
<th>Level</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic review and meta-analysis of randomised controlled trials</td>
</tr>
<tr>
<td>1b</td>
<td>At least one randomised controlled trial</td>
</tr>
<tr>
<td>2a</td>
<td>At least one well-designed controlled study without randomisation</td>
</tr>
<tr>
<td>2b</td>
<td>At least one other type of well-designed quasi-experimental study</td>
</tr>
<tr>
<td>3</td>
<td>Well-designed non-experimental descriptive studies, such as comparative studies, correlation studies or case studies</td>
</tr>
<tr>
<td>4</td>
<td>Expert committee reports or opinions and/or clinical experience of respected authorities</td>
</tr>
</tbody>
</table>

Table 2.8: Grading system for recommendations according to NICE
(National Collaborating Centre for Women’s and Children’s Health 2003)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Directly based on Level 1 evidence</td>
</tr>
<tr>
<td>B</td>
<td>Directly based on Level 2 evidence or extrapolated recommendation from Level 1 evidence</td>
</tr>
<tr>
<td>C</td>
<td>Directly based on Level 3 evidence or extrapolated recommendation from either Level 1 or 2 evidence</td>
</tr>
<tr>
<td>D</td>
<td>Directly based on Level 4 evidence or extrapolated recommendation from either Level 1, 2 or 3 evidence</td>
</tr>
</tbody>
</table>

However, it is also acknowledged that although such hierarchies and grading systems for evidence and recommendations appear straightforward and easy to use, there are important issues behind them. When addressing the quality of guidelines and the quality of their
recommendations, it is vital to examine the underlying evidence more critical than by just assigning a certain level according to the method, with which the evidence was generated. As this aspect plays a key role in the thesis, this will be critically explored in depth in the following.

Another finding from the analysis of grading systems to classify the robustness of recommendations was that evidence does not translate itself into guidelines. Each system defines the grade of a recommendation according to the level of evidence. The crucial factor for setting up guidelines is therefore a good strategy to deal with and weight the evidence. To clarify the underlying concepts, an in-depth analysis of the following was necessary:

Defining and critically discussing the nature and the hierarchy of evidence, exploring the different kinds of evidence, such as expertise and scientific evidence. This includes the definition of the advantages and limitations of the role of experts, the nature and role of scientifically generated evidence, as well as of what might be categorised as context-free and context-sensitive research evidence (Canadian Health Services Research Foundation 2005, 2006a + 2006b). However, before analysing the above concepts, the second quality aspect of guidelines needs consideration: the methodological quality and the instruments to assess it.

3.3.3 Instruments for the systematic appraisal of guidelines

When exploring the current state of the art for appraising the quality of clinical guidelines, it was found that efforts are made to develop instruments for the formal appraisal of guideline quality. These instruments address e.g. the methodological, the structural and the process quality of a guideline. For this part of the review, it was aimed at identifying different instruments to assess guideline quality, using them to explore potential strengths and weaknesses. To explore the concept completely, the early national stages of guideline appraisal are also demonstrated, using Germany as example. However, the main aim was to find an instrument, which can be used on guidelines from different countries. In addition to the entrance links mentioned above, links available under the heading 'evidence based medicine'
were searched (Appendix 1, Reference 9). In a next step, the specific strengths and weaknesses of the instruments are critically analysed.

In 1997, the German Medical Association and the National Association of Statutory Health Insurance Physicians published criteria for the evaluation of guidelines (Bundesarztekammer & Kassenärztliche Bundesvereinigung 1997). The criteria of this catalogue specify what information should be contained in guidelines. For example, which methods and materials were used to set up the guideline, how the material was analysed and which professions or groups were represented in the team that has developed the guideline. Unfortunately, within this catalogue there is no hint as to what standards have to be fulfilled to make up a good guideline. In order to standardise the conclusions drawn from the evaluation of a guideline, it seems to be necessary to specify not only which information should be contained, but also how this information has to be evaluated. For this, it is concluded that an important step is missing for a final conclusion about the quality of a guideline.

Another problem of the instrument is its format. In contrast to later instruments, this one has no inviting and structured format to extract and record the relevant information. The criteria are listed under subheadings, but there is no form on which the answers can be collated. As a consequence, this early catalogue of criteria might make more sense for those intending to set up a new guideline, rather than for those trying to evaluate an existing one.

Acknowledging the problems of the aforementioned instrument, another working group developed it further (Ollenschläger et al. 1998). The new instrument for a standardised evaluation of guideline quality came in the form of a checklist. It was not only based on the criteria of the previous instrument, but also included the ideas of seminal papers available at that time, e.g. from the Scottish Intercollegiate Guidelines Network. When using the new instrument, there is a more detailed analysis of the rigour of the development of a guideline, as well as a clear focus on the comments made by the issuing body concerning its applicability. However, although this instrument is structured more clearly, it still is not user friendly with 41 questions to be answered. In addition, it still does not invite the user to make a final decision
about the overall quality of a guideline. However, the major problem of this instrument is that it seems to have never been piloted in this version. All in all it might have been a good idea that the authors labelled their instrument with an expiry date, which was January 1999.

In contrast to these earlier attempts, the next instrument came in the form of a checklist with boxes to tick (ÄZQ 1999). This instrument was again a further development of the aforementioned one (Ollenschläger et al. 1998). However, also the work of other experts in the field, e.g. from the USA, Scotland and England was considered. As its predecessor, the instrument consists of 41 questions in three categories: 'Process quality', 'content and format' and 'applicability of the guideline'. However, also this instrument contains no summary of the quality of the assessed guideline. In this case, the user ends up with ticked boxes, rather than with an unambiguous comment or mark. Despite these limitations, the developers of the instrument named an expiry date for it, which is 11 December 2003. As with the dates for the previous instrument (Ollenschläger et al. 1998) it does not become clear whether these dates were chosen arbitrarily, or whether there is a certain system behind it. Moreover, there is also no comment to what should be done with the guideline after that date. It is just a speculation that the developers intended to do the same thing to their instrument, as they perceive as necessary for guidelines: to fix a date, at which it must be revisited. However, in this case it must be questioned whether Christmas time is suitable for such a process.

From 1999 onwards, the development of instruments for the systematic evaluation of guidelines started to gain an international dimension. The collaboration of Cluzeau and colleagues developed an instrument, which they claimed to be the basis of a common approach to assessing guideline quality in Europe (Cluzeau et al. 1999, Europarat 2001: 17ff). As a major step forward, this instrument was extensively tested by 120 raters on 60 guidelines for its face validity, reliability, internal consistency and its inter-rater agreement. It was found that the instrument led to acceptable results in all categories. With its 37 items in three dimensions, the instrument was well accepted by the raters. On the basis of this, the authors concluded that their instrument was useful for judging the extent to which the assessed
guidelines were developed systematically, and to what extent knowledge about the successful implementation of guidelines is considered. However, a major flaw of the system is that after assessing a guideline with this instrument, the scientific quality of the guideline's recommendations still needs to be assessed. Only comments can be made about the rigour of the guideline development process, the clarity of presentation and about issues on guideline implementation. Not least, again there is no graphical work up of the instrument as to facilitate the assessment procedure, and a scoring system is missing.

A major step away from the more or less national approaches to the appraisal of guideline quality is marked by the engagement of the Council of Europe. One of the key recommendations of the Council was that the governments of the member states should create a coherent and comprehensive framework which ensures that the national methods for drawing up and evaluating guidelines for optimum medical practice fulfil criteria that are internationally accepted and most up to date. In addition to that, the Council recommended that an internationally co-ordinated network of researchers should critically assess the methodology for evaluating guidelines, including the effects on cognitive processes and medical knowledge of health professionals (Europarat 2001:10-14). Probably as a consequence out of the engagement of EU institutions, between 1999 and 2000, a group of experts from most member states and the accession states of the EU has developed a recommendation for a framework on drawing up, evaluating, renewing and the active dissemination of evidence-based guidelines on a national basis (Europarat 2001: 6).

These efforts on EU-level brought together experts from national institutions that publish information about the development and evaluation of guidelines, such as from Scotland, Germany, the Netherlands and France. In order to boost these efforts, the EU funded an international group of experts, which developed an instrument for the evaluation of guidelines in Europe. This instrument is called AGREE, which stands for Appraisal of Guidelines for Research and Evaluation in Europe (The AGREE Collaboration 2001, Europarat 2001: 17ff).
3.3.4 Appraisal of Guidelines for Research and Evaluation in Europe

AGREE, which stands for Appraisal of Guidelines for Research and Evaluation in Europe, is an international collaboration of experts from 13 European countries, which was funded under the BIOMED-2 Programme of the European Commission (The AGREE Collaboration (n.d.), The AGREE Collaboration 2003). This collaboration developed a framework and an instrument for assessing the quality of clinical practice guidelines, by which a global assessment of a guideline can be achieved. In the end, marks should give evidence about the level of quality for each of the more formal criteria according to which a guideline can be judged. With the instrument and an additional collection of recommendations for people setting up guidelines, AGREE aimed at improving the quality and effectiveness of clinical practice guidelines.

The AGREE-instrument for assessing the overall quality of guidelines (The AGREE Collaboration 2001) is a further development of the instrument of the collaboration around Cluzeau, which has been described above (Cluzeau et al. 1999). Its predecessor was already designed for the use on guidelines from different countries, therefore laying the basis for assessing guideline quality in Europe. The previous instrument had been tested by 120 raters on 60 guidelines for its face validity, reliability, internal consistency and inter-rater agreement, and was found to provide appropriate results in all categories. As the further development of this, the AGREE-instrument was also intended to appraise not only guidelines in the country of its development (Europarat 2001: 17ff). Its abilities for this were proven in a test run on more than one hundred guidelines in twelve different countries, and also the development and refinement of the instrument are well documented (The AGREE Collaboration 2001 + 2003).

As another major advantage, the AGREE-instrument was designed and found to be appropriate for guidelines set up by multidisciplinary teams. This is of critical importance when appraising guidelines on antenatal care. Moreover, the AGREE-instrument can be used for guidelines on any disease area, including also such on diagnosis, health promotion, treatment or interventions. Also this is of critical importance for the appraisal of guidelines on antenatal care. Finally, it was reassuring to find that the AGREE-instrument was also judged by the
Council of Europe as the most promising instrument for the evaluation of guidelines in different countries that were set up by different groups of experts (Europarat 2001: 17ff). In order to demonstrate the instrument's abilities with regard to assessing national guidelines on antenatal care in the EU, it is discussed in greater detail. For this, the authors of the instrument were asked for additional material about the instrument itself, its development and testing (Burgers 2004, Cluzeau 2004, The AGREE Collaboration 2003).

Properties of the AGREE-instrument

The developers of the AGREE-instrument identified six domains, which define the overall quality of a guideline. Each of these domains was designed to shed light onto a separate dimension of guideline quality. The domains are:

Scope and purpose

Contain a specific statement about the overall objective(s), clinical questions, and describes the target population

Stakeholder involvement

Provide information about the composition, discipline, and relevant expertise of the guideline development group and involve patients in their development. They also clearly define the target users and have been piloted prior to publication.

Rigour of development

Provide detailed information on the search strategy, the inclusion and exclusion criteria for selecting the evidence, and the methods used to formulate the recommendations. The recommendations are explicitly linked to the supporting evidence and there is a discussion of the health benefits, side effects, and risks. They have been externally reviewed before publication and provide detailed information about the procedure for updating the guideline.
Clarity and presentation

Contain specific recommendations on appropriate patient care and consider different possible options. The key recommendations are easily found. A summary document and patients' leaflets are provided.

Applicability

Discuss the organisational changes and cost implications of applying the recommendations and present review criteria for monitoring the use of the guidelines.

Editorial independence

Include an explicit statement that the views or interests of the funding body have not influenced the final recommendations. Members of the guideline group have declared possible conflicts of interest.

In order to demonstrate the validity of the domains, the developers of the instrument assessed their face validity, construct validity and criterion validity individually. It could be demonstrated by a field test that 95% of the appraisers found the instrument useful and manageable to assess guidelines. Moreover, p-values below 0.05 and 0.01 confirmed the construct validity of the individual domains. The constructs used within the instrument are therefore sufficiently useful for discriminating high-quality guidelines from those with a lower overall quality. To assess criterion validity, it was necessary to use a rater's overall assessment of guideline quality as proxy measure. This revealed with correlation coefficients of p < 0.001 that a rater's overall assessment of guideline quality correlated with the scores of the individual domains.

All in all it can be concluded that the AGREE-instrument is useful for appraising the overall quality of a guideline, able to capture the properties of a guideline, relevant for its quality. However, this is only the case, when a guideline and its development process are well documented (The AGREE Collaboration 2003). It can not be excluded that a guideline fulfils all relevant criteria, but that a lack of documentation of these issues leads to a categorisation as a
guideline of minor quality.

In addition to the domains' properties for appraising the quality of a guideline, this categorisation helps the user of the AGREE-instrument to approach the appraisal of a guideline in a structured way. To critically assess a guidelines quality, 23 questions are used altogether. As an additional advantage, these questions are set up in the easy format of a 4 item Likert scale. Ratings from strongly agree to strongly disagree add a specific weight to the assessed criteria. In order to avoid wrong judgements and misunderstandings, additional information and explanation is available for most of the items. With this approach, for the first time the level can be assessed, to which a criterion is fulfilled. This is a definite improvement and goes far beyond the simple recognition of the presence or absence of a criterion, which was common in the earlier instruments. However, the question must be asked of whether the rating is subjective, or sufficiently reliable.

To demonstrate the internal consistency, the Cronbach α coefficient was calculated for each domain. The intra-class correlations were calculated to assess the reliability within each domain. This was done for single raters' ratings, and for the mean ratings of two, three and four appraisers. It was found that the internal consistency ranged between 0.64 and 0.88. This was judged as acceptable for most domains, and explained for the domain 'editorial independence', where this was not the case. However, also the domain 'clarity and presentation' scored low with a Cronbach α of 0.69 and an intra-class correlation of 0.57. Unfortunately, this was not explained. It is assumed that the findings from the appraisal of two national guidelines on antenatal care presented later in the thesis will shed light on this.

Another finding from the testing of the instrument's reliability demonstrated that the number of raters appraising a guideline affects the reliability of the intra-class correlations. Intra-class correlations substantially improved with the number of raters. As an example, the intra-class correlation for the domain 'scope and purpose' increased from 0.44 for one rater to 0.76 for four raters. The same effect was observable for all domains. With four raters, intra-class correlations from 0.57 to 0.91 are achievable. It was therefore concluded that at least four
raters should be used for appraising a guideline with the AGREE-instrument.

In its final section, one of the real advantages or the AGREE-instrument becomes clear. In contrast to the previously discussed instruments, AGREE leads to a conclusive decision about the level, to which the quality criteria are fulfilled for each of the domains. However, it is not wanted to make an overall numerical judgement of an appraised guideline by summing up the final judgements to a single mark (The AGREE Collaboration 2001). Instead, the appraisers of a guideline are requested to make a clear final judgement as to whether a guideline is of such quality that it should be implemented. Options for the final judgement range from “strongly recommend” to “would not recommend”. “Unsure” is also on offer, as it is “recommend (with provisos or alterations)”. This final step is the key to ensuring that only guidelines of sufficient quality are implemented, and to a real improvement of guideline quality. However, the AGREE-instrument does not provide a measure for estimating the impact of a guideline on patients’ outcomes (The AGREE Collaboration 2001).

However, also the best available instrument to assess guideline quality to date has its flaws. One of the difficulties involved with the application of the AGREE-instrument is that it should be used by more than one appraiser on the same guideline to achieve satisfactory reliability. For optimum results, four raters are recommended. This can be difficult to achieve, e.g. when it is used for a thesis. However, as this issue of reliability is made transparent by the authors of the instrument, measures can be taken to overcome this problem.

In addition to that problem, the idea of splitting up the quality of a guideline into several separate domains provides an in depth view into the quality of each domain, rather than of the entire guideline. In the user manual it is clearly stated that the scores of the individual domains should not be used to calculate a single quality score. Therefore, this in-depth procedure might hinder a comparison of guideline quality, when other guidelines have been assessed with instruments that judge quality by a less detailed procedure, ending up with a single score. The situation becomes even more complicated, when another instrument is used that uses other domains, or defined the same domains differently from those in the AGREE-instrument. As a
consequence, it is obvious that when it is the aim to compare different guidelines on the same topic, they have to be appraised with the same instrument.

Conclusions from the review on the appraisal of guideline quality

Drawing the findings of the review on the quality of recommendations and on Instruments for the systematic appraisal of guidelines together, the following conclusions were drawn:

- An important step in the evaluation of a guideline is the appraisal of the scientific quality of its recommendations.

- The scientific basis of a guideline is only one indicator for its overall quality, although of crucial importance.

- Other indicators, such as the scope and purpose of a guideline, stakeholder involvement, rigour of development, clarity and presentation, its applicability and editorial independence are important factors to indicate the quality of a guideline in Europe.

- An instrument is available, which was specifically designed to assess the overall quality of guidelines set up by multidisciplinary teams, and is applicable to guidelines from any disease area, including also such on diagnosis, health promotion, treatment or interventions. Moreover, this is the first instrument with proven validity and reliability for the appraisal of guidelines from different countries.

- The AGREE instrument has a lack of discriminating properties between guidelines of minor quality, which do not fulfil the set criteria, and guidelines with a lack of documentation (The AGREE Collaboration 2003). It could be therefore that a guideline fulfils all relevant criteria, but that a lack of documentation of these issues leads to a categorisation as a guideline of minor quality. This lack of discrimination might be overcome by using a high-quality guideline on a certain topic as a reference value, against which the recommendations of a guideline with a lack of documentation can be compared.

However, the most important conclusion to this analysis is that there is still the link missing
between the overall quality of a guideline, and the quality of its recommendations. In the best case, the overall quality of a guideline is excellent, and its recommendations are based on the best available evidence. However, it is unlikely that this is always the case. It seems to be possible that the recommendations of a guideline are based on scientific evidence, but that the overall quality of the guideline is poor. It might also be that the overall quality of a guideline is excellent, while the recommendations are lacking scientific backup. However, the relation between the methodological quality of a guideline and the quality of its recommendations needs to be explored. This will be addressed by the thesis.

3.4 Theoretical framework: Evidence-based policy

One of the conclusions from the previous analysis was that the quality of the evidence underlying a guideline's recommendations plays an important role for the quality of the entire guideline. Initial reading on the subject brought up a number of publications from professional and scientific journals, from textbooks and also from the public press, which inform the debate. From this it was found that the crucial theoretical problem when setting up guidelines is how to make decisions on recommendations in the light of limited or contradictory evidence, or on the basis of evidence from different sources and disciplines. After critically exploring this phenomenon it was found that evidence-based policy-making exactly deals with this problem. Which concepts exist under the paradigm of evidence-based policy will be critically analysed in the following description of the theoretical framework of the study.

At the beginning of the following section, it is demonstrated how debates about the evidence-base for any decision in health care have arisen from legislative changes, improvements in technology and changing professional practices. It becomes also evident that such debates and conflicts can be very complex and protracted, involving numerous agencies in their attempts to provide the best evidence to prove their case. Hence, the conflicts about what suitable evidence to inform professional practice and decisions on health care needed to be analysed. From this it was found that a thorough description of the nature of evidence is
needed to lay the basis for a profound discussion.

Despite the aim to ensure the generalisability of the discussion, it was decided to appraise all concepts and themes in direct relation to antenatal care as the subject of study. The study therefore started with an extensive review and analysis of the literature on the evidence for antenatal care in the EU and resulted in the development of a conceptual framework for analysing evidence that informs policy and practice. The framework uses evidence as a central theme. Assumptions about evidence-based health care and health policy, the nature and hierarchy of evidence, the use of evidence to set up guidelines, and factors with the potential to modify decisions will be discussed and used as the main frame of reference for this study.

3.4.1 Evidence-based health care and evidence-based policy

As will be discussed in greater depth in a separate chapter of this thesis, the member states of the European Union are striving to limit their health care expenses through the rational allocation of resources in order to avoid the rationing of health care. This is due to the fact that the demand for unlimited access to and use of health care can not be satisfied with the limited resources. Throughout Europe, this lead to an increase in management activities in the health sector (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999, Mossialos 1998: 2). Despite this increase in managerial activities, no country has as yet gone so far as to exclude services which go beyond a core package of vital care, which would be unwise as the commitment of the Union's citizens to a universal health system was found to be high (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1998: 15+18). However, there is a clear trend towards evidence-based medicine and the setting up of clinical guidelines in order to meet the challenges of some of the most prominent factors affecting the health care systems in the member states of the EU at present. These factors are (European Health Management Association 2000 citing multiple sources):
• The drive for cost-containment in the public sector against the background of economic pressures.

• The increasing pressures of 'consumerism' and of higher expectations from patients and citizens.

• The declining power of the medical profession, in particular its reduced role (or demands for a reduced role) in decision-making within health care systems and in resource allocation.

• The role of the health care industry in the national economy.

From the exposure to these forces, changes in medical care practice were suggested to have arisen (Harrison 2002a, Harrison et al. 2002c). The most prominent reaction is the movement for evidence-based medicine and evidence-based health policy. When searching the literature on evidence as such, and especially evidence-based health care and health policy, an excellent definition was found, which grasps the entire problem (Canadian Health Services Research Foundation 2006b):

“Evidence is information that comes closest to the facts of a matter. The form it takes depends on context. The findings of high-quality, methodologically appropriate research are the most accurate evidence. Because research is often incomplete and sometimes contradictory or unavailable, other kinds of information are necessary supplements to or stand-ins for research. The evidence base for a decision is the multiple forms of evidence combined to balance rigour with expedience - while privileging the former over the latter.”

Conceptualising the problem according to this definition and considering the previous analyses of the related concepts, the idea that evidence-based health care is an easy and straightforward task is clearly a misconception and underestimates the gaps and contradictions in knowledge. Nevertheless, in 1999, the European Commission Directorate General for Employment, Industrial Relations and Social Affairs still claimed that evidence-based health care is currently offering the most sophisticated approach for integrating these pieces of evidence in decision-making, as it is based on the systematic synthesis of existing information.
provided by primary studies. However, only three years later, other authors started to become more critical, acknowledging that it can be extremely difficult to produce sound evidence for complex health care situations (Forbes & Griffiths 2002). One suggestion to deal with these problems was to develop methods complementary to the principles of evidence-based health care, which deal with complexity and different forms of empirical expression (Forbes & Griffiths 2002). Guidance for dealing with what can be called ‘imperfect evidence’ is needed. Before this problem is addressed, evidence-based health care / practice / medicine and evidence-based policy need to be operationally defined. First of all it can be said that (Harrison 1998a),

"Evidence-based medicine is the doctrine that professional clinical practice ought to be based upon sound biomedical research evidence about the effectiveness of each diagnostic or therapeutic procedure."

This definition acknowledges that evidence-based medicine is the statement of a rule of how to deal with ‘evidence’. Moreover, it makes clear that biomedical research evidence should form the basis of clinical practice. However, the following definition relates this general comment to specific situations (Sackett et al. 1996 cited by European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 46):

"Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients."

This means that evidence-based medicine is the combination of clinical expertise with the best available scientific evidence (Sackett et al. 1997). In this context, scientific evidence is used to re-assess and improve the effectiveness and safety of commonly accepted therapeutic procedures and tests, or to replace them by more effective ones. In its pure meaning, evidence-based care is mainly about using best evidence for decisions on individual patients. However, evidence-based care can also be used to guide care for defined groups, e.g. women with normal pregnancies. In contrast to the years of euphoria at the beginning of the movement, today also critical texts on evidence-based care are published (Beller 2002,
Canadian Health Services Research Foundation 2005). When analysing the critique in greater
detail, it becomes evident that it relates to the approaches to the synthesis of and the
conclusions drawn from scientific evidence, rather than to EBM in general. From this it can be
concluded that evidence-based medicine is only as good as the reviewer, the scientific quality
of appraising the underlying evidence, and the consequences drawn from this.

However, the fact that there is a trend towards the introduction of evidence-based guidelines
for nearly all questions in health care demonstrates that there seem to be major doubts with
regard to the abilities or the willingness of individual practitioners to constantly appraise the
evidence-base. The introduction of evidence-based guidelines is one of the options to fulfil this
obligation without putting the burden of permanent scientific reviews onto the individual.
Therefore it has become more and more standard practice to set up clinical guidelines.
However, the quality of such guidelines is clearly a factor that needs to be addressed.

Another finding was that the advantages of evidence-based care are obviously judged to be
important enough to make it the basic principle for health care in Germany, as well as in the
UK. In Germany, the government decided that health care professionals holding contracts with
health insurances, health insurances and public hospitals are obliged to provide evidence-
based care. This was even put down in the German compendium of social laws (ÄZQ (n.d.) b,
Kunz et al. 2002). In the UK, evidence-based practice is official policy for the National Health
Service and has become institutionalised with the establishment of the National Institute for
Clinical Excellence (European Commission Directorate General for Employment, Industrial

Drawing the above definitions and aspects of EBM together, the crucial point is that evidence-
based care is the combination of the best available scientific evidence with the judgement of
the clinical expert. Although there might be strong or weak scientific evidence for a certain way
of action, a decision needs to be made about how to deal with this. The principles of evidence-
based health care/ practice/ medicine provide the basis for decisions, but do not make them
(Mossialos 1998). In general it can be said that the same principles that apply to evidence-
based health care/ practice/ medicine apply to evidence-based policy, which can be defined as

"the integration of experience, judgement and expertise with the best available external evidence from systematic research." (Davies 1991 cited by Policy Hub 2005).

Essentially, this definition again exemplifies the need and role of expert judgement in addition to research evidence for the making good policy. Good policy is not the mere dictate of research evidence, but needs the expert to weight the evidence and to define the wanted outcome with regard to effectiveness and efficiency, but also with regard to ethics, interrelations with other policies, and the available resources. By highlighting the need to include such factors in the decision-making process, the possible critique that the concept of evidence-based health care illegitimately reduces the complexity of clinical decisions and policy decision-making (Canadian Health Services Research Foundation 2005, European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 33 citing Carr-Hill 1998) can be avoided. Evidence-based policy essentially aims at improving the competence of decision makers and to strengthen the decision maker's motivation to use scientific methods when making decisions. It clearly addresses the question of how to make health policy and management decisions, including the fact that these must be based on scientific evidence (Muir Gray 1997: 7).

In yet another definition of evidence-based health care, the weighting of evidence is clearly addressed (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 33 citing Hicks 1998):

"Evidence-based health care takes place when (any) decisions that affect the care of patients and populations are taken with due weight accorded to all valid, relevant information".

This definition relates to evidence-based health care, and not to evidence-based policy, but it highlights a crucial factor within all evidence-based activities. Although none of the above definitions gives a hint as to how 'all valid and relevant' information is defined, this can be answered relatively easy and according to the standard rules of science. The person gathering
and assessing the evidence for a problem needs to define the scope and the strategy before searching for valid and relevant information, and has to perform the search accordingly. In contrast to that, the question of what is 'due weight' is much more difficult to answer. In the following, it is attempted to shed light onto this. To start with, the factors which might play a role in setting up evidence-based policies are assessed in greater depth.

3.4.2 The nature and the hierarchy of evidence

When searching the literature on evidence, the most important finding was that this concept is not as straightforward and one-dimensional as it was hoped for. In the previous section, the grading system of the National Collaborating Centre for Women's and Children's Health was already presented, as it is used for the study (National Collaborating Centre for Women's and Children's Health 2003). However, there are other popular examples for the grading of evidence, such as the 'Oxford-Scheme', provided by the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001), or the RCOG (Royal College of Obstetricians and Gynaecologists (n.d.)). Also a task force of the German association of gynaecologists and obstetricians issues a similar scheme (von Minckwitz et al. 2002). As the thesis is studying the baseline clinical care for healthy pregnant women, the evidence-levels of the categories 'Therapy/Prevention, Aetiology/Harm', 'Diagnosis' and 'Economic and decision analysis' might play a role. Therefore the evidence levels of these three are used as examples from the Oxford-scheme. Other schemes for grading evidence, such as that of Royal College of Obstetricians and Gynaecologists (n.d.), or that of the task force of the German association of gynaecologists and obstetricians (von Minckwitz et al. 2002) are not presented. Both associations provide much simpler schemes than the Oxford one, providing no additional information than the above mentioned one of the National Collaborating Centre for Women's and Children's Health. This makes them less useful with regard to discussing the concept 'evidence'. Although the Oxford-scheme might be too differentiated and too difficult to use for the average practitioner, it has excellent properties to frame the nature of evidence.
Table 2.9: Levels of evidence according to the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001): Therapy/Prevention, Aetiology/Harm

<table>
<thead>
<tr>
<th>Level</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic review (with homogeneity) of randomised controlled trials (RCT)</td>
</tr>
<tr>
<td>1b</td>
<td>Individual RCT (with narrow confidence interval)</td>
</tr>
<tr>
<td>1c</td>
<td>All or none case-series</td>
</tr>
<tr>
<td>2a</td>
<td>Systematic review (with homogeneity) of cohort studies</td>
</tr>
<tr>
<td>2b</td>
<td>Individual cohort study, including low-quality RCT</td>
</tr>
<tr>
<td>2c</td>
<td>Outcomes research; ecological studies</td>
</tr>
<tr>
<td>3a</td>
<td>Systematic review (with homogeneity) of case-control studies</td>
</tr>
<tr>
<td>3b</td>
<td>Individual case-control study</td>
</tr>
<tr>
<td>4</td>
<td>Case-series (and poor quality cohort and case-control studies)</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or ‘first principles’</td>
</tr>
</tbody>
</table>
Table 2.10: Levels of evidence according to the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001): Diagnosis

<table>
<thead>
<tr>
<th>Level</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic review (with homogeneity) level 1 diagnostic studies; clinical decision rule (algorithm or scoring system which lead to a prognostic estimation or a diagnostic category)</td>
</tr>
<tr>
<td>1b</td>
<td>Validating cohort study with good reference standards; or clinical decision rule tested within one clinical centre</td>
</tr>
<tr>
<td>1c</td>
<td>Absolute diagnostic findings whose specificity is so high that a positive result rules-in the diagnosis; or a diagnostic finding whose sensitivity is so high that a negative result rules-out the diagnosis</td>
</tr>
<tr>
<td>2a</td>
<td>Systematic review (with homogeneity) of more than 2 diagnostic studies</td>
</tr>
<tr>
<td>2b</td>
<td>Exploratory cohort study with good reference standards, clinical decision rule after derivation, or validated only on split-sample or databases</td>
</tr>
<tr>
<td>3a</td>
<td>Systematic review (with homogeneity) of 3b and better studies</td>
</tr>
<tr>
<td>3b</td>
<td>Non-consecutive study; or without consistently applied reference standards</td>
</tr>
<tr>
<td>4</td>
<td>Case-control study, poor or non-independent reference standard</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or ‘first principles’</td>
</tr>
</tbody>
</table>
Table 2.11: Levels of evidence according to the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001): Economic and decision analyses

<table>
<thead>
<tr>
<th>Level</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic review (with homogeneity) level 1 economic studies</td>
</tr>
<tr>
<td>1b</td>
<td>Analyses based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses</td>
</tr>
<tr>
<td>1c</td>
<td>Absolute better-value or worse-value analyses</td>
</tr>
<tr>
<td>2a</td>
<td>Systematic review (with homogeneity) of level 2 or better studies</td>
</tr>
<tr>
<td>2b</td>
<td>Analyses based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses</td>
</tr>
<tr>
<td>2c</td>
<td>Audit or outcomes research</td>
</tr>
<tr>
<td>3a</td>
<td>Systematic review (with homogeneity) of 3b and better studies</td>
</tr>
<tr>
<td>3b</td>
<td>Analyses based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations</td>
</tr>
<tr>
<td>4</td>
<td>Analyses with no sensitivity analysis</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or ‘first principles’</td>
</tr>
</tbody>
</table>

All these schemes have in common that they aim at grouping the evidence according to its scientific impact following from design and quality of the study in relation to the study type. That this is not a straightforward task becomes obvious from the Oxford-scheme (Phillips et al. 2001), which defines its levels of evidence individually for different questions. Evidence about therapy and prevention is graded slightly differently from evidence about prognosis, diagnostic procedures, and evidence from differential diagnosis/symptom prevalence studies. Also evidence from economic and decision analysis is graded according to own criteria. This demonstrates that different questions can not be answered optimally by one single type of study. Evidence is therefore dependent on the question to be answered.

A clear disadvantage of the Oxford-scheme is that it can be applied only to straightforward quantitative questions. This limits its use when addressing complex, qualitative, or non-
empirical questions, such as on patient satisfaction, or attitudes towards certain care schemes. Unfortunately, such questions play a major role when assessing the entire properties of antenatal care. For such questions, a grading system would be needed in which the scientific impact of qualitative methods is mirrored. When assessing this in relation to antenatal care, it was found that the debate on evidence became more differentiated over the last years. Especially professionals from other disciplines than medicine, such as nursing, argue that there are many interventions which can not be assessed easily with the instruments traditionally used for generating evidence for evidence-based care (Forbes & Griffiths 2002).

However, when addressing questions that can be answered by scientific studies, there is one crucial criterion for the inclusion of studies relating to any of the grading systems. This is the fact that there are trials of high and low scientific quality within each of the categories. There are randomised controlled trials, which are well designed and conducted. Unfortunately, there are also ones for which this is not the case. Nevertheless it has to be acknowledged that scientific research means that data have been gathered and critically appraised according to explicit and sound principles of scientific inquiry (Policy Hub 2005). However, these criteria are only useful for distinguishing evidence from within the same category, but not to define the level of evidence that should be added to a certain category of studies.

When the aim is to answer a clinical or research question, it is necessary to determine what study design is appropriate to answer the question best. On the basis of that decision, studies need to be identified, which fulfil the scientific standards for the respective category. Unfortunately, there are not only straightforward questions in health care. It can be extremely difficult to producing sound evidence for complex health care questions involving non-empirical questions (Forbes & Griffiths 2002). One suggestion to deal with these problems was to develop methods complementary to the principles of evidence-based health care have, which deal with complexity and different forms of empirical expression (Forbes & Griffiths 2002).
Different kinds of evidence

In order to get an idea about the kinds of evidence available, some examples for evidence used in policy-making are discussed in the following. However, it has to be noted that these examples have been chosen specifically for their relevance to policy-making, having in mind that the study relates to guidelines on antenatal care on the national policy level. The kinds of evidence available to assess the effectiveness of individual tests used for the recommendations within a guideline is represented well in the grading system of the National Collaborating Centre for Women's and Children's Health (2003), which has been demonstrated above. However, the kinds of evidence available for policy-making and for making decisions about medical practice are not mutually exclusive. The following examples therefore be regarded as additions to those commonly represented in the above mentioned grading systems. If not indicated otherwise, the following categories of evidence are derived from the internet service of the Government Chief social Researcher's Office (Policy Hub 2005).

Experimental and quasi-experimental evidence, such as randomised controlled trials, controlled before-and-after studies, or the various types of matched comparison studies. Such studies are useful when valid and reliable information about the effectiveness of an intervention is needed in comparison to other interventions or no intervention at all. This kind of evidence, together with the levels of the scientific impact of the respective methods, is commonly represented in grading systems, such as those presented previously (National Collaborating Centre for Women's and Children's Health 2003, Phillips et al. 2001). When searching for an overall classification of such approaches to generate knowledge, it can be said that it is the traditional approach to develop new or more detailed insights into a previously defined field. It has been described as more theoretically driven and linear, trying to discover independent and objective knowledge (Smith 2004: 114). Understanding the findings from the viewpoint of those involved, or in relation to a specific situation plays a minor role.

Survey and administrative evidence comes from social surveys and administration, such as the General Household Survey in the UK, or crime statistics. This evidence can either be used as
a basis for experimental or quasi-experimental studies, or as additional information about the
nature, size, frequency, and distribution of the subject under investigation.

Qualitative research evidence provides information about why something works in practice,
how this happens, for whom and under what conditions. It is also about perceptions, attitudes
and opinions of the key players involved and uses diverse methods, such as in-depth
interviews, ethnographies, or observational studies. With regard to the process of policy
making, this evidence becomes especially relevant during the implementation phase.

Economic evaluation evidence is needed when decisions have to be made of what can be
done to achieve an outcome in the light of limited resources, and about how the available
resources can be optimally used in order to achieve the highest benefit. Economic evaluation
evidence comes from studies such as cost-effectiveness analyses and cost-benefit analyses.

Philosophical and ethical evidence plays an important role in policy making, although it is
presumably not often applied in a formal manner. When using such evidence from needs
analyses or studies using consultative techniques, the values, beliefs, ideologies and
aspirations of those affected by a policy are considered. Also difficult decisions can be made
safer, when competing values are involved.

Systematic review evidence finally balances the methodological flaws of single studies through
the rigorous analysis of research evidence, using a previously defined scientific strategy for the
identification, selection and analysis of studies. Systematic reviews aim at making a reasoned
case about what the available evidence from other studies tells about a topic, or a policy area.
The different methods applied for this range from narrative reviews to meta-analyses.

Innovative approaches for generating evidence

Despite the inclusion of several different kinds of evidence and the ways of generating them,
the above list is not exhaustive. Evidence comes in different forms and from different
disciplines. It has been shown that scientific evidence is not only findings from randomised
controlled trials on clinical effectiveness, but also findings e.g. from social scientific research. Moreover, over the last years, innovative ways to generate new knowledge keep coming up. One main feature of these new ways is the idea to mix multiple stakeholders in order to use their specific knowledge and resources, such as competing values, agendas and expectations (Smith 2004: 114). Inter-disciplinary work, with regard to disciplines, as well as to institutions, is of key importance to such approaches. Projects following these new ways might e.g. include the ideas of users, commissioners and providers of research. Moreover, it seems to be observable that formerly uncommon research methods in health care are entering the scene, such as action research and participatory methods (Smith 2004: 120).

This need for more flexible and innovative approaches might be caused by the increasing complexity of questions, framing health care much more comprehensive than as being the treatment of illnesses alone. One example for attempts to overcome the limitations of the strong reliance on randomised controlled trials, the method of cross design synthesis was introduced by an expert panel for the United States General Accounting Office (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 70). This complementary method aims at extending the results of controlled studies to the conditions of medical practice. It combines randomised controlled trial designs with database analyses and was claimed to be an 'extension of the logic of meta-analysis'. Such innovative approaches towards the gathering of evidence for the production of new knowledge are very likely gaining importance in the years to come. However, at the moment they seem to be of minor relevance for the production of guidelines on screening tests for antenatal care. These new methods seem to be useful when innovative and flexible approaches are required, addressing particularly the needs of the recipients of care (Smith 2004: 121).

In addition, such innovative seem to be a good instrument to adopt the delivery and organisation of health care to specific circumstances. Also the European Commission concluded that there are many types of information that may be valid and relevant under particular circumstances (European Commission Directorate General for Employment,
Industrial Relations and Social Affairs 1999: 34). It is explicitly stated that there is no reason to exclude any particular type of information as long as an appraisal is made of its validity and relevance and the information is given ‘due weight’. What this due weight might be will be explored in the section of how to weight evidence. However, the new approaches to generate knowledge are not at the focus of this study, although these new ways for the production of knowledge might start to complement obstetric knowledge in the future. Up to now, the involved health professionals seem to be still in the phase of struggling with the medical or obstetrical evidence about the individual measures. At this stage, the interested reader is therefore referred to the current and the upcoming literature on the subject (Smith et al. 2004).

Although the wealth of possible sources of evidence is stimulating the thought that for every question there is potentially the appropriate evidence, there are limitations and dangers to scientific evidence. Especially when dealing with complex questions, leading to a wealth of literature of different scientific quality, it is often difficult to find simple recipes for complex questions. One example is a multidisciplinary group of experts trying to set up evidence-based quality standards in an online educational project. In the excellent description of the research process, the authors state that they found repeatedly that the collective reflection of their practical experience enabled them to test the validity and transferability of published evidence better than following formal processes, such as the course of critical appraisal checklists (Greenhalgh et al. 2003). This demonstrates well that evidence and experience are neither necessarily exclusive, nor are they necessarily congruent. However, experience, expert opinion and authority are definitely concepts worth assessing in greater depth. This is even more so, as this kind of evidence is continuously ranked lowest in the grading schemes for evidence. In the grading system of the National Collaborating Centre for Women’s and Children’s Health, expert committee reports or opinions and/or clinical experience of respected authorities marks as Level 4 the lowest grade in the ranking (National Collaborating Centre for Women’s and Children’s Health 2003). In the classification system of the Oxford Centre for Evidence-based Medicine (Phillips et al. 2001), expert opinion without explicit critical appraisal, or based on physiology, etc. ranks on Level 5 which is again the lowest possible level. For the
following exploration of the properties of expert opinion, not only scientific publications were reviewed. Also the case of a medical expert was analysed, which was fiercely discussed in the newspapers. This case was selected for its good documentation and its properties for providing an extreme example for the limitations and dangers of expert opinion.

3.4.3 Expert opinion

Within the legal system in the UK, “great store is laid by expert opinion which is regarded as synonymous with research-based evidence; both the prosecution and the defence try to find the best expert they can, that is, the best expert to support their case, ... after an expert has given an opinion, its generalisability and relevance to the individual case under judgement appears to be treated with remarkable naivety” (Muir Gray 1997: 192-93).

Although this statement relates to the legal system of the UK, it demonstrates well some of the major problems related to using expert opinion as evidence. First of all, the question of selection of the respective expert is relevant. As the above statement indicates, it is always possible to find an expert who supports one's own opinion. Although this does not mean that expert opinion should be ignored in any case, it rather guides to assess carefully how this expert was selected, and what the possible reason for his or her selection were.

Secondly, the above statement is a reminder of the fact that expert opinion might come from personal experiences under certain conditions. Therefore, expert opinion should be assessed with regard to the circumstances under which it was gained. The generalisability of expert opinion should be scrutinised as rigorously as this is done in empirical research. It has to be kept in mind that expert recommendations might be biased due to outdated assumptions, personal bias from clinical experience, training and self interest (Woolf 1998 cited by European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 45). However, it has to be noted that there are different kinds of expert opinion. It might be that expert opinion is only the opinion of a single expert, but it is also possible that it relates to
practical experience over many years. However, both are possible and the term 'expert' does not necessarily mean one or the other. It nevertheless makes a difference with regard to the validity of an expert's testimony.

With regard to setting up guidelines, also 'fatigue from the guidelines development process' was discussed as a possible source for biasing expert opinion. All this indicates that in order to equate expert opinion and research-based evidence, the same care must be taken to assess its quality and independence. Otherwise, the following case might happen again.

Only recently, the cases of Sally Clark, Angela Cannings and Trupti Patel demonstrated that expert opinion has to be handled with care. All three women were sentenced to serve life sentences for murdering two or three of their children shortly after their births on the basis of expert medical evidence. The court followed what is known under 'Meadow's Law'. According to Professor Roy Meadow, the former president of the British Paediatric Association, it was extremely unlikely that more than one child dies from cot death in one family. In short terms the 'law' states that "one sudden infant death is a tragedy, two is suspicious and three is murder, unless proven otherwise" (BBC 2003). As this unfortunately happened in all three cases, the above mentioned women were imprisoned on the basis of Professor Meadow's expert opinion. Only in 2003, Sally Clark came free after the revision of the case by the Court of Appeal. Before this, she was imprisoned for more than three years.

To restore his reputation, Professor Meadow explained that his expert testimony played a minor role in some cases. Nevertheless are convictions based on medical expert opinion only since his case now judged as unsafe practice (Doward 2005, Meadow 2002). Due to this, the Attorney General announced in January 2005 that nearly 300 cases in which parents were accused to have killed their children will be reviewed. An additional 30 000 cases in which children had been separated from their parents need to be reviewed as well (Doward 2005).

The tragedy in all these cases is that this expert and specialist for child protection and forensic paediatrics very likely had only the best intentions. He will have been convinced that he acted
in the best interest to protect children, and was even knighted for his efforts (BBC 2003, Doward 2005, Kite 2004). However, this case demonstrates well that special care has to be taken when one expert discovered a condition, was the first person to describe it, coined its name and was the only one who researched the subject. Even more care is needed if the theory helps to answer difficult questions, or closes uncomfortable gaps in knowledge. In such cases there is a chance that the whole phenomenon is an artefact and exists only in the logic system of this expert who describes it so convincingly that it appears real. A member of the House of Lords called this with regard to the case of Professor Meadow inventing a theory without science and fitting evidence into a diagnosis (Howe cited by BBC 2003), as never any real and peer-reviewed evidence was produced to prove the existence of the phenomenon.

In such a case, it could be useful not to accept a theory as true until the contrary is proven, but as provisional, until further evidence underpins it. In the case of Professor Meadow, it took 25 years until his theory was discredited. Before this, it was used as a diagnosis throughout the world (Kite 2004). However, in order to finally clarify the question of the reliability of Professor Meadow's theory, the UK Minister for Children named setting up an international panel to review it as an option (Kite 2004). Despite the fact that it would be interesting to know what happened to Professor Meadow, his theory and the families, whose cases have been reviewed after July 2005, this is not included in the thesis. It does not matter whether Professor Meadow was right or wrong with his theory, but his case itself can serve as a good example of what might happen, if expert opinion can or is not substantiated by scientific evidence, or is at least scrutinised according to similarly rigorous criteria as scientific research.

Similar to the consequences drawn from the Meadow-case, another source claims that "The opinions and judgements of experts that are based upon up-to-date scientific research clearly constitute high quality valid and reliable evidence. Those opinions that are not based upon such scientific evidence, but are unsubstantiated, subjective and opinionated viewpoints do not constitute high quality, valid and reliable evidence." (Policy Hub 2005). This implies that expert opinion can be accepted as evidence, as long as it is not too subjective and can be clearly
justified. Otherwise it can not be accepted as evidence and should not be used for any purposes but the decision that more research is necessary on a subject. The crucial question regarding expert opinion is therefore, whether the opinion is primarily based on personal experience only (Polit & Hungler 1999: 8). However, a major problem arises when an opinion is coming from more than one expert. In such cases, questions must be posed, such as whether an expert panel is independent, or whether it is influenced by a certain case, member of the panel, etc. The validity of expert opinion increases with other evidence supporting it, not necessarily with the number of experts stating that it is true.

However, if expert opinion needs to be substantiated with other evidence, what is it then? Is it a variation or synthesis of the findings from research, adding a personal flavour to it? Is it a mixture of experience and evidence? Or is it an interpretation or addition to limited scientific evidence? If these questions have some kind of truth in them, another question arises:

At which point and under what circumstances do scientific evidence and expert knowledge become congruent? It might be that scientific and expert evidence become congruent for old disciplines, to which little knew knowledge was added over the past decades. In this case, all members of the profession are up-to-date as the facts did not change since their professional training. They can therefore act as experts. Another explanation could be that a very active research community exists, which has the determination to keep the entire profession up-to-date, and is successful with its strategies. As a third option, it is a very small profession, so that the same persons are regarded as experts, who acquire the scientific evidence. However, if expert opinion must be based on the latest scientific evidence in order to be acceptable, it can just be an interpretation of the findings, adapting them to the individual circumstances.

3.4.4 Health policy

From the literature on guidelines and on evidence it was concluded that the crucial point in developing guidelines and health policy of the highest quality is the decision-making about how
to weigh up the available evidence. To explore this, some standard texts on policy-making in general (Parsons 1995), as well as on policy-making in health care were read to broaden the horizon, before setting off to a more focused search (Dolowitz & Marsh 2000, Klein 1990, Klein 1998, Muir Gray 1997, Walt 1994). Also texts relating to the special traits of making health policy in the EU were explored (Graham 2004, Radaelli 2000). For all searches, policy was operationally defined as the decision to act on some particular problem and the subsequent decisions relating to its implementation and enforcement (Walt 1994: 41). This means that policy includes statements of intent, of action, and of what will not be done. According to this, guidelines are policy. Another important point implied is that policy needs decisions.

Health policy, as a specific kind of policy, is a difficult matter. This can be demonstrated well by using a simple definition. It can be defined in the widest sense as any managerial activities striving to produce strategies to improve outcomes and quality in health care (Klein 1998). What makes this definition, and the issue itself difficult is the fact that it contains the notion of quality. Quality is a complex concept. It includes several sub-concepts and dimensions. At the very least, effectiveness, efficiency, equity, access, acceptability and appropriateness are regarded as the essential components entailed in the idea of quality (Maxwell cited by Klein 1998). On top of these, another author found respect, choice, availability of information, as well as technical competence to be a part of quality (Klein 1998). Although these ideas behind the concept of quality are not a problem in themselves, they shed light onto how complicated it is to make decisions for health care on the macro level.

First of all, already the amount of information is challenging, when complex situations are addressed. Evidence must be identified, gathered and examined with regard to its quality and applicability to the situation. An example for the possible extent, the reference list of the NICE-guideline on antenatal care can be used (National Collaborating Centre for Women’s and Children’s Health 2003). 631 studies were reviewed for the purpose of setting up a guideline on the basic care of healthy pregnant women. In addition to the mere extent of the literature, also strategies need to be developed on how to deal with the very likely situation of
contradictory evidence. However, these are issues relating to methodological quality in science, and are therefore manageable when adequate resources are available.

Another major problem when making health policy is that all of the above mentioned components can be assessed individually, but the assessment of each will lead to a result in its own right. Each component is likely to recommend a certain course of action, which can be the exact opposite as the course of action recommended by the findings on another component of quality on the same issue. In this case, decisions are required about what is the most desired outcome, what price can be paid for this, and at the cost of which other outcome variable. This type of problem can e.g. occur for monetary reasons, when there is only money for one intervention, but not for others. Another reason for such a problem could be that an intervention has positive effects for some people, but not for others, or at the cost of other desired outcomes. It can also happen that an intervention is found to be effective and affordable, but not acceptable for ethical, moral, or cultural reasons. Yet another problem can be the lack of either high-quality evidence or the lack of any evidence at all. In all these cases, managerial decisions are required about which weight should be given to the competing goals, and which course of action should be pursued accordingly. Furthermore, when making health policy, not only patient's interests and medical necessities have to be considered. Also the ideas of government and interest groups, such as professional organisations, the pharmaceutical industry and the church play a role. In addition to that, foreign interests or responsibilities, e.g. the regulations from the EU have to be taken into account (Walt 1994: 3).

Considering then that health itself and also the related research are highly sensitive issues with high public interest, decision-making is highly demanding in this area.

Despite the complexity of policy-making, the idea to use policy analysis as framework for the study was tempting. Initially, this framework appeared to provide tools to study the decision-making for specific strategies, such as guidelines on antenatal care. When reviewing the literature in greater depth, it was found that there is a full body of literature relating to diverse fields in which policy is made (Parsons 1995, Walt 1994). This was explored to a certain
degree, before it was concluded that policy analysis does not provide the right framework. The main problem with using this approach was that this study is concerned with one individual aspect of antenatal care only. The focus is not on the complete national strategies regarding antenatal care, but on the national guidelines which specify the content and sequence of measures directed at pregnant women. Relating this back to the definitions as discussed above, national guidelines can be regarded as programmes, which are set up in order to implement a part of the overriding strategy, defining the specific actions to be taken (Dolowitz & Marsh 2000). Using policy analysis as a framework would have meant to consider the organisation of antenatal care as well, e.g. who is providing it and who is paying for it. However, the original aspect which led to the consideration of the above framework still needs to be explored: the problem of deciding about the recommendations for a guideline or policy in the light of limited or contradictory scientific evidence and limited resources.

In order to gain insights into the processes of policy making, the literature was explored. Care was taken to address the problem in direct relation to the EU, as well as to link it to guidelines and antenatal care. At the outset of the analysis, the entire question was approached on the basis of the perception that decision-making on that level can follow two different paradigms.


Within the second paradigm, lessons drawn from former policies, or experience with the same approach in another setting (Dolowitz & Marsh 2000, Radaelli 2000).

The analysis of policy-making under these two paradigms intended to set the scene for the development of a model for decision-making, which can be used to set up a common minimum guideline on antenatal care in the EU. However, during the reading process it became clear that the second paradigm plays only a minor role in the context of the thesis. Antenatal care is a complex intervention, involving regular check-ups of maternal and foetal wellbeing, but also
preventive measures and social support. As outcomes of care are also determined by social
environment and lifestyle, genetics, health care and physical environment (Muir Gray 1997: 12)
it is extremely difficult to relate outcomes to care. For this, it is unlikely that the strategy of
‘learning from the best’, which would mean the transfer e.g. of the Swedish model of antenatal
care to a country with a higher perinatal mortality rate, plays a major role within the EU.

After this decision to concentrate on health care decisions under the first paradigm, the
literature around evidence-based medicine [EBM] was explored. However, it did not take long
to discover that medicine is not the only issue that claims to be evidence-based. Due to this,
the search was broadened to evidence-based health care, evidence-based practice and
evidence-based health policy. In addition to the internet links listed in the section on the review
of guidelines, again the links available under the heading ‘evidence based medicine’ were
used for this search (Appendix 1, Reference 9). In addition to those entrance links, the internet
service of the British Cabinet Office’s Government Social Research Unit was used, which also
regularly issues a bulletin to keep all interested parties informed (Appendix 1, Reference 2).

3.4.5 Factors with the potential to modify decisions

The fact that more than scientific evidence is needed for making decisions has been discussed
already. The British Cabinet Office’s Government Social Research Unit identified five important
factors to policy making, which are demonstrated in Figure 2.1.
Despite its straightforward appeal, this model of factors is not that simple. The problems related to the definition and to the utilisation of evidence have already been discussed. At this point, only some of the different kinds of evidence available for health care decisions should be briefly named. When making decisions on the population level, results from clinical research, epidemiology, health economics, health system research together with the systematic review of outcome-research, effectiveness research, appropriateness research, cross design synthesis, satisfaction research and quality development could and should be used (Canadian Health Services Research Foundation 2005, European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999). With this, also the key questions about equity, effectiveness, safety, patient satisfaction, cost-effectiveness, quality and appropriateness of measures have to be answered (Muir Gray 1997: 103ff).

When planning to implement a policy, it is necessary to transfer these factors to the strategic level of financing, organising and delivering health care. In this process, again other factors contribute to decisions, such as contracting, budgeting, accreditation, skill mix, education, etc. (European Commission Directorate General for Employment, Industrial Relations and Social
Affairs 1999). These factors could probably be summarised under the heading 'resources'.

However, although they need not be necessarily defined as evidence, there are several other factors to be considered, such as habits, tradition, reluctance to change, or radically different opinions. Such factors might fall into the realm of experience, or what can be defined as non-scientific colloquial evidence (Canadian Health Services Research Foundation 2005). As such factors definitely have an influence on decisions, models with less than the above mentioned five factors bear the danger of not grasping the entire situation. One example for such a model would be that of Muir Gray, who claimed that decisions about groups of patients or populations are made by combining three factors: evidence, values and resources (Muir Gray 1997: 1).

The fact that also the limited model mentions the factor 'values' indicates that it might have special importance for policy-making. Unfortunately, for this factor no universal definition can be given. Values are likely to vary from patient to patient and from practitioner to practitioner, but also from decision-maker to decision-maker. An example for this might be that Jehova's witnesses rather die than to accept blood from other people. Another example might be the decision to save the life of the mother, rather than that of the unborn foetus, if only one can be saved. It is especially important to acknowledge the factor values when making decisions for entire populations. The decision to allocate resources to maximize health gain in return for any given level of expenditure represents a very basic, but important value decision in the health sector (Harrison & Moran 2000: 498). Another example for such a general decision is the rule of rescue, which means to prioritise the treatment of live-threatening conditions, before serving other needs (Harrison & Moran 2000: 498). After that, other factors can be considered, such as effectiveness and uncertainties, cost-effectiveness, cost-utility, equity and equality, i.e. to enhance equity, which is to ameliorate the position of people who cannot afford the care form which they might benefit (Harrison & Moran 2000: 498). This indicates that usually an open or covered ranking of priorities exists, according to which resources are allocated.

Thinking the consequences of individual decisions through, it seems to be likely that the establishment of such a prioritising system is a sensitive and difficult task. This seems to be
even more difficult, when the values of diverse cultures, religions, ethnicities, classes, genders and age groups have to be considered, such as in the EU. Moreover, also the values and interests of governments and their policy makers, such as politicians and bureaucrats, of interest groups like the military, trade unions, church, professional groups, community structures, and foreign interests might play an important role (Walt 1994: 3). In addition to the above mentioned factors, there are other influences and decision-modulating factors in making context-sensitive health policy. Under the header ‘colloquial evidence’, not only values were named, but also political judgment, professional experience and expertise, habits and tradition, as well as pragmatics and contingencies were subsumed. Not the least, also lobbyists and pressure groups influence decision-making in health care (Canadian Health Services Research Foundation 2005). Despite the extensive review of the important contributing factors to policy making (Policy Hub 2005), a last one remains: judgement. This can be regarded in line with the crucial question of

3.4.6 How to weight the evidence to set up guidelines?

Integrating external evidence with experience, resources, values and judgement has been defined as the task of evidence-based health care and evidence-based policy. This means essentially balancing and weighting evidence from different sources, and the process of adapting the findings to the specific circumstances of a patient or a population. Although the methods of evidence-based health care help to make decisions transparent, they do not resolve the difficult value judgements, which are usually at the core of most strategic decisions (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 36 citing Hicks 1997). When setting up guidelines, also strategies are needed to deal with the absence of evidence. This constellation needs to be considered as some authors claim that to date only for 20% of all processes in medicine there is sufficient evidence (Kreienberg & Berg 2002). As a consequence, most of the currently available guidelines are based on limited scientific evidence and the judgement of the person or persons who made the
decisions. Scientific evidence and the judgement of the decision-maker are therefore intertwined, and not mutually exclusive. All the doubts and questions relating to any activity that claims to be evidence-based, are summarised in the following statement (Klein 1998):

"One of the fashionable mantras of our time is the demand that policy making should be 'evidence-based'. It is a misconceived demand. For the real trick of policy making is how to make sensible decisions, given inadequate, incomplete, and ambiguous evidence."

In this respect it is important to acknowledge the difficulties in making evidence-based decisions, but that inadequate, incomplete, and ambiguous evidence is not the only reason for this. One author tried to explain this, differentiating between decision making and decision taking (Muir Gray 1997: 210):

"Although evidence is influential in decision making, in decision taking the fears, anxieties and values of the patient may predominate."

Although this statement relates to an individual patient and a decision about a specific situation, it nevertheless makes clear that decisions are influenced by the person who is making it. Consequently, questions must be asked of who is in the position to make decisions, and how decisions should be reached. From the above statement it becomes clear that it is difficult to reach decisions for reasons of weak or ambiguous evidence, but also for reasons that lay within the decision-maker. This is likely to be the case not only if the decision-maker is personally affected by a decision, but also for political reasons, monetary constraints, as well as any kind of external and internal values and pressures. This applies at least for decisions, affecting more than one patient, e.g. when setting up guidelines. In such cases, an independent decision-maker is needed who is not directly affected by either course of action and is independent from the interests of third parties. When unpopular decisions have to be made, also a certain degree of courage is needed. Decision-making is therefore a demanding task, needing a well developed and independent personality. However, it might be questioned whether such a personality could be found in practice, irrespective of the issue under
consideration. It might be also questioned whether important decisions relating to entire populations should be made by one single person, or whether such decisions should be made by a group. Another issue to consider is whether the decision-makers should follow a certain strategy in order to make their decisions transparent.

With regard to decision-making, the European Commission makes a clear statement as to who should decide about the weight to be attached to the evidence for setting up guidelines (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 41). According to this, decisions should be based on scientific evidence and expert professional consensus on good medical care. It is made clear that decisions should finally lead to a ‘consensus’ and are to be made by ‘experts’. This means that not only an independent and well developed character is required to make decisions, but also expertise. Decisions should be based on scientific evidence, but should also be complemented by expertise. What becomes also clear is that it is obviously recommended to use a group, rather than a single person for the decision-making process. In the above statement, it is also highlighted that the group of experts should come to a ‘consensus’ decision. The definition of consensus is in this context a formal one (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 59), saying that

"Consensus methods are based on (usually expert) scientific panel conventions designed to establish recommendations on health-care practice within a professional community. Consensus methods utilise the process of group interaction to arrive at a collective opinion."

In order to shed light onto the rigour of the final stage of consensus finding in the guideline development process, the classification system of the German Cancer Society can be used as an example (Deutsche Krebsgesellschaft 2005). On top of certain requirements with regard to the underlying evidence and the way of gathering it, the German classification system prescribes how the final consensus must be reached, and therefore clarifies the role of experts. The grading system of the German Cancer Society is demonstrated in Table 2.12.
Table 2.12: Levels of guidelines according to the method of consensus finding
(Deutsche Krebsgesellschaft 2005)

<table>
<thead>
<tr>
<th>Level</th>
<th>Consensus finding procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>informal consensus reached by a ‘representatively chosen’ group of experts</td>
</tr>
<tr>
<td>2</td>
<td>formal consensus procedure, e.g. Delphi method, or consensus conference</td>
</tr>
<tr>
<td>3</td>
<td>extended formal consensus procedure, including decision-analysis and outcome analysis</td>
</tr>
</tbody>
</table>

Although the above system of reaching consensus seems to be clear and logical, there are several concerns with regard to consensus ratings. The first question relates to the idea of a ‘representatively chosen’ group of experts. From this statement it can be concluded that the composition of the panel influences the results, i.e. the decisions (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 58). As a consequence, the first task in the consensus procedure is to define the expertise needed in the panel. In a second step, persons with the respective expertise need to be identified, considering that they need to be independent, but representative for the group they stand for. Otherwise it might well be that they introduce more bias than expertise.

With regard to national guidelines on antenatal care, a consensus panel might involve not only obstetricians and midwives, but also general practitioners, pregnant women, women who have given birth, paediatricians, a representative from an ethics committee, economists, politicians, or funding bodies. Introducing the necessary obstetrical and scientific expertise for a consensus panel is important for the scientific quality and the practical applicability of a guideline. Introducing patient representatives is essentially an act of democracy and ensures that the guideline is acceptable to those affected by it. The inclusion of persons focusing on ethical and religious questions helps to ensure that the general values of the population are respected. Although a person from government is an elected representative of a nation, it would go too far to regard such a person as a representative of patient’s will. Such a person
might be e.g. useful to advocate the economic aspects in a tax-financed health system. However, with regard to selecting the best expertise for a consensus panel to set up national guidelines, the drive to limit bias can produce a myriad of research, literature and management activities. For this the amount of effort to be invested should be thought through.

When finally selecting the representatives of the respective groups, it has to be kept in mind that the method of consensus finding is vulnerable to the subjective opinion of forceful members of a panel (Phelps 1993 cited by European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 61). As a consequence, either the representatives have to be chosen to balance the forces within the panel, or strategies have to be developed to balance the influence of individuals. To name only three examples for such strategies, it might be an option to use more than one representative for weaker groups, to use consensus procedures, which blind the person opting for or against a decision, or to use an independent and well-skilled moderator to assist in balancing the influences. As conclusion to these considerations, it should be kept in mind that even formal consensus procedures do not guarantee the independence of the resulting decisions and guidelines, but that a critical eye is needed on what interest groups are represented, who the representatives are, and which consensus procedures are used. Yet another excellent idea is that not all possible views and positions must be represented in the decision-making panel, but consultations with relevant interest groups can be used instead (Canadian Health Services Research Foundation 2005 + 2006a). By this, the views of all relevant parties are included, but also weighed up against possible other interests. This process also enhances the independence of a decision-maker, or limits the potential of conflicting interests in a group of decision-makers.

Despite the above mentioned factors that can compromise the quality of guidelines, a distinct process of consensus finding has clear benefits compared to decision-making by a single person. This is especially the case, the more limited, ambiguous, or contradictory the underlying evidence is, or the more ethical or value judgements are to be made. If contradictory interests are involved, it is beneficial that a formal consensus procedure brings
together the competing disciplines with their respective goals to solve a certain health problem. This helps to balance views and to make the most out of the given resources under the respective circumstances and improves the acceptance of the final guideline across disciplines. For national guidelines, such a process helps to balance the needs of individual patients and those of the entire population, optimising the allocation of scarce resources by investing into the most beneficial measures.

In order to produce good guidelines, the consensus panel has to develop the most effective combination of their respective expertise, which stimulates co-operation and the co-ordination of efforts. This bringing together of ambitions contributes a lot to avoid redundancies as well as gaps in a certain health field and ensures that evidence from various sources and disciplines is considered. Moreover, such an approach produces synergies and stimulates continuing co-operation across disciplines. Drawing all the above arguments together, it becomes clear that the strategies and the 'experts' used to weigh up the evidence play a key role in setting up evidence-based guidelines. The need to select both the strategies for consensus finding as well as the experts with utmost care and critical thought can not be overestimated. However, it should be kept in mind that the first step in the process of developing a guideline is to decide on its aims, based on the existing evidence.

Evidence and expert opinion

The European Commission acknowledged in its document on developing a methodology for setting up guidelines for optimum medical practice that there are many types of information that may be valid and relevant in particular circumstances (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 34). It was also made clear that as long as the validity and relevance of any information are appraised, there is no particular type of information that should be excluded. However, also in this text it was again highlighted that the information should be given due weight. In the following, an attempt is made to grasp the notion of 'due weight' for developing guidelines on antenatal care in the EU.
In May 2005, the Canadian Health Services Research Foundation released "Conceptualizing and Combining Evidence for Health System Guidance", a systematic review on the meaning of evidence in health care (Canadian Health Services Research Foundation 2005). As a major intellectual step forward, the authors of the report developed two categories for the different kinds of scientific evidence: context-free and context-sensitive. This goes in line with the conclusions from the chapter on the hierarchy of evidence in this thesis, which were that evidence should not only be ranked according to the scientific impact following from the perceived independence from any external factors of a study, but that the levels of evidence and the study types need to be defined individually for different questions. The main conclusion was that it is necessary to determine which approach is suitable to answer the question under consideration best, and to rank the appropriate study designs accordingly.

However, setting up hierarchies of evidence implies that questions in health care are such as to be answered by one single set of studies. Unfortunately, questions in health care, and especially those relating to decisions about complex care situations, are usually not to be answered by one single question. When setting up guidelines, different questions and therefore evidence on different aspects of a situation play a role. Hence, it is therefore necessary to define what types of information should count as evidence, and how they can be appropriately combined to create guidance. Guidance is defined as recommendations for action. In the thesis, guidance is used synonymous to guidelines, although it generally refers to health care situations on a larger scale.

When setting up guidelines on antenatal care, evidence is needed e.g. about the effectiveness of individual screening tests, but also about the acceptability and the cost-effectiveness of them. While the first aspect might be addressed best by randomised controlled trials, the second one is likely to be covered well by a qualitative approach. Assessing the cost-effectiveness of a screening test again needs a totally different study design. After gathering scientifically valid evidence on all aspects of one individual test, a ranking of the evidence is needed in order to decide about the inclusion or exclusion of the test in the guideline. As the
evidence addresses different aspects of the same measure, a clear ranking of the evidence according to the scientific properties of the underlying studies is not possible. As a measure to resolve this problem, the idea of categorising research evidence as suggested by a systematic review provides a useful framework (Canadian Health Services Research Foundation 2005):

- context-free, i.e. evidence about what works in general, and what might be achieved under ideal circumstances, and

- context-sensitive, which is evidence about the conditions of implementation.

The classification into context-free and context-sensitive allows to assess the relevance of evidence, and to make decisions accordingly. After the general effectiveness of a measure has been proven by context-free evidence, context-sensitive evidence is especially useful. This is due to the twofold advantage of relating to both, studies testing how contextual factors moderate an effect, but also about the factors which influence the possibility to implement an intervention. This strategy helps to overcome the current heavy reliance on randomised controlled trials and meta-analyses of them, which have despite their high status also disadvantages. One of these problematic issues is that the subjects, who are part of the later target group of an intervention, are excluded by the study protocol to control for possible co-factors. Therefore, the findings of randomised controlled trials have a lower external validity compared to studies with non-randomised designs (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 47). This represents well the difference between effectiveness and theoretical efficacy, demanding ways to link the results of controlled trials to patient outcomes in the real world. Another issue that needs consideration in this respect is the fact that trials do not always measure all the outcomes of interest to patients and physicians (European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999: 65). This leads directly to the fact that there are several other factors that have an influence on decision-making in health care. This third kind of evidence, which represents any non-scientific information that plays a role in producing health systems guidance, was defined as 'colloquial evidence' (Canadian Health
services Research Foundation 2005 + 2006a).

However, although useful to weigh up the relevance of scientific research for health care practice, the above mentioned classification of evidence to produce guidance in health care is only the first step towards the final integration of the gathered evidence. Acknowledging this, the Canadian Health Services Research Foundation hosted an international workshop in September 2005, in which leaders from organisations providing guidance in health care discussed the role of different kinds of evidence for decision-making and explored models to combine multiple forms of evidence (Canadian Health Services Research Foundation 2006a). As this workshop seems to be the first of this kind on an international basis, it is concluded that the question of how to use different kinds of evidence to guide practice is just starting to come onto the agenda. Moreover, it is obviously a topic of interest at the highest level. This might be due to the fact that over the past years, the production of new evidence has reached a degree such as to require strategies to control the flood of information. In addition, there might be the perception that research is not a stand-alone measure and needs translation into practice.

Deliberative processes to combine multiple forms of evidence

In their international workshop in September 2005, leaders from organisations providing guidance in health care explored models for best practice to use deliberative processes to combine multiple forms of evidence (Canadian Health Services Research Foundation 2006a). Although the final conclusion of this workshop was that the review of deliberative processes did not reveal anything conclusive, it nevertheless led to important recommendations for good practice and laid the groundwork for future experimentation.

One of the central problems identified when producing guidance for health care was the need to develop clear strategies to consider all the contributing factors, which can be subsumed under the header ‘colloquial evidence’. In addition to the factors discussed in the previous sections, professional experience and expertise, political judgement, resources, values, habits and tradition, lobbyists and pressure groups as well as pragmatics and contingencies were
identified (Canadian Health Services Research Foundation 2006a). The term ‘colloquial
evidence’ is therefore used to subsume the many types of non-scientific forms of information,
which are likely to play a role in the process of decision-making. Despite its role in decision-
making, evidence on such decision-modulating factors is usually criticised for not being
gathered in a rigorous or systematic way (Canadian Health Services Research Foundation
2006a). In the rare case of having been gathered with replicable methods, e.g. from the social
sciences, evidence on values, habits, etc. can be treated as scientific evidence. Otherwise,
care has to be taken to consider the quality and nature of the information, but also the source
of the colloquial evidence and its potential to introduce bias (Canadian Health Services
Research Foundation 2006a). Unfortunately, the participants of the workshop concluded that
decision makers should consider the sources of colloquial evidence and give appropriate
weights. As the overriding question in decision-making and producing guidelines is how to
weigh up evidence, here the problem starts again. This clearly demonstrates that the workshop
ends with the same question that was asked at the beginning of the chapter, only at a later
stage in the decision-making process.

Despite this unsatisfactory final conclusion, the report of the workshop of the Canadian Health
Services Research Foundation provides a valuable discussion around the problem of weighing
up evidence. Most interestingly, it was discussed that research evidence could and should not
be combined with ‘colloquial evidence’. Colloquial evidence should be used inform scientific
evidence, guiding and interpreting science and filling gaps when they appear (The Canadian
Health Services Research Foundation 2006a). This means that colloquial evidence comes into
play when the step is made from summarising research findings to providing clear guidance for
practice. It is important to note that the evidence should be assessed first, and then expert
opinion, colloquial evidence and debate should be added. Colloquial evidence is not a
substitute for scientific evidence, but should be used to make sense of research findings.
However, it was also stated that colloquial evidence should inform the selection and
interpretation of research findings. At least the first notion must be viewed critically, as the
selection of evidence according to values might introduce selection bias. There is the danger
of considering only part of a problem or of selecting only part of the literature, rather than summarising the scientific evidence, and then assessing it in the light of additional factors.

Although technical methods for combining different kinds of evidence are possible and were identified by a systematic review, these approaches were criticised for their tendency to generalise, and the difficulty of including new knowledge and experience into the system (Canadian Health Services Research Foundation 2005). Moreover, it is likely that e.g. complex technical approaches using assigned relative weights improve the rigour of decision-making for straightforward questions, but it is not sure how such approaches work for complex questions in health care. As long as no rigorously developed and evaluated tool is available for complex questions in health care, a deliberative process, which is likely to ensure evidence-based guidance in the light of heterogeneous evidence is recommended (Canadian Health Services Research Foundation 2005 + 2006a). Although not neutral in their design, deliberative processes have the chance to enhance the acceptability of a guideline. However, it was found that there is still little evaluation of the best way to design such a process (Canadian Health Services Research Foundation 2005). After the workshop of the Canadian Health Services Research Foundation, the following core features of such a process were defined:

- the presence of a strong, skilful chairperson,

- consideration of different types of evidence,

- engagement between the scientific and decision-maker communities leading to a fair representation of scientists and stakeholders, and additional consultations with all parties affected by the outcome,

- an explicit process of exclusion and inclusion, together with a high-quality syntheses of the scientific evidence,

- face-to-face discussions,

- an appropriate timeline for questions, and finally
- a mechanism to elicit the values of the participants.

However, this process could and possibly should not be used only in the presence of heterogeneous evidence, but in all cases where evidence does not prescribe only one possible way of action. Whenever views, opinions, finances, prioritising, etc. come into play, this process appears to be suitable. The crucial sentence from the summary of the Canadian Health Services Research Foundation was that different forms of evidence do not combine of themselves into guidance (Canadian Health Services Research Foundation 2006a). A decision is always necessary in order to implement a certain course of action. However, although the deliberative process as recommended is convincing and principally plausible, it was also admitted that there is little evidence on its effectiveness to date.

The use of deliberative processes has several advantages, such as making decisions transparent. Although the documentation of such processes can help to demonstrate which influences led to decisions, and which weight was given to the individual components, it is extremely difficult to identify all the points at which non-scientific evidence has influenced a decision and in what ways. This is nevertheless the desirable level on which policy should be discussed. Moreover, the way of dealing with different forms of evidence determines how accurate, achievable and acceptable recommendations, e.g. guidelines are (Canadian Health Services Research Foundation 2005). The use of a deliberative process to deal with different kinds of evidence, as well as the idea to make the methods of policy-making explicit seems to be a possible way to achieve this. However, in the worst case, the entire process is only a new managerial variation of what has always been done in practice.

Although the above mentioned positive aspects of transparency in decision-making prevail, there are also possible problems related to it. One idea that has to be critically thought through is to whom and to what degree the decision-making is made transparent. Although there might be a wealth of pros and cons, only one example should be made in order to not be led too far from the core of the study here, but to demonstrate the need to address the issue. If a decision-making process is made transparent to all, individual parties might have insights that
help them to increase their influence in future decision-making processes.

Another concern is that despite the valuable attempts to make decisions transparent, such processes might not be reliable. It is questionable, whether another decision-making panel with the same expertise would arrive at the same decisions. Unfortunately, there is not yet any evidence on this aspect. Moreover, it is also questionable, if the same panel of experts would again arrive at the same decisions at another point in time. This is most likely not the case, but can be justified with the rapid advancements of science and the permanent increase in individual knowledge and experience of the group members. However, it needs to be proven whether this effect could be also a weakness of the methodology.

3.5 Implications of the analysis of concepts and of the literature

When finally synthesising the findings of the review on the quality of recommendations and on instruments for the systematic appraisal of guidelines, it was found that the appraisal of the scientific quality of a guideline’s recommendations is an important step in its overall evaluation, though not the only one. Instruments exist from national, as well as from international agencies, with which the overall methodological quality of guidelines can be assessed.

However, the most important conclusion to this analysis was that the link between the overall quality of a guideline and the quality of its recommendations, which is their actual content, is missing. As no tool was found to assess this relationship, the crucial question remains: do good guidelines make good recommendations?

To narrow this gap in knowledge, it is also necessary to critically appraise the implications that different methods for setting up guidelines have on guideline quality. This will be achieved by a critical in-depth assessment of two guidelines based on entirely different paradigms. For this, the evidence-based guideline from England and Wales (National Collaborating Centre for Women’s and Children’s Health 2003) and the expert-opinion based guideline from Germany will be appraised (Bundesausschuss der Ärzte und Krankenkassen 2003), using the
instrument developed by a collaboration funded by the European Commission (The AGREE Collaboration 2001). The detailed method for this appraisal will be described later in this chapter.

Developing a model to set up a common minimum guideline for the EU

Despite the above suggestions for using and weighing up different kinds of evidence to produce guidance in health care, it is important to note that up to now there has been no evidence to support any model of decision-making in such a process. It is therefore not surprising that there is also no data available as to whether such a process is suitable to produce guidance at national or even international level. Most importantly, it was found that as yet there is no tool available to set up genuine international guidelines. Instead, by the critical analysis of the concepts of evidence and the problems involved in dealing with it, the idea was reinforced that it is of critical importance to respect national decisions and values.

From the above analysis it was concluded that the existing national guidelines on antenatal care from the member states should be used as the basis to develop a model to set up a common minimum guideline for the EU. These guidelines represent the national decisions about what is wanted and acceptable for the respective populations, and what was found suitable in the light of the need to distribute limited resources. It would not be appropriate to disregard such important decisions. As the consequence of these considerations, a model is needed, by which the national ideas about good antenatal care will be respected, but which also enables the utilisation of the potential benefits of a common minimum guideline.

Developing a model for setting up a common minimum guideline for the EU on the basis of existing national guidelines contributes fundamentally to reach the aims of the thesis: it will further develop the theoretical basis for health policy on antenatal care on EU-level, it contributes to the acceptance of widely recommended methods for antenatal care and helps to optimise the allocation of resources. By integrating the decisions and values of the member states, the shaping of a common European set of values with regard to antenatal care might
be encouraged, opening up the opportunity for a true European approach in the future.

However, before such a model can be developed, deeper insights are needed into what models of decision-making already exist at the level of the European Union, and what the current public health policy of the EU is. Unfortunately, it was found to be nearly impossible to determine the current state of the art of this aspect. Official documents of the diverse institutions of the EU keep coming up on a daily basis. It was therefore attempted to identify the fundamental traits of health policy in the EU, and to present them together with comments on the current discussion in the field. However, completeness, i.e. topicality on a daily basis can not be achieved by the thesis. Despite these obstacles, in the following the current state of the art with regard to health care in the policy of the EU is critically assessed. Special care is taken to answer the question of whether a common minimum guideline on antenatal care would be possible under the regulations of the EU.

Guidelines on antenatal care in the EU

From the previous review of the literature, an area was identified, for which the gap in knowledge is sufficiently large as to justify further inquiry. This is the lack of knowledge about guidelines on antenatal care in the member states of the EU. As the extensive review of the literature did not bring up sufficient data, a survey design was regarded as most appropriate to establish the existence and the content of national guidelines on antenatal care in the EU states. The survey details will be specified at a later stage within this chapter. Before that, the public health policy of the EU is critically analysed together with the decision-making in the EU.
Part 2 - Health care and decision-making in the EU

3.6 Health care and decision-making at European level

The review on antenatal care, as well as the reviews on the making of health policy in general highlighted the need to undertake a review on decision-making within the institutions of the EU. This specific search was necessary to achieve the planned comprehensive overview of the concepts, their theoretical discussion and synthesis for the theoretical framework for the thesis. However, during the review on the current state of the art it became clear that EU health policy can not be separated from the underlying principles of the EU and the current debates. Hence, this part of the review needed to be more extensive than previously thought. However, it was most interesting to find how the institutions of the EU relate to each other, how attempts are made to solve the problems arising from ambiguity of mandates for health matters, and what the member states, as well as their individual inhabitants contribute to the constant change of the system. Special care was taken to assess the issues at the policy level, as well as at the organisational and output levels. With the review on the EU and its implications for health care it was aimed at identifying:

- all relevant institutions occupied with health care in the EU, together with the analysis of their interrelations and

- all key features of EU (health) policy that are important to the study.

To achieve this, the systematic and repeated search of diverse information systems was necessary. The sources searched were

- official websites of the EU, such as the homepage of the Department General for Health and Consumer Protection of the European Commission (Appendix 1, Reference 15),

- the official e-mailing service of the Department General for Health and Consumer Protection of the European Commission. These e-mails provide “timely, readable and reliable accounts of activities and developments in the field of consumer protection, public health, feed and food
safety, health and welfare of animals”, including official documents, speeches, minutes of meetings of working groups, etc. In addition to that, the internet links to documents are also provided (Appendix 1, Reference 7),

- ‘health and consumer voice’ and ‘consumer voice’, the monthly newsletters on public health, and also the yearly publication ‘EU public health information network newsletter’,

- the internet source of SCADPlus, i.e. Summaries of the Union’s legislation, which has a special section on public health (Appendix 1, Reference 18),

- CORDIS, the Community Research and Development Information System (Appendix 1, Reference 5), and

- other media, which provided comments on EU (health) policy.

However, it was a major challenge to find that health care issues are represented in nearly all policies of the EU. Some effort was taken to extract the public health issues from other EU activities, such as from the framework programmes on community action for research and development, in order to get an impression of how deep the strategies of the EU go to develop health care issues (CORDIS 2002). In contrast to the framework programmes and the programmes of community action in the field of public health, the yearly action plans within these programmes were not analysed, which break down the overriding programmes into manageable pieces. If publications of that level had been included it would have broadened the search to an unmanageable degree, without adding new information.

From the documents identified it was found that one of the major challenges of the European integration process is the right of free movement (Paton et al. 2002: 4), as was explained in the section on free movement of persons, services, capital and goods at the beginning of the thesis. On the basis of the latest ruling of the courts, the influence of patient and health care providers’ mobility on the health care systems of the member states was assessed. In particular the need to ensure the quality of care was examined.
The next issue assessed, which is presented in the following section, was public health as part of EU policy. Public health was for the purposes of this operationally defined as health care for populations, and the treatment of those who feel well. Another feature of public health is that decisions are made for entire populations, not for the individual (Muir Gray 1997: 11). From the starting point of a dedicated public health strategy, the development of public health policy in the EU was followed up as far as the Union’s current public health programme (CORDIS 2002, European Parliament and the Council of Europe 2002, Treaty of Maastricht 1992). In this context, also the impact of health policy on the European integration process was assessed.

Other findings explored to considerable depth were the organisational conflicts, task overlaps and gaps, as well as an ambiguity of mandates between the institutions of the EU and their programmes. In relation to this, the current attempts to overcome these problems were also assessed. Attempts to bring divided actions on the same issue together could be demonstrated well for the framework programmes for research and development and the Community’s public health programmes (Commission of the European Communities 1998, 2000 + 2005, CORDIS 2002, European Commission Research Directorate General 2001, European Parliament and the Council of Europe 2002, Watson 2001).

However, the most important question for this part of the review was, whether a common minimum guideline on antenatal care would be possible under the EU regulations. To assess this, the Treaties of Maastricht and Amsterdam (1992 + 1997) were analysed in depth, and the findings were complemented by the analysis of current initiatives at EU level.

As has been stated before, it was concluded from the literature on guidelines and on evidence, as well as from the literature on health policy that the making of health policy and the development of guidelines requires decisions about how to weight the available evidence. As a logical consequence, decisions have already been made in each member state of the EU about what should be contained in a national guideline on antenatal care. Evidence has been assessed and weighted according to the circumstances of the respective member state, e.g. the budget available for health care, prevalence rates of certain conditions, or the ethical
acceptability of diagnostic tests. On the basis of this fact, it was thought that a common minimum guideline should acknowledge and respect the national decisions about what was perceived as important for antenatal care. As a consequence, a model was needed to bring the national guidelines together, acknowledging that only the core measures should be included in a minimum guideline. For this, only measures should be considered which are acceptable to the majority of member states. It was acknowledged that special care needs to be taken not to exclude less wealthy states or states with many citizens to care for from the potential benefits of a common minimum guideline.

From the broad reading of official EU documents, it was evident that the European Union operates sophisticated models for decision-making in order to ensure the balance of interests of its member states. As it was thought that these models might be suitable for the selection of tests for the common minimum guideline, the decision-making structures within the EU needed to be critically analysed. It was hoped to find that with the help of a model for decision-making in the EU, a kind of consensus guideline could be established. Searching for information about decision-making in the EU, the Treaty of Nice and the proposed Constitution were found to be the most fruitful sources (The European Union Constitution 2004, Treaty of Nice 2001).

Due to the diverse aspects of antenatal care in the context of the EU, the framework for the study could not be based on a single concept from a pre-defined discipline, e.g. medicine, or the political sciences. Instead, a comprehensive critical descriptive review of the state of the art was necessary in order to disentangle the conceptual underpinnings of the research problem. This approach was recommended by experts as being appropriate for complex questions, especially when addressing conceptual literature. Such a search should demonstrate reasonable knowledge of the field, giving evidence of what was included and why, rather than being exhaustive with regard to having systematically accessed all publications on the topic of interest. Nevertheless, measures were taken to ensure the rigour of the searches. For this, pre-defined open questions, which were derived from the literature as well as from textbooks on research methods were used to ensure the rigour of the review method. This approach
allowed access to publications and intellectual ideas from various disciplines, on various aspects and of varying scientific quality, but ensured methodological rigour by providing a framework for inclusion/exclusion and analysis. By this, a comprehensive picture evolved, which clarified the underlying concepts of the study, their individual properties, and how they relate to each other. The critical descriptive review of the state of the art therefore provided a broader scope than a narrow systematic review, acknowledging the complexity of the situation.

Despite the obvious advantages of this approach, there are also possible points of critique. One of them could be that the search itself is not reproducible, and that a lack of rigour and depth in the retrieval of the relevant literature can easily lead to a haphazard collection of ideas. Attempts were made to overcome this by ensuring sufficient depth and breadth, using a pre-defined set of questions to guide the searches. Special care was taken to formulate the questions in order to

- ensure the methodological quality of the included material,

- determine, whether the evidence about a concept sufficed, and to

- conclude, whether there were additional concepts to be analysed.

In addition to the clear concept, a detailed track record of the review gives evidence of which sources of information were accessed, which criteria were used for scrutinising it, and how the concepts evolved. From this it becomes clear that substantial effort has been taken to include all relevant material, and all related concepts. Although the review of the state of the art is not reproducible with regard to identifying all the same publications, a search following the track record should end up with the same concepts. Until the searches are reproduced, only the face validity and plausibility can be judged, i.e. how convincing the review and the argument are.
3.7 European integration and public health in the EU

3.7.1 Health care in the Treaties of Maastricht and Amsterdam

Since July 1992, the European Union has had a dedicated common public health strategy. Public health in Europe has its legal basis within the Treaty on European Union, known as “The Maastricht Treaty” (Treaty of Maastricht 1992). Within Article 129, the member states of the EU decided to coordinate their health policies and programmes with the co-operation of the European Commission. By this, the member states wanted to ensure a high level of health protection and prevent widespread and severe illnesses.

In November 1997, Article 152 of the Treaty of Amsterdam enlarged the duties of the European Union in the field of health care compared to its predecessor (Treaty of Amsterdam 1997). Especially in Article 152 §1 of the Treaty of Amsterdam it is clearly stated that community action shall complement national health policies in order to improve public health, prevent human illness and diseases and to obviate sources of danger to human health. The position of public health was strengthened by the Treaty of Amsterdam, as from then on all Community policies and activities needed to consider what was called a high level of human health protection. Co-operation between the member states was highlighted as very important for the health sector, although the organisation and delivery of health services remains according to §5 of the Treaty entirely the responsibility of the member states (Paton et al. 2002: 4, Treaty of Amsterdam 1997). The harmonisation of laws and regulations of the member states is explicitly excluded in §4c.

Although direct arguments could not be found within the treaties for the decision to exclude a harmonisation of the health system within the EU, there might be several good reasons for this approach. One argument of the current Commissioner for Health and Consumer Protection is that the EU started as a Coal and Steel Community, an Atomic Energy Community and an Economic Integration Community, so the history on the EU shows no relation either to culture, or health (Kyprianou 2005b). Although this historical argument is plausible, there must be several other reasons for this clear decision to leave the health systems a national obligation.
One of them might be that limiting patient mobility helps to control costs and enables effective planning, as has been discussed in the section on implications of crossing borders (European Commission Health & Consumer Protection Directorate-General (n.d.)).

Costs

It is not unavoidable that allowing patients to cross borders means losing control over health care costs and rising expenses. When continuing to use antenatal care as an example, there could be another effective way to control costs. Cost control could be achieved by defining a common guideline for routine antenatal care applying to all member states. By this strategy, crossing borders would not enable pregnant women to extend their care, forcing their health system to pay for measures not financed otherwise. Effective planning would be possible, as there would be no additional benefit for pregnant women to seek care in a country other than their home country, purely for the sake of travelling there to access it. Another positive side-effect of a common guideline would be that crossing borders to a country with a common guideline in operation helps to avoid the dangers of omissions or duplications of care.

Checking this idea against Article 152 §4c of the Treaty of Amsterdam, it is concluded that only the superimposition of guidelines by a body of the EU is forbidden, but not the creation of freely accepted consensus guidelines delivered by the member states themselves. §4c can be interpreted as to exclude the harmonisation laws and regulations through the institutions of the EU, but to allow a common approach to public health problems. In accordance with §1 of the same article, a common guideline on antenatal care would contribute to ensuring that community action complements national health policies to improve public health. It might be also suitable to contribute to the Union's aim of providing added value in the sense of continuity of health protection provisions across the Union, by dissemination of "best practice information" (The European Commission (n.d.) c). A common guideline should therefore be facilitated by the institutions of the EU, though not prescribed or developed.

A negative side-effect of a common guideline would be that greater choice and access for
patients, which are the arguments in favour of increasing patient mobility (European Commission Health & Consumer Protection Directorate-General n.d.), are restricted in their positive effects. Choice would be limited to selecting where care is provided, but would exclude choice about the content of care. Better access to care would nevertheless be guaranteed by the fact that care can be provided wherever there is capacity, and not along national frontiers. International collaboration and interaction would therefore allow sharing of resources.

Another reason for leaving the organisation and delivery of health care to the member states could be that the health system is often regarded as a prestigious achievement of a state. The health care system in each member state is highly politicised and ideologically driven, like the National Health Service in the United Kingdom. National governments may perceive any directives from the EU as unnecessary interference with their national health policy. Opting too early for a harmonisation of sensitive areas during the process of growing together might cause unnecessarily increased resistance against the EU amongst its citizens. If harmonisation is achieved later on from within and on the basis of good arguments, the process of growing together is actually supported, rather than hampered. Hence, the exclusion of the harmonisation of health systems might be a strategic measure to promote harmonisation in the long run.

Conflict between national and European interests

The situation might be similar with the idea that health is always a sensitive issue and that member states and their citizens fear that their needs would not be adequately met when an impersonal body, such as the EU, is in charge of this sensitive matter. The perceived lack of control weighs especially heavily in vital areas of life. As a consequence, the harmonisation of health systems must be either avoided or postponed until the EU is no longer perceived as an impersonal body, but as a competent institution which brings additional benefit for its citizens.

The aforementioned conflict between independence and unification is not only evident in the health sector, but is also symptomatic of the entire European integration process (Wismar &
Busse 2002: 270). Despite these conflicting forces, the member states of the European Union are encouraged as well as obliged to co-ordinate their actions in the health sector (Treaty of Amsterdam 1997). Unfortunately, it was often concluded that the Treaties are more explicit with regard to what is not to be done, compared to more vague statements with regard to what should be done in the health sector (Graham 2004). Nevertheless, the European Union has a clearly defined strategy for public health. This is analysed below.

Health issues were acknowledged from the beginning of the European Union. As mentioned at the beginning of this section, the Maastricht Treaty required that the member states coordinated their health policies and programmes with the co-operation of the European Commission (Treaty of Maastricht 1992). With the Treaty of Amsterdam, the health care duties of the EU were extended (Treaty of Amsterdam 1997). This was partially ascribed to a new awareness of health policy at community level, originating from developments, such as the "mad cow" crisis (Commission of the European Communities 1998). From then on, it was required that specific action of the EU complements national health policies. By this strategy, public health should be improved, as well as human illnesses prevented. Moreover, also sources of danger to human health should be reduced. In addition to that, the position of public health was strengthened further, as according to the Treaty of Amsterdam all Community policies and activities needed to consider a high level of human health protection.

3.7.2 Framework programmes on community action for research and development

In order to achieve the overriding public health aims, the European Commission includes health issues in its framework programmes on community action for research and development to direct and coordinate activities in the member states (CORDIS 2002). It is the aim of these framework programmes to break overriding aims down into action plans, to coordinate and stimulate actions across the member states and to contribute to project funding. The framework programmes run for five years and specify what research and action is currently at the centre of interest to the EU.
The current 6th framework programme has a total budget of 17.5 billion Euros (10.94 billion GBP) for the years 2002 to 2006. The main objective of this programme is to contribute to the creation of the European Research Area by improving integration and co-ordination of research in Europe which is so far largely fragmented. At the same time research will be targeted at strengthening the competitiveness of the European economy, solving major questions of society and supporting the formulation and implementation of other EU policies.

"Life sciences, genomics and biotechnology" is one of the seven thematic priority areas of the current framework programme. The "thematic priority areas cover those areas where the EU in the medium term intends to become the most competitive and dynamic, knowledge-based economy in the world capable of sustainable economic growth with more and better jobs and greater social cohesion" (CORDIS 2002). Public health is one of the sub-titles under the header of life sciences, with a budget of 353.8 million Euros (221 million GBP).

Despite the inclusion of public health matters in the framework programmes for research and development, the framework programmes have two main strategic objectives which are not directly related to health. These objectives are strengthening the scientific and technological bases of industry and encouraging its international competitiveness while promoting research activities in support of other EU policies. These two objectives are setting the general scene for choosing priorities and instruments (CORDIS 2002). Health is acknowledged insofar as a high level of health protection has to be ensured in all Community policies (Treaty of Amsterdam 1997), and that it is one of the sub-interests addressed by the programmes. Although health matters are explicitly addressed under the framework programmes, it must be kept in mind that its primary objectives are directed at international competitiveness and the interests of industry. Although research activities under the framework programmes should support other EU policies, health matters also have to be examined from the competitiveness perspective.

When analysing the framework programmes it becomes clear that the focus is on new technologies and competitiveness also in the health sector. Less interest is put on structural, political and low-profit issues, which are important issues in public health. This weighting of
interest becomes even clearer when realising that projects are much more likely to be funded when small or medium sized enterprises [SMEs] are part of the consortium applying for funding. In addition to generally better funding opportunities when SMEs are included in the consortium, at least 15% of the budget of the 6th framework programme is reserved for SMEs only (CORDIS 2002). In addition to these financial aspects, the 6th framework programme has been heavily criticised for favouring of biomedical and basic sciences, often with the more or less hidden objective of industrial production development (Saracci et al. 2005) As a consequence, the framework programmes for research and development can not be judged as an optimum, or sufficient means to improve public health in the EU. Nevertheless they make an important contribution with regard to strengthening international co-operation amongst scientists, and channelling research interests in the direction of the overall health aims.

However, in order to strengthen the framework programmes' capacity to improve the health of the European population, international experts have recommended an epidemiological approach for the 7th framework programme, targeting the whole population (Saracci et al. 2005). The authors subsumed the suggested population based investigations on genetic, social, environmental, and economic determinants of health under the header of "health systems research". For this health systems research, the authors claim at least 20% of the total financial means provided under the section of life sciences in the framework programmes. This might be judged as adequate when bearing in mind that under the 7th framework programme for research and development, health is the first and foremost theme (Saracci et al. 2005).

3.7.3 Organisational conflicts, task overlaps and ambiguities of mandates

Following the ratification of the Treaty of Maastricht in 1992, the European Union implemented a public health strategy with five specific action programmes in 1993: Cancer, Acquired Immune Deficiency Syndrome, drug dependence, health promotion and health monitoring (Commission of the European Communities 1998). After the completion of the 1993 action
programme, the Commission has drawn two main conclusions. The use of action programmes was positively judged as making a difference in making public health priorities between the member states manageable. Unfortunately, the advantages were limited by the considerable administrative burden and a lack of flexibility (Commission of the European Communities 1998). Despite these disadvantages, the strategy of the EU with regard to the management of its public health objectives stayed unaltered.

The Union's current public health programme (CORDIS 2002, European Parliament and the Council of Europe 2002) provides a strategy aiming on the one hand at ensuring a high level of health protection in all Community policies, which is still a direct quotation from the Treaty of Amsterdam (Treaty of Amsterdam 1997). On the other hand, the public health strategy of the EU strives to supplement and co-ordinate policies and actions carried out by the member states in the field of health surveillance and information systems (Commission of the European Communities 2005: 51). By introducing the corresponding measures, the EU tries to build up and strengthen the capacity to react to health threats in a coordinated manner (European Parliament and the Council of Europe 2002).

Most interesting is, however, the overlap in tasks between the framework programmes for research and development and the Union's public health programmes (Commission of the European Communities 1998, CORDIS 2002, European Parliament and the Council of Europe 2002). It must to be noted that both programmes are launched and answered for by the European Commission. However, the framework programmes are under the administrative control of the Directorate General for Research, situated in Brussels, and the public health programme is under the control of the Directorate General for Health and Consumer Protection in Luxembourg. What this means in practice will be demonstrated in the following discussion.

• improving health information and knowledge
• responding rapidly to health threats
• addressing health determinants

The same aims and strands of action were set out in an extended version in the Union’s public health programme for the years 2003 to 2008 (Commission of the European Communities 1998, European Parliament and the Council of Europe 2002, Watson 2001):

• improving information and knowledge with a view to promoting public health and health systems (this is to be achieved by developing a comprehensive system for collecting, analysing and evaluating data and knowledge)

• boosting the ability to respond rapidly and coherently to health threats (this relates mainly to infectious diseases and should be achieved by installing an inter-linked surveillance and early-warning system together with rapid-reaction mechanisms on EU-level)

• addressing health determinants (this involves health-promotion activities on a broad basis to reduce premature deaths. This should be achieved by addressing some of the underlying causes of major illnesses, such as lifestyle behaviours, socio-economic circumstances and the environment).

This overlap of aims leads to the situation that on top of the financial means available from the framework programmes for research and development, a budget of another 312 million Euros (195 million GBP) is available to achieve the same aims under the Union’s public health programme. As a consequence, the budget is split up between two sub-organisations of the European Commission. This might lead to a waste of resources, as a duplication of activities can occur under the different programmes, instead of one activity building upon the findings of another under the same programme. Moreover, there is less control over the financial means being used most effectively or not, and it is additionally problematic that there are two administrations to be financed and run to reach the same aims. There might be open or covert rivalry between the two organisations, which can prove to be most counter-productive. Finances of the EU might then be wasted through internal conflicts, rather than used to make
the most effective contribution towards the public health aims of the Union.

In a critical remark, the European Commission once acknowledged the problem of ambiguity of task division, using health information systems as an example (European Commission Health & Consumer Protection Directorate General 2004). The statement says that the cause of malfunctioning of the health information system in the EC could be the rather haphazard and blurred assignment of tasks among units and organizations responsible for data collection, analysis and diffusion. There is a major ambiguity of task division, e.g. between Eurostat, DG SANCO, DG EMPL, the Community Agencies, PHP surveillance networks, etc. This ambiguity of task division was found to go with overlapping authority and responsibilities among different organisations. It is most likely that the same problems apply to the programmes.

A first step towards resolving the ambiguity of mandates

Acknowledging these problems, the latest development towards a true common European strategy for public health is the introduction of an executive agency for the Union’s public health programme in Luxembourg. Since 1 January 2005, the executive agency for the management of Community action in the field of public health is recruiting personnel in order to support the Commission in raising the profile of the European Community in the field of public health. This agency should start to work together with the European Commission in order to implement and reach the aims of the public health strategy from autumn 2005 onwards (European Parliament and the Council of Europe 2002). On the basis of a decision of the Council of Europe, the new agency is responsible for the implementation of the public health programme for the years 2003 to 2008 and the respective annual work programmes. It manages the calls for proposals and evaluates them. In addition to that, the agency has the executive power over the programmes’ finances. It awards contracts and grants on behalf of the Commission and manages all payments. In contrast to the current situation, the new executive agency provides additional logistical, scientific and technical support to the funded projects.
The agency's main source of funding is a subsidy from the European Union's general budget, and the operating budget for its six years of existence is 28 229 million Euros (about 17 600 million GBP). This sum is intended to cover staff and infrastructure costs, administrative expenditure and management expenditure. As this budget is on top of that allocated to the public health programme, it has to be hoped that the new agency is not just another actor in the bureaucracy of the EU, but that it ends up as the one and only coordinating institution for all public health matters of the Community. Otherwise it would be an unnecessary waste of resources, undermining the trust of Europe's citizens in the Union.

3.7.4 A new strategy to overcome the ambiguity of mandates

In order to overcome the problems arising from the aforementioned overlap of mandates between the Directorate General for Research and the Directorate General for Health and Consumer Protection, the European Commission has proposed to combine the programmes of both directorates under one framework (Commission of the European Communities 2005). This idea finally acknowledges the fact that EU health and consumer protection policies have core joint objectives, which are to

- protect citizens from risks and threats which are beyond the control of individuals and that cannot be effectively be tackled by individual member states alone
- increase the ability of citizens to take better decisions about their health and consumer interests
- mainstream health and consumer policy objectives across all Community policies in order to put health and consumer issues at the centre of policy-making.

In the proposed Programme of Community action in the field of Health and Consumer Protection, the intention to improve health and consumer confidence has been specified as overall aim of both directorates for the years 2007 to 2013 (Commission of the European Communities 2005: 2). Common objectives are to enhance European citizens' health, make them feel safer and more confident, but also to bring Europe closer to its citizens and to
contribute to enhance the competitiveness of the EU. These ideas were recently reinforced by the Commissioner for Health and Consumer Protection (Kyprianou 2005b).

To reach the above mentioned aims and objectives, for the first time, it is proposed that public health and consumer protection policies and programmes be combined under one framework with one executive agency (Commission of the European Communities 2005). From the public health side, the aims of promoting health protection, information and education, safety and integration of health and consumer concerns into all EU policies are brought into the common framework. However, within the common framework, specific structural objectives were also defined, which were judged as important by the Health Ministers of the member states, representatives of Europe's health care sector, the European Parliament and the Commission (Commission of the European Communities 2005, South East Partners 2003). The objectives are to

- contribute to the development of more effective and efficient health systems
- support the objectives above by providing health information and analysis.
- improve European co-operation to enable better use of resources
- meet the information requirements of patients, professionals and policy-makers
- enable access to and to improve the quality of care
- reconcile national health policy with European obligations.

In general, health and consumer policies have similar objectives, e.g. information to citizens. Therefore, it is not only recommended that a common framework is established, but also a common executive agency which streamlines administrative and budgetary procedures. This agency is planned to consist of an extended version of the existing public health programme's executive agency, then including a consumer institute. The agency shall then be divided into a health department and a consumer department, and only common actions are to be managed jointly between the two departments. These strategies were introduced to increase policy coherence, bringing up synergies for the benefit of all member states. As a controlling
institution to the envisaged agency, the Commission will stay in charge of all policy decisions related to defining and managing policy priorities and action, including the annual work plans. The proposed programme budget is 1203 million Euros, which as about 750 million GBP.

However, although the new common framework for health and consumer protection is a promising step into the direction towards a single strategy for health in the EU, there is still a long way to go. As has been demonstrated before, the same issues are not only tackled by the institutions in charge of health and consumer protection, but also by the Department General for Research. This problem will not be solved by a common agency and framework for health and consumer protection. The 7th Framework programme for research will again interact with some of the interests proposed by the draft Programme of Community action in the field of Health and Consumer Protection 2007-2013 (Commission of the European Communities 2005). Some of the issues which arise from the perspective of health and consumer protection were recommended to be tackled with instruments and resources provided under the 7th framework programme for research. This recommendation demonstrates well that there is still an overlap in mandates and tasks, which enables the different parties to transfer unwanted tasks to the other party, or to produce an unnecessary and ineffective duplication of work.

3.7.5 General trends, reproductive health and patient mobility

Over the last years, a shift of paradigms has been recognisable in the EU health strategy. Initially, the European Commission and its subordinate committees were hesitant to exert influence on health systems and organisation, mainly due to §§4c and 5 of Article 152 of the Treaty of Amsterdam, and their clear statements that the organisation and delivery of health services has to be left with the member states (Treaty of Amsterdam 1997). However, the European Commission now takes a more active role and contributes to bringing public health from the member states alone to a true European level. This puts more weight onto §1 of Article 152 of the Treaty of Amsterdam, which highlights the co-ordinating and complementary role of the Community to improve public health. This has been particularly with regard to health
information and infectious diseases as proposed by the combined programmes of health and consumer protection (Commission of the European Communities 2005: 42). In this nearly revolutionary proposal, two important issues with regard to harmonisation in health care are named as fields for common action. Actions within the new framework should

"... enhance scientific advice and expert risk assessment, e.g. ... fostering harmonised approaches to risk assessment and promoting training for assessors."

In contrast to previous documents, harmonisation is no longer excluded, but explicitly sought after. The same change in strategies can be observed in the area of product safety and the regulations for substances of human origin. Here, the document recommends the use of common guidelines and standards, fostering best practice exchange.

Despite these remarkable changes in strategy, these are only little steps towards a true European approach towards major health problems. For the rest of the health systems, there is still a long way to go. As the European Parliament and the Council of Europe stated, due to the principle of subsidiarity and the restrictions placed on the Union in health matters, the active co-operation and wholehearted commitment of all the member states is still essential to the smooth running of the public health programme and to achieving its objectives (European Parliament and the Council of Europe 2002). This situation opens up a window of opportunity for bottom-up approaches, which have their origin in the member states. A common guideline for antenatal care, which is not superimposed and reinforced by the institutions of the EU, would fit exactly into the current legislation and situation of the EU and its public health aims.

At present, there are major efforts to demonstrate to the member states and Europe's citizens how promising it is to learn from other European countries. The current European Commissioner for health and consumer protection, Markos Kyprianou, gave some impressive examples for lessons, which could be learned from other member states (Kyprianou 2005b). One was that the five-year survival rates from breast cancer range from 81% in Sweden to 58% in Poland and Slovakia. Similar differences exist for malignant melanoma. Kyprianou
used these examples to illustrate the potential for co-operation and learning from the strategies of the more successful states within the EU. Such examples might force some member states to act actively for their citizens and to start international co-operations. However, it might also lead to the situation that patients travel to another country with better treatment options for their disease. This would have major impact on the health system of their home country.

Shifting emphasis

Similar to this shift of paradigms in the role of the EU in public health, a shift of emphasis is also observable with regard to the content of health policy. From the former stance of treating illnesses as prime responsibility of health care, the trend goes now towards the promotion of good health, taking a pro-active stance. The idea is to prevent ill health before it develops (Byrne 2004). This is acknowledged in the 6th framework programme for research and development for the years 2002 to 2006 (CORDIS 2002) as well as in the Community's public health programme for the years 2003 to 2008 (Commission of the European Communities 1998, European Parliament and the Council of Europe 2002). Addressing health determinants is one of the three core objectives of both programmes. However, although the public health programme highlights that this involves health-promotion activities on a broad basis, the specifications relate mainly to major illnesses of older people, rather than taking a more general stance. Premature deaths due to major illnesses of adulthood should be reduced by addressing some of the underlying causes of major illnesses, such as lifestyle behaviours, socio-economic circumstances and the environment. However, in this respect, antenatal care has to make an important contribution. By considering a healthy lifestyle and a healthy diet throughout pregnancy, as well as by treating unfavourable conditions immediately, best starting conditions are ensured for the future citizens of the Union. As recent research suggests, conditions during the intra-uterine period can programme an organism for the rest of its life, mainly regarding a predisposition for coronary heart disease (Barker 1995, Barker et al. 1993, Paneth et al. 1995, Eriksson 2005). As this is especially evident for cardio-vascular diseases – one of the major health problems in Europe – antenatal care fits perfectly into the
third major public health theme of the EU: Addressing health determinants (Commission of the European Communities 2005: 44). This is acknowledged under strand 3.2.3 of the proposed Programme of Community action in the field of Health and Consumer Protection for the years 2007 to 2013 (Commission of the European Communities 2005). The aim of this strand is to promote health by tackling its determinants. In this section, reproductive health is also listed amongst lifestyle factors and infectious diseases. Unfortunately, reproductive health is mentioned only as a keyword, with other issues being discussed in much greater detail. If there is any specification at all, sexual health and reproductive health are mentioned with regard to preventing sexually transmitted diseases, rather than putting the focus on improving intrauterine conditions to exert a positive influence on the health of future citizens.

When exploring the themes in the public health policy of the EU, it becomes evident that there is an imbalance with regard to the weighting of issues (Kyprianou 2005b). The focus of interest is on the prevention of

- the “big killer” cancer,
- communicable diseases, and of
- problems arising from demographic ageing.

The same neglect of children’s issues and factors related to reproductive health becomes obvious for the 6th framework programme on Community action for research and development (CORDIS 2002). This programme reflects a heavy reliance on genomics and major illnesses, such as cancer and communicable diseases. Children’s health is not mentioned at all, although it would be of utmost importance to include this into the EU’s public health policy. Children are citizens in their own right, but not yet able to act as self-advocates, particularly at the population level (Rigby et al. 2003). Here, the European Union could make a unique contribution towards the health of its citizens, as children’s health determines the health of the future population (Rigby et al. 2003).

In addition to the neglect of children’s health, other important public health themes just start to
be acknowledged in the proposed programme for a combined framework for health and consumer protection policies (Commission of the European Communities 2005: 5). Issues of inequality, which influence citizens' health and life expectancy, such as poverty, housing and nutrition are now starting to be set on the agenda. However, although it is acknowledged that these are issues of equality, the focus is put on the idea that good health is needed for Europe's competitiveness and economy by ensuring the health of the workforce.

The above described slow and still unsatisfactory shift of emphasis can be judged especially critically, as public health has been, e.g. in the UK, a specialist field of practice since the middle of the 19th Century (Department of Health, Chief Medical Officer 2003). Although it developed out of the field of hygiene, today the definition and the focus of modern public health is much more than that (Allin et al. 2004). The official definition used for England, which is also widely used throughout the EU is the one by Sir Donald Acheson (Allin et al. 2004, Department of Health, Chief Medical Officer 2003 citing Acheson), which grasps the essential focus of modern public health. It says that public health was the science and the art of preventing disease, prolonging life, and promoting health through the organised efforts of society.

According to the UK Department of Health (Department of Health, Chief Medical Officer 2003), the tasks are to monitor the health status of the community, identify health needs, develop programmes to reduce risk and screen for early disease, control communicable disease, foster policies which promote health, plan and evaluate the provision of health care, and manage and implement change. This demonstrates well that antenatal care fits perfectly into this, but demonstrates also that the public health strategy of the EU is very limited and far from developing its full potential. However, it is assumed that there is a decision behind the limited use of public health in the EU. One reason might be the focus on economic competitiveness, trying to produce additional benefit for e.g. small or medium sized enterprises from what is only labelled to be public health. However, if public health is to be used according to a deeper understanding of the subject, antenatal care has to make a real contribution. Offering the best starting conditions to Europe's future citizens by improving intra-uterine conditions has the
potential to prevent disease in later life and thus enhance the health of an entire population (Barker 1995, Barker et al. 1993, Paneth et al. 1995, Eriksson 2005).

3.8 Perceived legitimacy, health policy and harmonisation

At present, the European Union suffers from a major lack of confidence from its citizens and drifts towards an existential crisis. Problems with the ratification of the new Constitution perfectly illustrate this (The European Commission (n.d.) d). It is still not clear whether the new Constitution will ever be signed by all member states of the Union. The citizens of the founding members France and the Netherlands already voted against the new Constitution in 2005. Although it might be that the citizens of these member states simply did not like the idea of sharing a constitution with all other members of the Union, there might be other underlying reasons for their rejection, and there might be a more general rejection of the idea of a European Union as well (Falksohn et al. 2006). It becomes likely that financial considerations have played a role in rejecting the Constitution, when it is considered that two net payer states have voted against the new Constitution. From this perspective it is possible that the citizens of the rejecting member states have voted against the devalued Euro and the resulting problems, or too much regulation and publicised corruption scandals, rather than against the Constitution.

Effects of enlargement

A related motive could be the fear of cheap workers from less wealthy EU states who are now having more or less free access to the markets of the wealthier member states. Or the other way round, 84 percent of the Germans were found to be afraid of jobs being transferred to other member states, where labour is cheaper (Falksohn et al. 2006). For Italy and France, similarly high rates of this anxiety were reported. There is a realistic fear that this mechanism could leave many people of the states where labour is more expensive unemployed and the enterprises bankrupt. However, there could be also a xenophobic reason, with their negative attitudes arising from an underlying fear from the progressing enlargement of the Union, taking
‘unwanted’, or culturally different states on board.

The aforementioned reasons and probably many others seem to leave Europe’s citizens mistrustful about the new Constitution, which might be an indicator for the citizens’ general lack of trust into the EU. The most interesting feature in this respect is that the European Commission obviously has identified this phenomenon. In its proposed framework programme for public health for the years 2007 to 2013, the Commission states that the recommended measures and the common agency for public health and consumer protection will improve efficiency and effectiveness of EU actions and make them more visible (Commission of the European Communities 2005). In this document, it is especially highlighted that this will help to reconnect the EU with its citizens. It is evident that by revising their strategies, the European Commission seeks ways to bring the advantages of the European Union closer to its citizens by demonstrating an additional benefit in fields which lie at the heart of the citizens. One idea is to enhance activities for improving health, to improve policy coherence in the health sector, and to communicate these efforts effectively. The second idea is to make EU policies more visible and transparent by drawing efforts of different institutions together. One example of this is the attempt to bring public health and consumer protection under one framework (Commission of the European Communities 2005), as has been discussed in section 2.6.4.

The proposed measures are advantageous as they create the image of a de-bureaucratisation and the harmonisation of issues which share many objectives anyway. These are promoting health protection, information and education, safety and integration of health and consumer concerns into all policies of the EU (Commission of the European Communities 2005).

The new Constitution would have reinforced the European Union’s mandate for health and consumer protection by explicitly acknowledging its role in promoting the well-being of its peoples in article I-3 (The European Commission 2005, The European Union Constitution 2004). In addition to that, the Treaty of Amsterdam (1997) is reinforced by Article II-95 of the new Constitution in its idea that all Union policies and activities should ensure high human health protection. In addition, the new Constitution states in its charter of fundamental rights
that everyone has the right of access to preventive health care and the right to benefit from medical treatment (The European Commission 2005, Article II-95). It would also extend the European Union’s powers in health matters (Article III-278). Unfortunately, the reinforcements and revisions with regard to the health policy of the EU have not yet brought the envisaged effects, as the new Constitution has already been rejected by the citizens of France and the Netherlands.

A factor which might have acted counterproductively could be the progression of the harmonisation in the EU. One example for this is the development of common health indicators (Bauer et al. 2003), which should be used also at member state level. Moreover, these indicators should subsequently also apply to the applicant countries, selling them with the entire package of issues to be fulfilled before entering the Union. Although such a harmonisation is appealing to the rational thinker, the pressure to harmonise increases for the established member states as well as for future members of the Union. This can be a psychological problem for citizens who feel already over-regulated by the EU.

3.9 Decision-making in the Council of Europe

Serious efforts for improving public acceptance of the European Union are made by changing the decision-making structures in the Council of Europe towards a faster, more flexible and more transparent system. Whether these efforts are effective has to be shown by the reactions of the citizens of the European Union with regard to the proposed Constitution for Europe (The European Union Constitution 2004).

As a basic principle, voting by qualified majorities is the way in which the European Council and the Council of Ministers are required to make their decisions. Despite its straightforward appeal, the notion of qualified majority needs exploration. What qualified majorities are at present is specified in the Treaty of Nice (2001), and what they might be from 1 November 2009 on is specified in the Constitution for Europe (The European Union Constitution 2004).
However, if the Constitution is finally not ratified, the Treaty of Nice will continue to regulate decision-making in the European Union (Berbalk et al. 2005). Unfortunately, the decision-making system of the Treaty of Nice is even more complicated than that of the Constitution.

3.9.1 Decision-making according to the Treaty of Nice

In order to take a qualified majority vote under the current system specified Nice Treaty

- a threshold of weighted votes,
- a majority of member states, and
- 62% of the population of the Union

are necessary to take decisions (Treaty of Nice 2001). Reaching a decision therefore requires the effort to calculate the numbers of weighted votes which are assigned to the member states in accordance with Article 205(2) of the Maastricht Treaty (Treaty of Maastricht 1992). The member states with the highest number of weighted votes are Germany, the United Kingdom, France and Italy with 29 votes each, followed by Spain and Poland, which have 27 votes each. Malta is the country with the lowest number of weighted votes, with only 3 votes assigned (Treaty of Nice 2001). The number of weighted votes basically reflects the number of seats occupied by the individual member states in the European Parliament. Although the assignment of weighted votes according to the seats occupied seems simple, it was reported that often problematic and time consuming negotiation is required (SCADPlus (n.d.)).

The current threshold for decision-making in the European Council is at least 258 votes in favour of a decision. These votes must stem from a majority of the member states, when the proposal to be decided on comes from the European Commission. However, in all other cases the threshold is 258 votes, but at least two-thirds of the member states must be in favour.

In addition to passing the applying threshold for weighted votes and the minimum number of member states for a decision, calculations of the represented population sizes have to be made. For a motion to be passed, a 62% threshold is also required for every decision.
When reflecting on this complicated procedure for decision-making, it appears likely to have an acceptance problem in the public of the European Union. On the one hand, the weight attached to the votes can be perceived as unjust, and on the other hand the decision-making process can not easily be followed by the public. In addition to problems with transparency and justice, this voting system has become much more complicated and involves even more effort since the enlargement of the European Union from fifteen to twenty-five states in May 2004.

3.9.2 Decision-making according to the Constitution for Europe

As a reaction to the weaknesses of the Nice Treaty regarding the decision-making in a Union of twenty-five members, the Constitution acknowledges this altered situation and seeks to improve the functioning of the enlarged Union (Treaty of Nice 2001). The voting system as specified in Article I-25 of the Constitutional Treaty still considers the population sizes of the member states and therefore constitutes another form of qualified majority voting (Berbalk et al. 2005, SCADPlus (n.d.), The European Union Constitution 2004). However, the Constitution proposes a new system of qualified majority voting. The suggested system within the Constitution is ‘double majorities’ and means that the majority of the member states must vote in favour of a decision and that they must represent the majority of the population of the EU.

However, instead of mathematically defining double majorities for a qualified majority vote as 50% of the member states and 50% of the entire population of the European Union, the Constitution goes only half the way. The controversial weighting of votes is abandoned, but for a decision under the Constitution for Europe at least

- 55% of member states (including at least 15 of them) and
- at least 65% of the Union’s population

have to be in favour of a decision, before it is accepted.

The interesting feature is that as long as the EU consists of 25 member states, 15 states represent 60% of the total. However, when a 26th state joins the Union, 15 states automatically
add up to 55%. Hence, the addendum of “including at least 15 of them” presently contradicts the 55% rule, and will become superfluous with the further enlargement of the Union (SCADPlus (n.d.)). However, as it might happen that a current member state leaves the Union, leaving this addendum in place assures stability in the decision-making system.

In order to further complicate procedures, the above mentioned criteria for qualified majorities under the Constitution again apply only for decisions about issues proposed by the European Commission, as this is the case for the Treaty of Nice. However, the same applies then for decisions about issues proposed by the under the Constitution existing Union Minister of Foreign Affairs (SCADPlus (n.d.)). In all other cases, the threshold for qualified majorities is higher and decisions must be supported by

- 72% of member states, representing
- at least 65% of the Union’s population.

In addition to the above mentioned criteria for decision-making by qualified majorities, the Constitution as well as the Treaty of Nice contain additional regulations for the case of decisions reached by narrow minorities only, and specify regulations for blocking minorities.

However, the problem starts with the content of the Constitution and its advantages being not properly explained to the public. The votes against the Constitution by the French and the Dutch in 2005 demonstrate the deep mistrust in the functioning, the administration and the current orientation of the European Union (Falksohn et al. 2006). Important advantages of the proposed Constitution have obviously not been translated to the public, which might have been simply caused by the sheer volume of the document. One example is the vital simplification of the decision-making procedure in the enlarged Union, and the resulting opportunity for far more combinations of member states to take decisions (SCADPlus (n.d.)). This lack of information to the public might finally lead to a rejection of the new Constitution for Europe, leaving the Treaty of Nice and all its flaws regarding the decision-making process in operation.
3.10 Implications of the reviews for the study

From the findings of the above reviews it was concluded that there might be a window of opportunity to introduce a common minimum guideline on antenatal care in the EU. However, it was perceived that it would be neither useful, nor appropriate to ignore the work and expertise that has already been invested to set up national guidelines in the member states. For this, a model was needed to synthesise the existing national guidelines, acknowledging the fact that special care needs to be taken not to exclude less wealthy states or states with many citizens to care for from the potential benefits of a common minimum guideline. In addition, only measures should be considered, which are acceptable to the majority of member states. This led to the idea that the models for decision-making within the EU might provide a good basis for this. Therefore, the treaties establishing the EU, as well as documents commenting on these structures, were analysed with regard to their properties which would inform the study.

The analysis of the current state of the art with regard to decision-making in the Council of Europe revealed that a voting system by qualified majorities is the basic principle for decision-making. At present, the Treaty of Nice defined qualified majorities. However, the review gave also evidence of serious efforts to change the decision-making structures towards a faster, more flexible and more transparent system. What decision-making might be from 1 November 2009 on is specified in the proposed Constitution for Europe. Both systems were assessed in depth for their properties regarding the subject of the thesis.

As a reaction to the weaknesses of the current system, the proposed Constitution seeks to improve the functionality of decision-making. The proposed voting system considers the population sizes of the member states and therefore constitutes a new system of qualified majority voting. The suggested system is called ‘double majorities’ and means that the majority of the member states must vote in favour of a decision and that they must represent the majority of the population of the Union. However, instead of mathematically defining a majority as a subset of a group, which is more than half of the entire group (Wikipedia 2006c), the Constitution goes only half way. For a decision under the Constitution for Europe, at least 55%
of member states, including at least 15 of them, and at least 65% of the Union’s population have to be in favour of a decision. In this context it is interesting to note that 15 members currently represent 60% of the States, which leads the 55%-rule ad absurdum. However, after the critical review, it was concluded that the qualified majority voting under the new Constitution for Europe according to ‘double majorities’ is worth exploration. It was decided that a more easily applicable model of ‘double majorities’ is needed for use on a broad basis.

In accordance with basic mathematics, a majority is defined as being more than 50%, and a double majority is defined as being a majority of votes according to two separate criteria (Wikipedia 2006c + d). For the study, ‘double majorities’ is therefore defined as more than 50% of the member states, and more than 50% of the inhabitants of the European Union.

To finally answer the question of whether a common EU minimum guideline on antenatal care is suitable, and what its content should be, the findings from the survey need to be considered. Also the question of whether such a guideline is financially possible for the less wealthy member states needs further exploration. The above analysis, complemented by texts on research methods, led to the detailed research plan. This will be described in the following section on the methods for the study presented in the thesis.

In the following section, the methods to reach the objectives of Phases 2 and 3 of the study are addressed. The first objective is to find out whether the basic principle for decision-making on the basis of ‘double majorities’ is suitable for making decisions on antenatal care, when national guidelines already exist. The second objective is to develop a framework according to which the existing national guidelines can be used to efficiently setting up a common minimum guideline, without compromising its quality.

For this, the hypothesis that a guideline based on ‘double majorities’ of all national guidelines on antenatal care in the EU contains the same recommendations as a guideline based on the principles of evidence-based medicine needs to be tested. To achieve this, the research plan was devised:
• develop and conduct a survey on the content of national guidelines on antenatal care in all member states of the EU.

• extract all measures which are recommended by at least 50% of the member states and apply to at least 50% of the inhabitants of the EU.

• compare the measures recommended by double majorities to the measures recommended by the most comprehensive evidence-based guideline on antenatal care to date.

To find out whether such a guideline is achievable, exploration is needed as to whether a common minimum guideline based on the findings from the aforementioned investigations would mean insurmountable hardships for the less wealthy member states of the EU.

The following chapter therefore specifies the methods used to shed light onto these areas. Before this, the ethical implications of the collection of primary data, as well as of the findings of the study are discussed. After this, detailed reference is made to the development of the survey tool, its properties, and how the survey was finally conducted. How the findings will be analysed and used is also made clear. The second part of the chapter focuses on the methodology for the in-depth appraisal of the two national guidelines.

Part 3 - The research plan

3.11 Ethical considerations

As for all research projects, ethical aspects also have to be considered for this study. Although at first sight only the ethical implications of the survey might apply, it was considered necessary to address also the ethical implications of screening procedures and the potential ethical challenges when setting up a common minimum guideline for the EU. This section therefore ties into the discussion of the ethical implications of guidelines, which were critically analysed in sections 2.4.1 and 2.4.2 of the thesis.
Before exploring the remaining ethical implications in greater detail, it has to be noted that it is the aim of this work to develop the theoretical background on antenatal care on the European level further. This will improve the EU-wide acceptance of widely recommended methods, optimising the use of limited resources. It is clearly aimed at improving services for pregnant women. Hence, the thesis itself intends to take a genuinely ethical approach. Moreover, the thesis is in line with the European Union's general strategy to coordinate the health policies and programmes of the member states (Treaty of Maastricht 1992), and to complement national health policies in order to improve public health, prevent human illness and diseases and to obviate sources of danger to human health (Treaty of Amsterdam 1997). Especially in an extended version of the Union's public health programme for the years 2003 to 2008, it was highlighted that information and knowledge should be improved with a view to promoting public health and health systems by developing a system for collecting, analysing and evaluating knowledge (Commission of the European Communities 1998, European Parliament and the Council of Europe 2002, Watson 2001). The planned overview and comparison of the national guidelines of the member states enhances transparency and therefore contributes to reaching these aims for antenatal care.

Ethical implications of the research process

Although the topic under investigation is antenatal care, i.e. the care for pregnant women, it does not directly involve any vulnerable persons, such as pregnant women or neonates. Moreover, no subjects are involved in the study at all, although contact is made with representatives of governments and professional organisations. Health and professional administrators are accessed, but asked to fulfil part of their normal tasks, one of which is the distribution of guidelines. These administrators are not used as research subjects, and they are not asked for any personal data, or their opinion. The persons addressed therefore fulfil none of the criteria, which are specified in the Declaration of Helsinki, e.g. those who are economically disadvantaged, those who cannot give or refuse consent for themselves, or who may be subject to giving consent under duress (World Medical Association 2000).
As this study does not research on human subjects or animals, the classical criteria for involving research ethics committees are not fulfilled (Neuberger 1992: 22, University of Surrey 2001). Although the Declaration of Helsinki does not apply for the same reasons, it provides valuable input about what aspects need to be considered irrespective of the direct involvement of human subjects (World Medical Association 2000). When developing the questionnaire, care was taken to inform potential participants about the aims and methods of the survey, institutional affiliations of the investigator, the anticipated benefits of the study and the discomfort it may entail. Finally, the survey questionnaires are of a non-offensive and non-distrressing nature, two of the main ethical issues that need to be addressed when using questionnaires (University of Surrey 2001).

Another reason for not seeking major ethical clearance is that this work is based on already published material, but does not involve access to medical records (University of Surrey 2001). No confidential or personal data is accessed, processed or stored. Instead, guidelines are used, which were specifically developed for publication and distribution.

However, although no formal ethical clearance from an ethics committee was sought, care was taken to fulfil the requirements of fundamental research ethics. One of these basic principles is that research must, for example, conform to generally accepted scientific principles and be based on a thorough knowledge of the scientific literature and other relevant sources of information (World Medical Association 2000). The thorough knowledge of the field and the research problem has already been demonstrated in Chapter 2 of the thesis, as well as in Parts 1 and 2 of the current chapter. To ensure the scientific quality of the research, the project was designed to comply with the charters on good research practice of the University of Surrey/United Kingdom and of the University of Bielefeld/Germany. With this, the principles and precautions of two research institutions in member states of the EU were considered. In addition, the project was closely supervised by experienced researchers from the UK, Germany and The Netherlands. Due to this close and also multi-national supervision, major and minor flaws of the study could be detected and eliminated at an early stage. Special care
was taken to ensure the validity of the methods, as this is a fundamental criterion that makes research ethical (Neuberger 1992: 22).

Ethical implications of screening procedures

According to the UK National Screening Committee (National Screening Committee (n.d.)), screening is defined as follows:

"Screening is a public health service in which members of a defined population, who do not necessarily perceive they are at risk of, or are already affected by a disease or its complications, are asked a question or offered a test, to identify those individuals who are more likely to be helped than harmed by further tests or treatment to reduce the risk of a disease or its complications."

Following to this definition it is highlighted that screening has important ethical differences from clinical practice, as it is targeting apparently healthy people. This is especially important if there are risks involved. It was therefore highlighted that people should have realistic expectations of what a screening programme can deliver (National Screening Committee (n.d.)).

The United States Commission on Chronic Illness Conference (cited by Bouvier et al. 1995: 17) defines screening as:

"...the presumptive identification of unrecognised disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who apparently have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment."

According to the above definitions, screening intends to reduce complications that arise from certain illnesses or conditions, including the reduction of mortality from those illnesses (Bouvier et al. 1995: 15). Another aim of screening is to decrease the overall incidence of an illness in a certain population. From these intentions of screening, questions arise.
When complications of conditions should be avoided by detecting a possible disease at an early stage, how should a test be judged, which screens for a condition for which no cure exists? As an example, genetic disorders, such as Down's syndrome might be named. In such cases, neither a difference is made by an early detection during pregnancy, nor can the prevalence be reduced. There are only two possible reactions to the diagnosis of such conditions during pregnancy: either to induce abortion or to accept life with an affected child. Before screening for such conditions, the issue needs to be explored with the pregnant woman and her significant others. Even more so, if the screening itself carries risks.

This leads directly to another important problem of screening: what about invasive, dangerous and psychologically irritating procedures, especially when the prevalence of the condition to be screened for is low? Is it ethically justifiable to risk the life of an unaffected foetus, when the condition screened for is either lethal, extremely rare, or cannot be treated despite its early diagnosis? All these issues need consideration before a decision is made about introducing a screening test for routine practice, for a selected publication, or not offering it at all.

One of the most important ethical issues in screening is that the quality of the screening test should be high enough to minimise the number of false positive and false negative results, which can not be avoided completely. The UK National Screening Committee commented on this that although screening has the potential to save lives or improve quality of life through early diagnosis of serious conditions, it is not a foolproof process, and cannot offer a guarantee of protection (National Screening Committee (n.d.)). It was also highlighted that "In any screening programme, there is an irreducible minimum of false positive results (wrongly reported as having the condition) and false negative results (wrongly reported as not having the condition)." However, positive results from tests with a low predictive value and low specificity are usually followed by further, more invasive tests, if these are available. This was also highlighted by the definition of screening by the United States Commission on Chronic Illness Conference, which states that a screening test is not intended to be diagnostic, and that persons with positive or suspicious findings must be referred to their physicians for diagnosis
and treatment (United States Commission on Chronic Illness Conference cited by Bouvier et al. 1995: 17). A good example for this is the triple-test, a test for three markers in maternal blood and a subsequent statistical calculation of risk for e.g. Down’s syndrome. If the result passes any subjective threshold of acceptable statistical risk, the procedure might be followed by an amniocentesis to assess the foetal chromosomes from a sample of amniotic fluid.

However, the following consequences of an initial positive test result can be manifold. By the positive test result from the triple-test, fears are caused, which can disrupt bonding between the pregnant woman and the foetus. This affects both the affected as well as the unaffected foetus. In addition to that, a positive result from the initial screening test will lead to an invasive procedure, which carries additional risks, such as miscarriage. This again applies to an affected as well as to an unaffected foetus. Safety in general, as well as the safety of measures applied on pregnant women is therefore a major issue in the ICM international code of ethics for midwives (International Confederation of Midwives 1999).

Another scenario might be that no further tests are performed, e.g. as no such test is available or wanted, and that it is therefore not possible to exclude any false positives. This might lead to the induction of abortion although the foetus is unaffected by the condition screened for. However, this goes beyond the screening as performed for routine antenatal care for healthy women. The ethical as well as legal implications of prenatal diagnosis are widely discussed in the literature and are far too complex and specific to be discussed in the context of the thesis (Deutsche Gesellschaft für Gynäkologie und Geburtshilfe 2002). However, the thesis will contribute to a more critical use of some of the tests for routine antenatal care.

An issue not to be ignored is the reactions of the pregnant woman’s social environment when she either opts against screening itself, or against treatment/abortion, after a certain condition has been diagnosed. Opting for a probably handicapped child might lead to stigmatisation and a lack of social as well as financial support. The mother/the parents run the risk of being blamed for their decision and are prone to social isolation and disadvantages. It has to be considered also that people in different societies and social circumstances tend to react
differently to such decisions and procedures. What is culturally acceptable to some might not be acceptable to others. Religious aspects also play an important role. Therefore, screening in the antenatal period has to be sensitive to cultural and personal circumstances.

Another way of dealing with screening tests is to respect or support the autonomy of the pregnant woman, as is suggested by the ICM international code of ethics for midwives (International Confederation of Midwives 1999). This code of ethics clearly states that midwives should respect a woman's informed right of choice and promote the woman's acceptance of responsibility for the outcomes of her choices (International Confederation of Midwives 1999). This would mean introducing a screening test for a population, but leaving the final decision about what is acceptable with the pregnant woman herself. This is also acknowledged in the definition of screening of the UK National Screening Committee, with the statement that that the candidate for screening should be offered a test, but not tested without being asked (National Screening Committee (n.d.)).

However, the claim to enable informed decisions calls for professional preparation of the person to be screened. Before deciding about a screening procedure, pregnant women need to be aware of the implications of the procedure and of the possible following consequences from positive as well as negative test results. Also the issue of false negative and false positive results needs to be discussed. When opting for or against screening, the personal conviction of the person to be screened should be respected. The wish for a certain screening, possible fears of the screening itself or of the condition for which is screened need to be explored. It depends much on personal values, if feelings of relief and gratitude result from a positive as well as a negative test result, or whether resentment occurs after finding out about a condition, which the screened person did not want to know about. Enabling informed decisions therefore requires knowledge and skills, as well as the resources to discuss the relevant issues. However, this becomes more important with the severity of the condition and possible consequences following a screening test. This applies more to tests from the realm of prenatal diagnosis, rather than to screening tests performed for the routine care of healthy pregnant
women. The extensive procedures described above would be inappropriate e.g. when screening the mother for iron deficiency, which has minor or no ethical implications at all.

Consequences for a common minimum guideline on the EU-level

Another important aspect to consider is the position that it is the task of health professionals to treat individuals according to the latest evidence, but not necessarily to have entire populations in sight (Hunter 1997: 76, citing Sheldon & Long). However, advances in health care have their costs, and only by the strategic, intelligent and effective allocation of resources is it possible to keep the current system of social justice and equal treatment. It is therefore absolutely necessary to avoid an arbitrary and inconsistent allocation of limited resources (Hunter 1997: 62). Here, high-quality and broadly approved guidelines help to plan strategically and to allocate resources most effectively. How difficult it is to set up guidelines of high quality, which are also sensitive to values, culture, ethics, and other decision-modifying factors has been discussed at an earlier stage of the thesis. However, at the end of the extensive analysis it was concluded that the need for sensitivity for specific national situations do not stand against a common minimum guideline on antenatal care in the EU. In contrast to that, a common minimum guideline lays a sound basis for the care of all pregnant women, which can subsequently topped up according to national, regional or even individual preferences.

3.12 The survey

There were indications that guidelines on antenatal care exist in most of the member states of the EU, as has been discussed in the corresponding section in the literature review (Backe & Buhag 1994, Blondel et al. 1985, Hemminki & Blondel 2001, World Health Organization 1987). However, the respective national guidelines are not available outside the countries and could not be located by means of a literature review using the databases of the medical and related professions, such as midwifery. If it was possible to trace them, e.g. via a national society of midwives or the society of obstetricians, the documents were published in the
national languages only. As an alternative, part of the content of the guidelines could be extracted from studies on certain aspects of care (Backe & Buhaug 1994, Bundesausschuss der Ärzte und Krankenkassen 2003, Den Almindelige Danske Jordemoderforening 1998, Gobena I Tricas et al. 1996 [Spanish], Kristensen et al. 1995, Ministry of Health and Social Affairs 1984 [Finnish], cited by Hemminki & Gissler 1993, The Ministry of Health 1997). The existing guidelines are therefore neither reasonably accessible for international researchers or decision-makers, nor is their content known to health professionals in the other EU states.

On account of these problems, it was decided to conduct a survey to get a clear picture of whether guidelines on antenatal care already exist in the member states of the EU, and what their content is. The answer to this question was planned to show whether guidelines are a common and accepted way to guide practice in the member states. Moreover, the content of the national guidelines was established in order to use it as the basis of a common minimum guideline at a later stage. It was assumed that using the national recommendations as the basis would enhance the acceptability of the common minimum guideline.

To gain a complete overview and enhance the comparability of data, a structured questionnaire was used. However, as no such survey had been conducted on this topic previously, a questionnaire needed to be developed from which guidelines could be extracted and commented on directly in the member states. As the survey was planned to include only guidelines for the routine care for normal pregnancies, antenatal care was defined as the baseline clinical care of all pregnancies of a healthy woman with an uncomplicated singleton pregnancy (National Collaborating Centre for Women’s and Children’s Health 2003). Baseline clinical care was limited to the individual recommendations of what test should be performed and at which stage of pregnancy. Moreover, only antenatal care up to the estimated date of delivery was considered. The research question was: What national guidelines exist and what tests are recommended in each member state for antenatal care?
3.12.1 Designing the questionnaire

To extract the individual recommendations from guidelines, tables with tick-boxes were used. All 37 routine tests identified from the critical review of the literature were listed in the tables, including physical tests, such as the measurement of blood pressure, technical tests, such as abdominal or vaginal ultrasound, and laboratory tests, such as taking blood samples for an HIV test. As all tests identified by the literature review were included, it was made explicit that this did not imply a value judgement as to whether these tests were effective and/or should be offered. To facilitate the extraction of data from the original guidelines, tests were organised under the aforementioned sub-headings: physical tests, technical tests and laboratory tests. In each category, space was provided to record tests additionally recommended in each country. The complete set of tests asked for is presented in Appendix 2, where the final questionnaire is provided together with the covering letter and consent form.

Another important feature in the tables introduced the time factor. Each test had to be located on a time line, which represents gestational weeks. By this it could be made clear if a test should be performed at a specific gestational week, or if there is a time span, within which a test should be performed. However, when the tables were used for gathering the data on national guidelines during the literature review, it was found that they needed to be refined and tailored for extracting the relevant data. Reporting tests to be performed at the first visit during pregnancy required the introduction of the section “1st visit” in addition to the time line representing weeks of pregnancy. Starting with the 6th gestational week was not appropriate, as this would have forced the respondent to determine a virtual first visit on the timeline.

In addition to the tables to extract the recommendations, questions were introduced to document which country is represented by the respondent and for which organisation he or she works. The next question clarifies whether there is a national guideline, or not. If there is one, the issuing body should be named and the respondents were asked to provide the full reference of it. No further questions were introduced, as the information extracted from this questionnaire was judged as adequate with regard to the aims of the survey. It was also
concluded that any additional questions might have had a negative influence on the response rate by introducing more dimensions and requiring more time to complete the questionnaire.

Finally, the survey tool was completed with instructions on how to use it, an accompanying letter stating the purpose of the study, and a consent form. The latter were required and designed in order to fulfil the University of Surrey’s Ethical Guidelines for Teaching and Research (University of Surrey 2001). The format and content of the survey and the associated project procedures did not require a formal submission to the Advisory Committee on Ethics. The question of whether formal ethical clearance was required was discussed and decided upon in consultation with the secretary of the University’s Advisory Committee on Ethics and the PhD supervisors.

3.12.2 Pilot study and properties of the questionnaire

In order to optimise the questionnaire’s user-friendliness, validity and reliability, it was piloted and revised before its distribution. In accordance with the statistician of the European Institute of Health and Medical Sciences it was decided that it was neither necessary, nor appropriate to calculate inter- or intra-rater reliabilities by means of statistics according to the purposes of the study. A test-and re-test run with the questionnaire after four weeks, together with a comparison of the answers against the original guideline and the thorough and detailed description of findings was regarded as appropriate. Moreover, this procedure provided the opportunity to comment in depth on the questionnaire’s advantages and disadvantages. To achieve this, the questionnaire was tested and re-tested against the German guideline (Bundesausschuss der Ärzte und Krankenkassen 2003). The German guideline was selected for the pilot, as it avoided any language difficulties for the German test raters. As the final questionnaire will be also used in English, but on guidelines published in the national language of the raters, it was regarded as appropriate to apply the same principle for the pilot run.
The test raters were:

- one professor of gynaecology and obstetrics,
- one scientific assistant in nursing sciences,
- one midwife, and
- the person conducting the study.

These test raters were chosen according to their professional specialty together with their level of professional education, which represents the target groups for the survey. Sending the questionnaires to Ministries of Health, and the societies of midwives and the societies of obstetricians, it was assumed that the following persons could be involved in filling in the questionnaires: health professionals from varying specialties including physicians, but also nurses or other health care personnel. In addition to them, also obstetricians and midwives as experts in the field could be the respondents, but representing different professional grades. Therefore, the persons piloting the questionnaire were selected to represent the target groups. They were planned to include a health professional without any knowledge of obstetrics, i.e. the scientific assistant in nursing sciences, a staff midwife as the specialist for normal pregnancies, and a professor of obstetrics to represent the highest level of specialisation. In addition to their profession and level of expertise, the raters needed to be able to work in German and English, and had to be willing to give feedback on the questionnaire.

With regard to user-friendliness and time needed, the test raters found that the questionnaire was easily understandable and applicable. Although the test raters gave generally positive feedback, the example given on the instruction sheet was made more explicit and clear, and the sentence limiting the answers to normal uncomplicated pregnancies was re-worded and highlighted. On average, the time needed to complete the questionnaire was calculated as 15 minutes. The professor of gynaecology and obstetrics needed the least time. The scientific assistant from the nursing sciences needed the longest, as he was not familiar with antenatal care and the corresponding terminology. However, although the scientific assistant from a discipline other than obstetrics needed longer and found it more difficult to fill the questionnaire in, this rater was also able to extract the German guideline correctly, when comparing it to the
original guideline. This finding was important, as the questionnaire needed to be applicable by persons from diverse disciplines, and particularly such without expert knowledge of antenatal care. It was also necessary that the recommendations from the guidelines were not altered by the level of expertise of the respondent.

Finally, all test raters found no unclear or ambiguous items. Moreover, they could not identify any important information within the German guideline, which could not be reported in the questionnaire. To demonstrate that this was also the case for a guideline from another country, the questionnaire was also used on the guideline from England and Wales by the principal investigator of the study (National Collaborating Centre for Women’s and Children’s Health 2003). For language reasons, the English guideline was selected. In addition to that, the guidelines from Germany and England and Wales are based on different models for setting up guidelines. The German one is based on expert opinion, making no referral as to what sources of evidence have been used and how they have been integrated. In contrast to this, the guideline from England and Wales has been developed according to a clearly specified scheme of identifying, selecting and integrating scientific evidence, therefore representing the evidence-based model. This step of testing the questionnaire on the guideline of another member state based on another paradigm was introduced, as the final questionnaire needed to be able to extract recommendations of guidelines from different countries based on different paradigms. However, as the information from the guideline from England and Wales also could be extracted, the questionnaire was found to have the property to model the recommendations adequately from guidelines of different member states and based on different methods of setting them up.

3.12.3 Intra-rater reliability

In addition to user-friendliness and validity, the intra-rater reliability of the questionnaire needed to be proven. Answers were judged to be identical when they reported the same tests and the same number of these tests. Their distribution was accepted to be identical when the
frequency was the same, irrespective of the exact gestational week, as this depends on when care commences. Responses were counted as similar when they reported the same tests, but with a different number of repetitions for them. In order to ascertain whether the answers of the test raters were identical to the original, they were compared to the recommendations in the guideline. When comparing the tests and re-tests of the raters according to the aforementioned criteria, the following was found:

- In the two questionnaires filled in by the professor of gynaecology and obstetrics, all items were coded identically (100%). Obviously, this rater's interpretation of the guideline as well as his reporting of it was stable over the four-week period. Also the recommendations of the guideline were extracted correctly.

- The scientific assistant from the nursing sciences also coded identically at both instances (100%). However, as he had no professional view on the subject, he extracted the recommendations from the guideline without clinical judgement. This led to a specific problem. He overlooked a page, on which the blood tests were listed that should be performed at the first visit. The most likely reason for this problem is probably the misleading order of subjects within the German guideline. In this, a section on how to detect and monitor high risk pregnancies is inserted between the general section on normal pregnancies and the section on laboratory tests for normal pregnancies. The raters with a background in obstetrics found this page, as they knew that information on laboratory tests was missing. The scientific assistant did not have this knowledge and therefore did not look especially for this page. As a consequence, the scientific assistant's answers were reliable, but not complete.

- The midwife coded 32 of 37 tests identically (86.5%). Differences occurred when the midwife used her clinical judgement in the re-test, but relied entirely on what was written in the guideline in the first run. The problem occurred when a test was recommended by the guideline for each visit, but starts to make sense in clinical terms later on. Palpating the foetal position at each visit is an example for this from the
German guideline. However, recommending a test to be performed at each visit while common clinical knowledge of midwives says that it starts to make sense only in the second half of pregnancy is a weakness of the guideline. The unreliable finding is therefore not a problem of the questionnaire, but due to an unclear or clinically questionable recommendation in the original guideline.

- 35 of 37 items (94.6%) were coded identically by the person conducting the study. As this person is a midwife, too, the same error occurred as described for the midwife above. On one occasion, the rater reported strictly what was written in the guideline, but used professional judgement in the second run.

Drawing these findings together, it was found that problems related to the intra-rater reliability seem to be due to ambiguous formulations and the misleading order of subjects in the guideline, rather than to weaknesses of the questionnaire. Errors occurred systematically, when a test was recommended to be conducted at each visit, but does not make sense clinically in the early weeks of pregnancy. With regard to this issue, the raters seemed to be in doubt whether they should report what was written in the guideline, or whether they should apply their clinical knowledge to the text. This problem could not be solved by altering the questionnaire. However, it has to be noted that the error occurred in both instances at the re-test, i.e. that clinical knowledge started to compromise the one-to-one extraction of recommendations when the questionnaire was familiar to the raters - and the instructions sheet was not read a second time. Considering this in greater depth, it might be that the questionnaire models a phenomenon of clinical practice. It would be worthwhile to explore whether more experienced clinicians are more likely to deviate from guidelines than their inexperienced counterparts. However, for the questionnaire this finding means that its reliability is good, when the questionnaire is filled in for the first time.
3.12.4 Inter-rater reliability

In order to comment on the inter-rater reliability, all test questionnaires completed on the first occasion by the four raters were compared to each other and to the original guideline, in order to establish potential differences. A decision was made in favour of the responses provided in the first run, as the respondents in the survey will use the guideline only once, hence when seeing it for the first time. As a consequence, the responses need to have sufficient inter-rater reliability at the first application of the questionnaire, which was to be tested.

In this comparison it was found that the midwife reported a vaginal examination at each visit, which is not within the guideline, but common practice in Germany. As discussed in the section on intra-rater reliability, the scientific assistant overlooked one page containing laboratory tests. However, the major inconsistency between the raters was the different onset of care. Based on the rater’s idea of when the first visit takes place, the recommended weeks for visits changed. When care was theoretically started at week 6 instead of week 8 or 10, more visits and tests were reported, although they were based on the same frequency and schedule of care. Therefore, the gestational weeks at which tests are recorded should not be regarded as absolute measures. This applies at least to all measures recommended for each visit. There is a variance of two weeks according to time at which antenatal care was thought to commence.

This finding has to be interpreted in the light of the knowledge, that the gestational week to start with antenatal care is not necessarily specified in each guideline. In practice, the starting point represents entirely the decision of the woman about when to seek care. It should not be taken therefore as an absolute measure. Despite this, the frequency of visits can be regarded as reliable, with the gestational week showing a variance of two weeks. However, as the experts setting up the guideline were obviously aware of this problem, for specific measures which are recommended only once or at certain stages of a pregnancy, time spans are always given within which the test should be performed. As an example, in the German guideline abdominal ultrasound scans are recommended to take place between 9 and 12 gestational weeks, between 19 and 22 weeks, and again between weeks 29 and 32 (Bundesausschuss
The ultrasound examinations can therefore be performed within the normal course of antenatal appointments. However, it has still to be noted that the authors of the guideline were specific in this point, but recommend other tests, such as the measurement of maternal blood pressure, to be performed “every four weeks”.

On the basis of the findings from the pilot study and the tests for intra-rater in inter-rater reliability, the user-friendliness, and validity as well as the reliability of the final questionnaire was judged as appropriate for the planned survey.

3.12.5 Distribution of the questionnaire

As the institutions issuing national guidelines, Ministries of Health, as well as the societies of obstetrics and the societies of midwives were selected as the target groups for the survey. These organisations were identified through lists provided by the German Society of Gynaecology and Obstetrics, the European Board and College of Obstetrics and Gynaecology, the German Association of Midwives, and the International Confederation of Midwives. The Ministries of Health were identified via the internet. In addition to addressing the organisations mentioned above, the president of the European Board and College of Obstetrics and Gynaecology (EBCOG), Professor André van Assche, contributed to the study. He was prepared to contribute after the investigator presented the study and the survey questionnaire briefly to him at a conference of EBCOG in Athens in May 2004. As he considered the survey worth supporting, he sent a letter of reference and the survey material by e-mail to the presidents and the delegates of the national societies of obstetrics and gynaecology. Using the president of EBCOG as a distributor and referee for the study was seen as of key importance, as it was regarded as unlikely to being able to include data from all member states of the EU without a trusted person as an advocate. At the same conference, at which professor van Assche was approached, it was also possible to distribute three questionnaires to obstetricians from Sweden, the Netherlands and Greece.
As initially planned, 132 questionnaires were sent to the Ministries of Health, and to the national societies of obstetricians and midwives of the EU-15 States. However, as the first answers were obtained and a preliminary analysis was conducted, it seemed inappropriate to exclude the new member states from 1st May 2004, when the study would be published in 2005. Therefore, the survey was extended to the 10 accession countries. With these states being now fully integrated into the Union, it could not be justified that a recommendation for the entire EU could be based on the data of the pre-2004 states only. Including the accession states, the survey was finally conducted between May and August 2004 by surface and e-mail.

Considering all 25 member states of the EU together, a total of 155 questionnaires and 61 reminders were distributed. If after a reminder no response was obtained, individual health professionals were approached, who were well known beyond the borders of their countries, identified through their publications, or through personal contacts. As a whole population approach was chosen for the study, this was continued until there was at least one response from each member state.

With regard to potential compromise of findings, approaching individual health professionals might introduce bias. Interpreting guidelines in the light of one's own professional practice does have an influence on responses, which has been discussed for the midwife in the section on intra-rater reliability. However, the same error occurs when a specialist working for a society of obstetricians or a midwife from a society of midwives fills in the questionnaire. Hence, the error is similar across the institutions approached, but could not be alleviated. It can not be avoided that the thoughts and interpretations influence results when using a survey design. Using individual experts for the survey if no answer could be obtained from an institution was regarded as appropriate, as it does not compromise the reliability of results.

Another error that could be introduced by approaching individual health professionals could have been that they do not have the latest version of their national guideline, such as a respondent from an institution occupied with antenatal care might have. However, as all health professionals are accountable for their professional practice, this is unlikely.
3.12.6 Analysis of the survey data

The selection of questionnaires

For analysis, one questionnaire was selected per member state. If there was more than one answer from a country, the one from the most official source was used in case there were differences in the reported recommendations. As the most official source, the issuing body of the guideline was defined. In case of being issued by a multidisciplinary group, the Ministries were seen as the most official source, followed by the professional societies of obstetricians and midwives. Answers were judged to be identical when they reported the same tests and the same number of these tests. Their distribution was accepted to be identical when the frequency was the same, irrespective of the exact gestational week, as this depends on when care commences. Responses were counted as similar when they reported the same tests, but with a different number of repetitions of them. When the response from the most official source needed to be selected, the following ranking was used: a governmental body, a national professional society, an individual health professional.

In case of differences in the reported recommendations, an explanation should be found for them by the investigator. If differences can not be explained, clarification is to be sought from the respondents directly. Moreover, it should be assessed, what implications these differences might have. However, in case that no explanation was found, and the response could therefore not be included in the analysis, it was planned to consult the original guideline from which the recommendations were extracted. To enable this, the respondents were asked to provide the reference of their national guideline. If there were different recommendations for first and subsequent pregnancies, the care scheme for first pregnancies is considered.

After this selection procedure, the questionnaires were analysed with regard to how many countries recommended a test and as to how many people were affected by this. As reference numbers, the population sizes of the official statistics of the European Commission were used (Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften 2004). Numbers from the year 2000 were used, as they are the most up-to-date numbers that are based on
population census, rather than on estimations.

The development of a common minimum guideline

Based on the findings from the procedures described above, a common minimum guideline for the member states of the EU will be developed. To find a model to integrate the guidelines from the member states, the models for decision-making in the Council of Europe were explored (The European Union Constitution 2004, Treaty of Nice 2001). After the critical analyses of these models, which was discussed in the chapter "decision-making in the Council of Europe", it was concluded that at present there are serious efforts to change the decision-making structures towards a faster, more flexible and more transparent system (The European Union Constitution 2004). It was therefore decided, not to use the currently operated model of qualified majority votes, which is difficult to calculate and suffers from a lack of acceptance (SCADPlus (n.d.), Treaty of Nice 2001).

In contrast to this, the model as suggested in the proposed Constitution seeks to improve the functioning of decision-making (The European Union Constitution 2004), but still considers the population sizes of the member states. It therefore constitutes a new system of qualified majority voting. The suggested system was called 'double majorities' and means that the majority of the member states must vote in favour of a decision and that they must represent the majority of the population of the EU. However, instead of mathematically defining a majority as more than 50%, the Constitution was found to go only half the way.

Triggered by these problems, but also by its broad basis in the Council of Europe, voting according to the principle of double majorities was explored further. Finally, a more easily applicable model was needed to ensure sufficient transparency. In accordance with basic mathematics, a majority is therefore defined as being more than 50% (Wikipedia 2006c), and 'double majorities' refers to more than 50% of the member states, and more than 50% of the inhabitants of the European Union (Wikipedia 2006d). In order to be included in the common minimum guideline, a test needs to fulfil two criteria: being recommended by at least 50% of
the member states with a guideline, and applying to at least 50% of the inhabitants of these countries. This approach of ‘double majorities’ was chosen to take account of two principles of the EU, i.e. that each member state counts, and therefore should have a vote irrespective of its size, but also that the member states have responsibilities for different population sizes, which should be acknowledged when making decisions. Therefore the principle of 50% of member states and 50% of inhabitants appeared as a reasoned case to ensure that a common minimum guideline can be accepted and implemented by all member states.

With regard to defining double majorities in the mathematical sense it was found more than reassuring that Wolfgang Schüssel, the President of the Council of Europe from January to June 2006, recommended exactly this model of decision-making for the Constitution of Europe (Falksohn et al. 2006, Heil 2006). In June 2006, Schüssel pledged to hold a plebiscite throughout all member states to pass the common constitution for Europe. In this context Schüssel specified that the constitution should be accepted, if the majority of the European population and the majority of member states voted in favour of the constitution (Heil 2006).

The link between the Gross National Product and the number of tests

The next step was aimed at finding out whether there is a correlation between the Gross National Product/GNP and the number of tests recommended. For this, the figures of the World Bank for the year 2003 were used (World Bank 2004). However, the GNP was provided in US dollars and subsequently calculated in Euros, as this is a purely European study. In order to make reference to the different price levels in the countries, the same calculations were conducted using the figures of the official statistics of the European Commission for the year 2003 (Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften 2004). These were not used primarily, as for 2003 only prognoses were available and the purchasing power parities given instead of the GNP per capita in Euro.

For this part of the study, a null hypothesis was formulated as follows: That there would be no difference between the Gross National Product of each member state and the number of tests
recommended for antenatal care. This additional analysis was introduced, as it was assumed that a common minimum guideline could only be implemented, if it does not mean insurmountable financial hardship for the less wealthy member states of the European Union. For this it was regarded necessary to explore whether there is a relation between the financial status of a Member State and the number of tests recommended in the respective guideline.

National recommendations and published evidence

To explore the robustness of the common minimum guideline, its recommendations will be compared to the available evidence about the individual tests. A hypothesis is also used here: a guideline based on double majorities of all national guidelines contains the same recommendations as a guideline based on the current evidence-base. For comparison, the recommendations of the guideline from England and Wales are used (National Collaborating Centre for Women’s and Children’s Health 2003). These recommendations represent the most extensive critical literature review on the subject to date. All in all, 631 pieces of original publication and meta-analyses were cited and analysed by a multi-professional group of nine specialists and two consumers. The literature review was systematic, and the methods used for identifying, selecting, analysing and integrating the evidence were clearly specified. Where available and appropriate, economic data about cost-effectiveness was also introduced.

However, despite this extensive review, the guideline development group of the National Collaborating Centre could not find sufficient evidence to come to a clear recommendation for all measures without a final judgement of the group. Therefore, recommendations were classified from grades A to D, depending on the robustness of the underlying evidence. In addition to these grades, also the label "good practice point" was introduced for recommendations based on the view of the guideline development group. As this grading system for guidelines does not demonstrate the lack of skill of the development group, but mirrors the current state of knowledge, it was concluded that this report is appropriate as a comparison for the findings of the survey. Most useful is in this respect that the guideline for England and Wales provides excellent transparency so that a skilled critical reader is able to
follow the argument, access the underlying literature, and draw his or her own conclusions (National Collaborating Centre for Women’s and Children’s Health 2003). As evidence is rarely clear and unambiguous and guidelines represent decisions based on the best available evidence, it was decided to use the transparent decisions and underlying evidence of the UK’s National Collaborating Centre for Women’s and Children’s Health for comparison.

Another reason for using the guidelines from England and Wales as a comparator follows the argument of the WHO, which claimed that many of the studies underlying the WHO basic model for antenatal care have been carried out in different settings and countries, and that generalising from these results may be very difficult (World Health Organization 2003).

However, as the WHO recommends its basic model also for the European region, it was critically assessed in a previous chapter entitled “the WHO new model of antenatal care”. Only one of the reasons for considering the model of the WHO as inappropriate for the EU was that it had not been tested in any single European state, but in Cuba, Thailand, Saudi Arabia and Argentina (Villar et al. 2001). The guideline from England and Wales is a better comparator, as its recommendations were made on the background of a genuine member state of the EU.

The recommended frequency of tests

The last analysis of the survey data will complete the list of tests recommended for the routine care of healthy pregnant women in the EU and extract the frequencies with which individual tests are recommended during each pregnancy.

3.13 In-depth appraisal of two national guidelines

Another important issue that evolved from the review of the state of the art was the need to distinguish clearly between recommendations and guidelines. This is clarified and discussed in a concept analysis in the chapter on guidelines and their quality, and especially in the section “evaluation of recommendations”. In addition to this, other indicators of the quality of guidelines are explored. After clarifying the operational definitions as well as the underlying concepts, and
after the critical appraisal of the identified instruments, it was concluded that there is a useful instrument for the purposes of the study (The AGREE Collaboration 2001).

Objectives of the in-depth appraisal

The objective of Phase 1 of the study is to critically appraise the relation between evidence-based guidelines and guidelines based on expert opinion for the defined public health task 'antenatal care'. To achieve this, the paradigms underlying and the methods for setting up evidence-based guidelines and guidelines based on expert opinion have already been theoretically discerned and analysed in the previous chapter. Within this, the nature of evidence was critically discussed, and the differences and similarities between scientifically generated evidence and evidence referred to by experts were highlighted.

From the critical review on guidelines, their underlying evidence and the challenges of decision-making it was found that the appraisal of the scientific quality of a guideline's recommendations is an important step in its evaluation, though not the only one. Other indicators, such as the methodological and structural quality of the guideline itself also play a role. However, it was also concluded that there is still the missing link between the overall quality of a guideline, and the quality of its recommendations. Although the AGREE-Collaboration claims that there is a relation between the overall quality of a guideline and the quality of its recommendations (The AGREE Collaboration 2001), clarity is missing as to whether there is a stable relationship between the two, and to what extent.

Selection of an instrument for appraising two national guidelines

Answering this question for guidelines on antenatal care is regarded as one of the cornerstones with regard to the completion of Phase 1 of this study, and therefore for its contribution to the body of knowledge. In order to achieve this, deeper insights are needed into the implications that different methods for setting up guidelines have on their overall quality. Two guidelines based on entirely different paradigms will be analysed in depth for their methodological and structural quality, using the instrument of the AGREE-Collaboration (The
AGREE Collaboration 2001). Why this instrument was selected has been already discussed extensively in the sections entitled "Instruments for the systematic appraisal of guidelines ", and "Appraisal of guidelines for research and evaluation in Europe" of the thesis. In these sections, the findings from a systematic review of the literature are presented together with the critical analysis of the identified instruments (The AGREE collaboration 2001, ÄZQ Zentralstelle der Deutschen Ärztenschaft zur Qualitätssicherung in der Medizin 1999, Bundesärztekammer & Kassenärztliche Bundesvereinigung 1997, Cluzeau et al. 1999, Ollenschläger et al. 1998, Scottish Intercollegiate Guidelines Network 2002).

In brief, it was found that the instrument of the AGREE-Collaboration was specifically designed to assess the overall quality of guidelines set up by multidisciplinary teams, and is applicable to guidelines from any disease area, including also those on diagnosis, health promotion, treatment or interventions. Moreover, this is the first instrument with proven validity and reliability for the appraisal of guidelines from different countries. All these factors qualified the AGREE-instrument to be used for appraising guidelines on antenatal care for the purposes of this study. Finally, it was reassuring to find that the AGREE-instrument was also judged by the Council of Europe as the most promising instrument for the evaluation of guidelines in different countries that were set up by different groups of experts and recommends the use of the instrument (Europarat 2001: 17ff, Europarat 2001: 30 + 44).

In addition to these advantages, the instrument has been reported to be in use by well respected national and international organisations occupied with quality assurance in medicine, such as the National Institute for Clinical Excellence in the UK, the National Federation of Cancer Centres in France, The Agency for Quality in Medicine in Germany, the Scottish Intercollegiate Guidelines Network and also the World Health Organization (Cluzeau 2004). Using this instrument will therefore enhance the international comparability of guideline quality, as well as improve the consistency and quality of reporting them.
The appraisers

Tests of the AGREE instrument's reliability demonstrated that the number of raters appraising a guideline affects the reliability of the intra-class correlations. Intra-class correlations substantially improved with the number of raters. To achieve acceptable Cronbach α values for the individual quality domains as specified by the instrument, four raters are recommended (Cluzeau 2004, The AGREE Collaboration 2001 + 2003). In order to ensure the reliability of the appraisal, the recommended number of four raters will be used.

However, although the AGREE-instrument is used for the appraisal of the guidelines, the method for this part of the study goes beyond the simple appraisal of two national guidelines using this tool. As was discussed in section 3.3.4 of the thesis, the AGREE-collaboration reported the questionable internal consistency for the domain ‘clarity and presentation’. Unfortunately, this was not explained. To shed light on this, the method for the in-depth appraisal of the two national guidelines was amended. The appraisers using the AGREE-instrument were treated as a panel of experts, which was selected purposefully in order to elicit informed judgements about the quality of the two national guidelines. Special care was taken to relate the interpretations of the experts back to the original guidelines in order to shed light onto their individual professional views.

Although the estimated time needed for appraising a guideline is 1.5 hours (Cluzeau 2004), the professionals who had already piloted the survey questionnaire were willing to support the study again. This was regarded as a good mixture of skills and knowledge of the topic. All appraisers had demonstrated that they are able to work in German and English. Moreover, they had proven that they can interpret guidelines, can now be more easily instructed about the difference between a guideline's recommendations and the other criteria to be assessed. However, most important for their selection was that the appraisers cover different disciplines, but are all involved either with using, or the process of developing or appraising guidelines. Again, one person has no special knowledge of antenatal care.
The selected appraisers therefore are:

- one professor of gynaecology and obstetrics,
- one scientific assistant in nursing sciences,
- one midwife, and
- the person conducting the study.

Guidelines selected for the in-depth appraisal

As guidelines for the in-depth appraisal, the evidence-based guideline from England and Wales (National Collaborating Centre for Women's and Children's Health 2003) and the expert-opinion based guideline from Germany (Bundesausschuss der Ärzte und Krankenkassen 2003) will be used. These two guidelines have been selected, as the guideline for England and Wales is a document of a unique and extensive multidisciplinary process according to the principles of evidence-based care and excellence in guideline development. Just published in October 2003, it represents the most recent approach towards excellence in antenatal care for the normal pregnant woman.

In contrast to that, the German guideline represents a legally binding guideline which has been used for several years now, based on less formal criteria, agreed on and developed by the Federal Joint Committee of the self-governing body of physicians and health insurance funds (Bundesausschuss der Ärzte und Krankenkassen 2003). Therefore, the guidelines from the England and Wales and Germany are based on completely different approaches to the same problem, hence offering best opportunities to comment on strengths and weaknesses. Moreover, both guidelines were either set up, or amended in 2003, which makes them also comparable with regard to the state of the evidence at the time of issuing the guidelines.

The overall quality of the guidelines and the quality of their recommendations

At the outset of this paragraph, the objective of this part of the study has been specified as to add knowledge about the link between the overall quality of a guideline, and the quality of its
recommendations. After the in-depth appraisal of the overall quality of the guidelines from England and Wales and Germany, it is finally necessary to link these findings to the quality of the recommendations of both guidelines. However, the developers of the AGREE-instrument state clearly that their instrument is neither intended to assess the clinical content of a guideline, nor the evidence that underpins the recommendations (The AGREE Collaboration 2003). This means that there is still no tool to assess this relation, and the crucial question remains of whether good guidelines make good recommendations.

Due to this lack of a suitable tool, it was decided to additionally compare the recommendations from the two national guidelines against each other. However, it has to be noted that due to its outstanding qualities the guideline for England and Wales is used as a comparator for the findings from the survey. Despite this, the guideline from England and Wales is not used as an untouchable gold standard. As has been demonstrated in Table 2.8, the recommendations of the guideline for England and Wales are graded from A to D depending on the robustness of the underlying evidence. Therefore, the critical discussion of the recommendations from this guideline will be used to compare the German guideline. In case of the underlying evidence being weak or ambiguous, it is acknowledged that different recommendations have to be accepted without necessarily judging one as more appropriate.

With the approach demonstrated above, it is hoped to

- answer the question of whether different approaches to set up guidelines on antenatal care lead to similar recommendations, or not.
- demonstrate the difficulties involved with the assessment of the overall quality of currently available national guidelines on antenatal care.
- provide insights into the relation between the recommendations of a guideline and its overall quality according to internationally accepted criteria.
- shed light onto the question of whether scientific evidence and expert knowledge on antenatal care are similar or congruent on the highest national level, i.e. whether the guideline based on expert opinion contains the same recommendations as the one set up on the basis of scientific evidence.
Chapter 4 – Results

4.1 Results from the survey

4.1.1 Response rates

Although the aim of obtaining at least one answer from each member state of the EU was achieved, not all questionnaires sent out led to a response. A total of 36 responses to the survey were obtained from all 25 member states of the European Union. This represents an overall response rate of 23.22% of all 155 addressed institutions and individuals. As the questionnaires were distributed separately to the EU-15 States and the States joining the EU on 1 May 2004, separate response rates were calculated for these groups. Twenty-three answers from 132 institutions and individuals were received from the EU-15 countries, which is equivalent to a response rate of 18.18% in the EU-15 group. Twelve responses were obtained from the 10 new member states after sending 23 letters. This means a response rate of 52.17% in this group. In Table 4.1, all completed questionnaires are listed per member state and with the respective respondent. Thirteen questionnaires (36.11%) were filled in by obstetricians, with one Czech obstetrician working for the Institute for the Care of Mother and Child. Nine questionnaires (25%) were completed by personnel from Ministries of Health and midwives respectively, and another five (13.89%) by representatives of other institutions. Seven member states provided more than one response and are described in more detail below.
Table 4.1: Completed questionnaires

<table>
<thead>
<tr>
<th>Respondent</th>
<th>A</th>
<th>B</th>
<th>C2</th>
<th>DK</th>
<th>EST</th>
<th>F</th>
<th>GER</th>
<th>GR</th>
<th>H</th>
<th>I</th>
<th>IE</th>
<th>LT</th>
<th>LUX</th>
<th>LV</th>
<th>M</th>
<th>NL</th>
<th>P</th>
<th>PL</th>
<th>SE</th>
<th>SK</th>
<th>SI</th>
<th>SL</th>
<th>SW</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ministry of Health</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ministry of Social Affairs</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Society of obstetricians</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obstetrician</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Society of midwives</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Independent) midwife</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Council / Centre for Research &amp; Development</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institute for the care of mother &amp; child</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institute of Public Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University professor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using professional contacts to distribute the questionnaires to persons willing and able to contribute to the study as described in Chapter 3.12.5 was necessary for Austria and Greece, where there would otherwise have been no response. As the official institutions of Portugal overran the deadline significantly, two more questionnaires were sent through professional contacts, which led to two additional responses. Although the Italian Ministry of Health also overran the deadline, the completed questionnaire was finally returned after an official of the Ministry had asked the investigator by telephone, if the response was required.

4.1.2 Multiple responses

From the Republic of Ireland, both respondents reported the non-availability of national guidelines. From the UK, one letter referred to the new guideline for England and Wales
(National Collaborating Centre for Women and Children's Health 2003), the other respondent transferred the content of the same guideline into the survey table. This extract of the guideline was subsequently included in the analysis.

Denmark provided answers from the society of obstetricians, the Ministry of Interior and Health and the society of midwives. The responses reflected the situation of shared antenatal care between general practitioners, midwifery centres and hospital departments. Each professional group may have reported mainly on the measures for which they are responsible. The response from the Ministry was a summary of the answers from both professional societies. As the recommendations cited by the professional societies were included in the response from the Ministry, and as it was assumed that this is the least biased source with regard to professional judgements about individual tests, the response from the Ministry of Interior and Health was included in the analysis, and those of the professional societies excluded.

Portugal provided answers from the Ministry of Health, a University Professor and a midwife. In this case, the answers from the Ministry were included in the analysis as representing the data from the most official source as discussed above. In addition to that, the Portuguese guideline has been issued by the Ministry of Health. Therefore, the answer from the institution most closely related to the guideline was selected as the one accessing the latest version of the document, and being less prone to misinterpreting the recommendations.

The answers from a University-based midwife in Sweden and an obstetrician from a University Hospital were almost identical. The Swedish Research Council correctly responded that it does not issue a guideline, as the guideline is issued by the National Board of Health and Welfare.

Institutions from the Czech Republic and Lithuania provided two answers each. Their answers were again similar in content, with one Czech institution reporting to start care at ten gestational weeks, and one at twelve. However, as the onset of care depends on both the realisation of the woman of being pregnant, and her decision of when to contact health professionals, care will inevitably start at different times. Even if a guideline recommended
starting care as early as possible, not every woman will discover that she is pregnant at the same stage. Therefore it would be unrealistic if a guideline recommended to always commencing care at the same defined gestational week. Apart from the different starting date, the recommendations were the same from each source.

4.1.3 Sources of national guidelines

When analysing the selected questionnaires according to the framework described in Chapter 3, it was found that 20 member states (80%) have a national guideline on antenatal care. For these 20 member states with a national guideline it was found that 13 States have a guideline issued by the government, mainly the Ministry of Health. Four states have a guideline issued by the national society of obstetricians, and 3 states have a guideline published by a governmental institution in co-operation with the society of obstetricians. However, although most guidelines are issued by a governmental body, they are set up by multidisciplinary teams, always including obstetricians, but also experts from health insurances and other related disciplines. The guidelines from the United Kingdom and Germany serve as examples of this. The German guideline has been developed and issued by the Federal Joint Committee of the self-governing body of physicians and health insurance funds, but now has legal status. The guideline for England and Wales has been set up by a multidisciplinary team of experts from obstetrics, but also consumers, a radiographer and other specialties. However, the final guideline has been issued by the National Collaborating Centre for Women's and Children's Health, commissioned by the National Institute for Clinical Excellence, which is part of the National Health Service. From these findings it became clear that each guideline has an issuing body, but from the issuing body a conclusion can not necessarily be drawn about who set the guideline up. It is therefore concluded that the issuing body has an impact on the status of the guideline and on how (legally) binding it is, rather than reporting on the quality of the guideline.
4.1.4 Tests recommended for antenatal care

Five member states, i.e. Belgium, Cyprus, Greece, Ireland and Malta, do not have a guideline on antenatal care. However, only 5.75% of the total inhabitants of the 25 member states live in these countries. The data from the other 80% of member states is presented in Tables 4.2 to 4.7. In Table 4.2 the data shows in descending order, which and how many member states recommend each test for antenatal care. Similarly, Table 4.3 shows the size of the population to which the individual recommendations apply.
Table 4.2: National recommendations for individual tests. Number of countries.

<table>
<thead>
<tr>
<th>Test</th>
<th>Recommendation</th>
<th>Country</th>
<th>No. of countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood group</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>20</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>20</td>
</tr>
<tr>
<td>Rhesus factor determination</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>20</td>
</tr>
<tr>
<td>Maternal weight</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>19</td>
</tr>
<tr>
<td>Urinalysis / Bacteria</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>19</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>18</td>
</tr>
<tr>
<td>Lues</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>18</td>
</tr>
<tr>
<td>Urinalysis / Protein</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>18</td>
</tr>
<tr>
<td>Auscultation of foetal heart rate</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>17</td>
</tr>
<tr>
<td>Foetal position</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>17</td>
</tr>
<tr>
<td>Fundal height</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>17</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td>&lt;CZ&gt;</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Ultrasound, abdominal</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>17</td>
</tr>
<tr>
<td>Urinalysis / Glucose</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>17</td>
</tr>
<tr>
<td>Vaginal examination</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>16</td>
</tr>
<tr>
<td>Alpha-Feto-Proteine or Triple</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>14</td>
</tr>
<tr>
<td>Atypical red cell antibodies</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>14</td>
</tr>
<tr>
<td>Formal risk scoring</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>14</td>
</tr>
<tr>
<td>Full physical examination</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>14</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>&lt;CZ&gt;</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>HIV</td>
<td>&lt;CZ&gt;</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Rubella litter</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>12</td>
</tr>
<tr>
<td>Breast examination</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>11</td>
</tr>
<tr>
<td>Gestational diabetes - OGTT</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>11</td>
</tr>
<tr>
<td>PapanicoLoau smear</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>11</td>
</tr>
<tr>
<td>Ultrasound, transvaginal</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>11</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td>&lt;A&gt;</td>
<td>&lt;CZ&gt;</td>
<td>10</td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td>&lt;CZ&gt;</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Doppler ultrasound</td>
<td>&lt;CZ&gt;</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Fetal movement count</td>
<td>&lt;E&gt;</td>
<td>&lt;EST&gt;</td>
<td>8</td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td>&lt;CZ&gt;</td>
<td>&lt;EST&gt;</td>
<td>7</td>
</tr>
<tr>
<td>Gonorrhoea</td>
<td>&lt;CZ&gt;</td>
<td>&lt;EST&gt;</td>
<td>7</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td>&lt;CZ&gt;</td>
<td>&lt;EST&gt;</td>
<td>7</td>
</tr>
<tr>
<td>Streptococcus group B</td>
<td>&lt;CZ&gt;</td>
<td>&lt;EST&gt;</td>
<td>5</td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td>&lt;A&gt;</td>
<td>&lt;E&gt;</td>
<td>4</td>
</tr>
<tr>
<td>Placental hormones</td>
<td>&lt;GK&gt;</td>
<td>&lt;E&gt;</td>
<td>3</td>
</tr>
<tr>
<td>Foetal fibronectin</td>
<td>&lt;GK&gt;</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Test</td>
<td>Recommendation</td>
<td>Country</td>
<td>No. of inhabitants</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Blood group</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>425 842 500</td>
</tr>
<tr>
<td>Blood pressure</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>425 842 500</td>
</tr>
<tr>
<td>Rhesus factor determination</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>425 842 500</td>
</tr>
<tr>
<td>Foetal position</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>413 838 600</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>412 340 400</td>
</tr>
<tr>
<td>Urinalysis / Bacteria</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>409 978 500</td>
</tr>
<tr>
<td>Maternal weight</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>395 881 800</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>383 894 000</td>
</tr>
<tr>
<td>Ultrasound, abdominal</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>370 891 300</td>
</tr>
<tr>
<td>Auscultation of foetal heart rate</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>363 403 800</td>
</tr>
<tr>
<td>Lues</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>361 763 800</td>
</tr>
<tr>
<td>Rubella titer</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>356 984 900</td>
</tr>
<tr>
<td>Urinalysis / Protein</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>352 296 600</td>
</tr>
<tr>
<td>Alpha-Feto-Proteine or Triple</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>346 700 900</td>
</tr>
<tr>
<td>Atypical red cell antibodies</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>346 700 900</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>331 536 100</td>
</tr>
<tr>
<td>Fundal height</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>318 003 700</td>
</tr>
<tr>
<td>Urinalysis / Glucose</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>292 674 700</td>
</tr>
<tr>
<td>Vaginal examination</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>292 241 600</td>
</tr>
<tr>
<td>HIV</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>256 501 800</td>
</tr>
<tr>
<td>Ultrasound, transvaginal</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>248 289 500</td>
</tr>
<tr>
<td>Gestational diabetes - OGTT</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>246 544 200</td>
</tr>
<tr>
<td>Formal risk scoring</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>214 607 400</td>
</tr>
<tr>
<td>Full physical examination</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>201 962 700</td>
</tr>
<tr>
<td>Doppler ultrasound</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>198 308 200</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>197 859 100</td>
</tr>
<tr>
<td>Breast examination</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>192 526 800</td>
</tr>
<tr>
<td>Papanicolaou smear</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>180 825 900</td>
</tr>
<tr>
<td>Streptococcus group B</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>172 066 100</td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>132 792 700</td>
</tr>
<tr>
<td>Fetal movement count</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>128 157 000</td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>123 443 300</td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>84 116 700</td>
</tr>
<tr>
<td>Gonorrhoea</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>73 101 600</td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>68 958 500</td>
</tr>
<tr>
<td>Placental hormones</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>50 689 400</td>
</tr>
<tr>
<td>Foetal fibronectin</td>
<td></td>
<td>CA &lt; GR &lt; EE &lt; EST &lt; FR &lt; DE &lt; IT &lt; LT &lt; LV &lt; NL &lt; P &lt; PL &lt; SC &lt; SK &lt; SL0 &lt; SW ? UK</td>
<td>5 396 700</td>
</tr>
</tbody>
</table>
Drawing the findings of these two tables together, it was found that 23 tests are recommended by more than 50% of the member states and additionally apply to more than 50% of inhabitants. This approach has already been explained in the paragraph on data analysis in the methodology section. The thus identified tests are shown in Table 4.4 together with the frequency with which they are recommended. The other 14 tests asked for in the survey failed to fulfil either one or both of the criteria for double majorities and are shown in Table 4.5. In both tables the tests are ranked with the test recommended by most member states on top and the test recommended by the lowest number of member states at the bottom.
Table 4.4: Recommended number of individual tests per pregnancy (by more than 50% of countries applying to more than 50% of inhabitants)

<table>
<thead>
<tr>
<th>Test</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood group</td>
<td>1</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>2</td>
</tr>
<tr>
<td>Rhesus factor determination</td>
<td>3</td>
</tr>
<tr>
<td>Maternal weight</td>
<td>4</td>
</tr>
<tr>
<td>Urinalysis / Bacteria</td>
<td>5</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td>6</td>
</tr>
<tr>
<td>Lues</td>
<td>7</td>
</tr>
<tr>
<td>Urinalysis / Protein</td>
<td>8</td>
</tr>
<tr>
<td>Auscultation of foetal heart</td>
<td>9</td>
</tr>
<tr>
<td>Foetal position</td>
<td>10</td>
</tr>
<tr>
<td>Fundal height</td>
<td>11</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td>12</td>
</tr>
<tr>
<td>Ultrasound, abdominal</td>
<td>13</td>
</tr>
<tr>
<td>Urinalysis / Glucose</td>
<td>14</td>
</tr>
<tr>
<td>Vaginal examination</td>
<td>15</td>
</tr>
<tr>
<td>Alpha-Feto-Proteine or Triple</td>
<td>16</td>
</tr>
<tr>
<td>Atypical red cell antibodies</td>
<td>17</td>
</tr>
<tr>
<td>Formal risk scoring</td>
<td>18</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>19</td>
</tr>
<tr>
<td>HIV</td>
<td>20</td>
</tr>
<tr>
<td>Rubella titer</td>
<td>21</td>
</tr>
<tr>
<td>Gestational diabetes - OGTT</td>
<td>22</td>
</tr>
<tr>
<td>Ultrasound, transvaginal</td>
<td>23</td>
</tr>
</tbody>
</table>

**min** | **mean** | **max**

25% below mean | 25% above mean
Table 4.5: Recommended number of individual tests per pregnancy (by less than 50% of countries or applying to less than 50% of inhabitants)

<table>
<thead>
<tr>
<th>Test</th>
<th>Test Number</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full physical examination</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Breast examination</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Papanicolaou smear</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Doppler ultrasound</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Foetal movement count</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gonorrhoea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus group B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placental hormones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal fibronectin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although it was found that all 37 tests asked for in the survey are recommended by at least one member state, only 3 tests (8.1%) are recommended by all 25 member states of the European Union. These universally recommended tests are measuring the maternal blood pressure and determining her blood group and Rhesus factor. However, respondents from four member states made additional use of the opportunity to report tests from their guidelines which were not asked for in the survey. The respondents from these countries added 10 more tests so that the entire number of tests used for routine antenatal care in the EU rose to 47. There were no overlaps in the additional recommendations. This means that all added tests are recommended by the respective member state only. These are listed below:
• Austria additionally recommends measurement of maternal height as well as abdominal circumference. Regular checks for oedema and varicose veins are also recommended. Up to the 16th gestational week, a gynaecological examination should take place, during which also a vaginal smear is taken and checked for pathogens. An erythrocyte count or the determination of the haematocrit also takes place once up to the 16th gestational week, and for a second time between weeks 25 and 28.

• Hungary recommends a non-stress test between 36 and 38 weeks, and another one between the 38th and the 40th gestational week. In addition to that, the Hungarian guideline recommends an oxytocin challenge test between 38 and 40 weeks, which is again not recommended for normal pregnancies in the other member states.

• Luxembourg recommends between weeks 21 and 25 a test for glycaemia, which means the measurement of blood sugar without any controlled intake of glucose.

• Poland recommends the testing of the vaginal pH at weeks 10, 20, 32 and 37.

4.1.5 National recommendations and published evidence

When the tests recommended by at least 50% of member states and applying to at least 50% of inhabitants (double majorities) were finally compared to the published evidence, the idea of using these recommendations for a common minimum guideline gained support. It was found that 19 tests were recommended by the member states with double majorities, and can additionally be recommended according to the published scientific evidence as documented in the evidence tables of the National Collaborating Centre for Women’s and Children’s Health (National Collaborating Centre for Women’s and Children’s Health 2003). Fourteen tests were neither recommended with double majorities, nor are they supported by the literature (Table 4.5). Most interesting is the finding that only four tests were recommended by double majorities, but are currently not sufficiently supported or discouraged by the literature as specified in the guideline from England and Wales (National Collaborating Centre for Women’s
and Children’s Health 2003). For vaginal examinations, there is strong evidence that it is ineffective to predict a premature ripening of the cervix. The auscultation of the foetal heart rate is not recommended as it has no further clinical or predictive value than to confirm that the foetus is alive at the time of auscultation. Performing urine glucose tests is discouraged on the basis of its low sensitivity and high number of false positives. The routine use of oral glucose tolerance tests for gestational diabetes is also discouraged, as currently there is neither any consent about the definition, nor about the management of gestational diabetes mellitus. The National Collaborating Centre for Women’s and Children’s Health came to the conclusion that the evidence does not support routine screening for gestational diabetes mellitus.

4.1.6 The link between the gross national product and the number of tests

The answer to the question of whether there is a correlation between the Gross National Product and the number of recommended tests is provided in Tables 4.6 and 4.7. From these tables it becomes evident that member states with a GNP above average (GNP50+) do in fact recommend fewer types of tests and a smaller total number of tests per pregnancy than the member states with a GNP below average (GNP50-).
Table 4.6: Correlation between Gross National Product (GNP) per capita in Euro and intensity of care

<table>
<thead>
<tr>
<th>Country</th>
<th>GNP in €</th>
<th>Recommendation</th>
<th>Types of tests</th>
<th>Total No. of tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxembourg</td>
<td>34 598</td>
<td></td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>Denmark</td>
<td>26 575</td>
<td></td>
<td>15</td>
<td>49</td>
</tr>
<tr>
<td>Sweden</td>
<td>22 709</td>
<td></td>
<td>25</td>
<td>65</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>22 323</td>
<td></td>
<td>18</td>
<td>52</td>
</tr>
<tr>
<td>Finland</td>
<td>21 276</td>
<td></td>
<td>26</td>
<td>118</td>
</tr>
<tr>
<td>Austria</td>
<td>21 099</td>
<td></td>
<td>30</td>
<td>68</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>20 716</td>
<td></td>
<td>19</td>
<td>38</td>
</tr>
<tr>
<td>Germany</td>
<td>19 882</td>
<td></td>
<td>20</td>
<td>114</td>
</tr>
<tr>
<td>France</td>
<td>19 504</td>
<td></td>
<td>25</td>
<td>62</td>
</tr>
<tr>
<td>Italy</td>
<td>18 819</td>
<td></td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>Spain</td>
<td>13 378</td>
<td></td>
<td>32</td>
<td>120</td>
</tr>
<tr>
<td>Portugal</td>
<td>9 551</td>
<td></td>
<td>24</td>
<td>111</td>
</tr>
<tr>
<td>Slovenia</td>
<td>9 315</td>
<td></td>
<td>22</td>
<td>140</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>5 307</td>
<td></td>
<td>32</td>
<td>106</td>
</tr>
<tr>
<td>Hungary</td>
<td>4 984</td>
<td></td>
<td>25</td>
<td>136</td>
</tr>
<tr>
<td>Poland</td>
<td>4 150</td>
<td></td>
<td>22</td>
<td>96</td>
</tr>
<tr>
<td>Estonia</td>
<td>3 906</td>
<td></td>
<td>25</td>
<td>127</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>3 874</td>
<td></td>
<td>37</td>
<td>171</td>
</tr>
<tr>
<td>Lithuania</td>
<td>3 535</td>
<td></td>
<td>22</td>
<td>61</td>
</tr>
<tr>
<td>Latvia</td>
<td>3 205</td>
<td></td>
<td>27</td>
<td>70</td>
</tr>
</tbody>
</table>

The GNP50- countries recommend on average 1.3-times more types of tests compared to the GNP50+ countries. For the average total number of tests for one pregnancy, the results are even more impressive. The GNP50- countries recommend 1.8-times as many tests as the GNP50+ group. The same associations were found when the GNP was adapted to the national purchasing power parities (data not shown). Of the member states without a guideline, Belgium and Ireland have a GNP above the average of € 14 432.30. Cyprus, Greece and Malta have a GNP below.

Table 4.7: Correlation between average Gross National Product (GNP) and intensity of care

<table>
<thead>
<tr>
<th>Group of countries</th>
<th>Recommendation</th>
<th>Types of tests</th>
<th>Total No. of tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNP above average</td>
<td></td>
<td>21.3</td>
<td>64.1</td>
</tr>
<tr>
<td>GNP below average</td>
<td></td>
<td>26.8</td>
<td>113.8</td>
</tr>
</tbody>
</table>
4.1.7 Frequency of tests

Tables 4.4 and 4.5 show how often tests are recommended during pregnancy and that most of the tests are recommended to be performed more than once during pregnancy. Using a box-and-whiskers type of diagram, the means are clear, as well as the respective 25% ranges below and above the mean. In addition to this, the highest and lowest number of repetitions can also be seen. However, although these tables provide a useful overview, it has to be considered that the reported repetitions of a test depend on the point at which the care began. This differs widely and is a product of the national recommendations of when to start care, but also of when pregnant women decide to first contact health professionals.

4.2 Results from the in-depth appraisal of two national guidelines

In accordance with what was described in Chapter 3 on methodology, the in-depth appraisal of the national guidelines from England and Wales and Germany was conducted in two stages. In a first step, the methodological quality of the two guidelines was assessed with the AGREE-instrument by four raters with purposefully selected different professional backgrounds. In a second step, the recommendations of the guidelines were compared to each other. The findings of both stages are presented in the following section.

4.2.1 Appraisal of the methodological quality of the two national guidelines

The following critical appraisal of the methodological quality of the guidelines from England and Wales and Germany is intended to contribute to reach the objective of Phase 1 of the study, which is to critically appraise the relation between evidence-based guidelines and guidelines based on expert opinion for the defined public health task 'antenatal care'. As planned, the appraisals of both guidelines were conducted between May and August 2004, using the instrument of the AGREE-Collaboration. They could not be conducted simultaneously, as it was not possible to fix a date with all appraisers at the same time.
However, the appraisers neither knew each other, nor had any contact. Moreover, the appraisers’ names were kept confidential. Any influence from discussions or other contact was therefore excluded. In line with the recommendations of the developers of the AGREE-instrument, four appraisers were chosen to critically assess the guidelines. However, as an amendment to using four unspecific appraisers, the method was expanded by purposefully selecting four raters with different professional backgrounds, involved with different stages of using or setting up guidelines. By this, a deeper and more critical view on the assessed guidelines should be obtained.

Although appraising the guidelines demanded time and effort from the appraisers, the guidelines were finally assessed by the designated four appraisers. Each appraiser answered all 23 questions for both guidelines and made a final comment on whether the guideline should be recommended. To ease analysis and to ensure anonymity, the following numbers were assigned to the appraisers:

- A1: Professor of gynaecology and obstetrics
- A2: Scientific assistant in nursing sciences
- A3: Midwife
- A4: The person conducting the study.

After all appraisers returned their material, the analysis was conducted according to the instructions of the AGREE-Collaboration, which are explained in the following. The guidelines are therefore assessed in the domains ‘scope and purpose’, ‘stakeholder involvement’, ‘rigour of development’, ‘clarity and presentation’, ‘applicability’ and ‘editorial independence’. According to the instructions for analysis, the scores for each domain were calculated separately by summing up all scores on the individual questions in a domain and subsequently calculating the total as a percentage of the maximum possible score for that domain. Using the following formula it is possible to calculate every domain score for each individual appraiser, but also to calculate the standardised domain score out of the ratings of all appraisers together:
obtained score – minimum possible score

maximum possible score – minimum possible score

= (standardised) domain score in %

This formula explains also, why for example in Table 4.8 a score of 12 means a domain score of 100%, but a score of 6 is equal to 33%, rather than to 50%, which would be normally expected. However, if the minimum score of a domain with three questions is 3 and the maximum score is 12, a score of 6 means a domain score of 33%. This would be only different if the lowest score were 0. As this might be a source of misunderstandings, this logical problem has to be kept in mind while reading all following tables and the results section on the appraisal of the methodological quality of the two national guidelines. Another important point to be noted is that it is not possible to aggregate the domain scores into a single quality score, but to use the scores separately to comment on the particular strengths and weaknesses of a guideline. Unfortunately, the developers of the AGREE-instrument could not set thresholds for marking good or bad guidelines. As this does not lend itself directly to an explanation, and due to the fact that the AGREE-collaboration has not provided an explanation either, the reason for this might be discovered when using the instrument to assess the two national guidelines. Due to the lack of thresholds for good or bad guidelines, the ratings of the appraisers have to be considered together with the additional comments and the final overall assessment they made, which might be regarded as a subjective factor in the appraisal system. However, also the different rating patterns of the different professionals are used to gain additional insights.

In order to ease the assimilation and understanding of the findings, first of all the ratings of all four appraisers will be analysed for the individual domains of the England and Wales guideline, and in the next step the corresponding domain of the guideline from Germany. At the same time, the ratings of the individual appraisers will be assessed with regard to specific traits of their ratings. It will be critically analysed whether the raters with a background in obstetrics show a different rating pattern than the scientific assistant from another discipline. In addition
to the individual scores, the written comments of the appraisers will also be considered.

After that, the findings for the individual domains will be brought together to ease a clear view on the overall quality of each guideline. In a last step, the findings from this will then be compared to the overall assessment of each guideline as provided by the appraisers.

Scope and purpose

As becomes clear from Table 4.8, all four appraisers were convinced by the descriptions of the scope and the purpose of the guideline from England and Wales. All three items were rated with the maximum score. All appraisers found that the objectives, as well as the target group and the clinical condition covered by the guideline are clearly described. The description is clear and understandable to readers without prior knowledge of the field, but according to the high ratings of raters A1, A3 and A4 also acceptable to specialists.

Table 4.8: Scope and purpose of the guideline from England and Wales

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>1. The overall objective(s) of the guideline is(are) specifically described</td>
<td>4</td>
</tr>
<tr>
<td>2. The clinical question(s) covered by the guideline is(are) specifically described</td>
<td>4</td>
</tr>
<tr>
<td>3. The patients to whom the guideline is meant to apply are specifically described</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>100</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>100%</td>
</tr>
</tbody>
</table>
In contrast to that, the purpose of the German guideline was probably easier to understand for the appraisers with a background in obstetrics, which was explicitly stated by appraiser A2. The scientific assistant from the nursing sciences commented that the patients to whom the guideline is meant to apply are not explicitly described, and that it is rather difficult to understand without deeper knowledge of obstetrics. As can be seen from Table 4.9, the appraiser without prior knowledge rated the domain scope and purpose with only 33%, which is by far the lowest rating. In particular, the description of the clinical questions covered by the guideline did not convince this appraiser. However, the practising midwife was also not entirely satisfied with this. She stated that the overall objective of the guideline was not specifically described, as only two sentences referred to the benefits of medical treatment during pregnancy in general. Appraiser A2 commented that the overall objectives are mentioned, but rather unspecific. However, appraiser A3 stated that the guideline gives detailed information about when a medical check up during pregnancy has to take place. In addition, she found that the guideline provides information about normal results and how to act on irregular findings.

Table 4.9: Scope and purpose of the German guideline

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>1. The overall objective(s) of the guideline is(are) specifically described</td>
<td>4</td>
</tr>
<tr>
<td>2. The clinical question(s) covered by the guideline is(are) specifically described</td>
<td>4</td>
</tr>
<tr>
<td>3. The patients to whom the guideline is meant to apply are specifically described</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>100</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>69%</td>
</tr>
</tbody>
</table>
Although the ratings from the appraisers with a professional background were generally higher, only the professor of gynaecology and obstetrics rated all items within this domain with the maximum score. This might be due to additional information provided by the professional organisation, which is not known to the other appraisers. Another possible explanation is that the high rating is associated with having practical experience of working with the guideline.

Although appraiser A4 attached a high score to all items, an additional critical remark was made with regard to the specification of patients to whom the guideline is meant to apply. Appraiser A4 commented that the guideline seemed to apply to all pregnant women, but that it is not specified whether the guideline applies irrespective of the number of the pregnancy or maternal age etc. Moreover, it was also stated that the guideline applies to all pregnant women who are covered by general health insurances and not to pregnant women per se. Appraiser A3 commented also that there was no doubt about who the patient group were, i.e. pregnant women, while reading the guideline, while appraiser A2 mentioned that the patients were not described in greater detail and that the guideline was very physician-oriented. However, despite some criticism, a standardised domain score of 69% was calculated for the German guideline when finally drawing the individual ratings of the appraisers together.

Stakeholder involvement

Table 4.10 shows that all appraisers were convinced that the guideline development group of the guideline for England and Wales was adequately chosen, including individuals from all relevant professional groups. The same level of satisfaction was demonstrated with regard to the clear definition of the target users of the guideline. Moreover, all but appraiser A4 were completely satisfied with the degree to which patients' views and preferences have been sought. Appraiser A4 reported that two consumers were in the guideline development group. However, only appraiser A3 was entirely satisfied with the piloting of the guideline. This finding is important, as appraiser A3 has experience as a user of guidelines, but virtually no knowledge about the principles of scientific work. The three raters with knowledge and
experience in dealing with scientific questions and with guideline development were not convinced that the guideline from England and Wales was sufficiently piloted. However, appraiser A4 acknowledged that the guideline reported on pages 3 to 7 had undergone an "external review process". Appraiser A2 also stated that he did not find any report of piloting, but that the guideline was externally reviewed. Despite the doubts about the appropriate piloting of the guideline, the final standardised domain score of 88% demonstrates a high overall level of satisfaction with regard to stakeholder involvement in the guideline development process.

Table 4.10: Stakeholder involvement of the guideline from England and Wales

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>4. The guideline development group included individuals from all the relevant professional groups</td>
<td>4</td>
</tr>
<tr>
<td>5. The patients' views and preferences have been sought</td>
<td>4</td>
</tr>
<tr>
<td>6. The target users of the guideline are clearly defined</td>
<td>4</td>
</tr>
<tr>
<td>7. The guideline has been piloted among target users</td>
<td>3</td>
</tr>
</tbody>
</table>

Score type                                      Score
Domain score per appraiser (%)                  92  83  100  75
Standardised domain score                       88%

With a standardised domain score of 25%, the German guideline is rated extremely low with
regard to stakeholder involvement in the guideline development process (Table 4.11).

Especially appraiser A2, i.e. the scientific assistant from the nursing sciences, was critical and found no single criterion of the domain fulfilled. He found that for the development of the German guideline neither the relevant professional groups were chosen, nor were patients' views sought. He even found that the target users of the guideline were not clearly defined. However, the ratings of the other appraisers indicate the same extremely critical view only on the issues that patients' views were obviously not sought and that the guideline does not contain any hint as to whether it was piloted at all.

Table 4.11: Stakeholder involvement of the German guideline

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>A2</td>
</tr>
<tr>
<td>4. The guideline development group included individuals from all the relevant professional groups</td>
<td>3</td>
</tr>
<tr>
<td>5. The patients' views and preferences have been sought</td>
<td>1</td>
</tr>
<tr>
<td>6. The target users of the guideline are clearly defined</td>
<td>4</td>
</tr>
<tr>
<td>7. The guideline has been piloted among target users</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>42</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td></td>
</tr>
</tbody>
</table>

An interesting finding was made with regard to the judgment about the appropriate selection of
the guideline development group. According to the original document, the German guideline was decided upon by the Federal Joint Committee of the self-governing body of physicians and health insurance funds. No further reference can be found in the document as to who the actual individuals were, and what were their professions, or positions in the guideline development group. The practising midwife, i.e. appraiser A3, did not challenge the fact that the contribution of neither a midwife, nor of a patient representative is reported in the guideline. The professor of gynaecology and obstetrics (appraiser A1) was also satisfied with the selection of the development group, which might be explained with a possibly biased view onto the subject, as only physicians and health insurances were represented in the group. However, appraisers A2 and A4 were more critical about this issue and were not convinced that all relevant professional groups were involved. Even the scientific assistant from another discipline, i.e. appraiser A2, noted that the recommendations were defined by physicians only.

In contrast to appraisers A2 and A3, who found that the target users of the guideline were not clearly defined, appraisers A1 and A4 awarded the guideline the highest possible score in this area. As this rating could not be explained by possible prior knowledge or experience, the original guideline was consulted in order to clarify the issue. According to the explanation from within the AGREE-instrument, the question was whether the target users of the guideline are clearly defined, so that they can immediately determine if the guideline is relevant to them. In the first section of the original guideline it is stated that the guideline was introduced to ensure sufficient, appropriate and cost-effective antenatal and postnatal care for women insured under general health insurance, and refers to care which is provided by physicians. In a later section on the same page, it is further specified that the guideline applies to those physicians, who are qualified to do so according to their knowledge, experience and equipment, and who are allowed to do so according to professional regulations. Only under paragraph seven on page 6 is reference made to the fact that some of the recommended measures can also be performed by a midwife, as long as the physician has either given the order to do so, or if a physician has diagnosed a normal pregnancy and has no objections to the continuation of antenatal care by a midwife. However, it has to be noted that the guideline refers to antenatal care as funded by
general health insurances only. Despite this, the guideline is used as a gold standard by all professions involved in antenatal care in Germany. Why appraiser A2 did not count the above information as a clear definition of the target users of the guideline is not clear. It might be that the information was not provided in a manner which can be easily picked up, and the appraiser put the focus on target users needing to be "clearly" defined. It might have helped, if the guideline provided the information in a structured and more straightforward manner. The practising midwife commented that she was not satisfied with definitions. She was critical that it was not written in the guideline, "what kind of doctor is responsible for the medical check ups during pregnancy". The practising midwife realised from her experience that it is normally the obstetricians practising independently in the community, who provide antenatal care. This is a special trait of the German system, and appraiser A3 holds obviously the opinion that those physicians who are normally providing the care should be called by their name.

Rigour of development

Also in the domain 'rigour of development', the guideline from England and Wales was rated high with a standardised domain score of 88%. The detailed numbers as compiled in Table 4.12 demonstrate that all four appraisers found that systematic methods were used to search for evidence and that the criteria for selecting it did become clear. All appraisers were also satisfied with the level to which the envisaged positive, but also the possible negative effects of the recommendations have been considered. They were also convinced that there is an explicit link between the recommendations and the supporting evidence. When addressing question number 13 of whether the guideline has been externally reviewed, the rating of the appraisers was again uniform and attained the highest possible score. This reinforces what was stated in the previous section on stakeholder involvement, where the appraisers commented that there was no evidence of piloting, but that the guideline was externally reviewed. Appraisers A2 and A4 in particular acknowledged that the guideline reported undergoing an external review process. For the question of whether the methods used for formulating the recommendations are clearly described, the ratings were generally high,
although two raters did not give the highest possible score.

Table 4.12: Rigour of development of the guideline from England and Wales

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>8. Systematic methods were used to search for evidence</td>
<td>4</td>
</tr>
<tr>
<td>9. The criteria for selecting the evidence are clearly described</td>
<td>4</td>
</tr>
<tr>
<td>10. The methods used for formulating the recommendations are clearly described</td>
<td>3</td>
</tr>
<tr>
<td>11. The health benefits, side effects and risks have been considered in formulating the recommendations</td>
<td>4</td>
</tr>
<tr>
<td>12. There is an explicit link between the recommendations and the supporting evidence</td>
<td>4</td>
</tr>
<tr>
<td>13. The guideline has been externally reviewed by experts prior to its publication</td>
<td>4</td>
</tr>
<tr>
<td>14. A procedure for updating the guideline is provided</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>81</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>88%</td>
</tr>
</tbody>
</table>

In contrast to the homogenous rating for the other items, there was major disagreement on whether a procedure for updating the guideline is provided. The practising midwife thought that it was provided, whereas the other three appraisers were not convinced that this was the case. Appraiser A4 might have brought up the reason for this by stating that no date and no specific
process for updating were specified, but that on pages 19 and 20 of the guideline recommendations are made for future research. While the midwife without deeper knowledge about the theoretical ideas behind guidelines has accepted this as an appropriate procedure, the more experienced appraisers might have regarded this as appropriate for a piece of academic work, but would have expected a specific date for updating a guideline.

Table 4.13: Rigour of development of the German guideline

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>8. Systematic methods were used to search for evidence</td>
<td>1</td>
</tr>
<tr>
<td>9. The criteria for selecting the evidence are clearly described</td>
<td>1</td>
</tr>
<tr>
<td>10. The methods used for formulating the recommendations are clearly described</td>
<td>1</td>
</tr>
<tr>
<td>11. The health benefits, side effects and risks have been considered in formulating the recommendations</td>
<td>2</td>
</tr>
<tr>
<td>12. There is an explicit link between the recommendations and the supporting evidence</td>
<td>1</td>
</tr>
<tr>
<td>13. The guideline has been externally reviewed by experts prior to its publication</td>
<td>1</td>
</tr>
<tr>
<td>14. A procedure for updating the guideline is provided</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>5</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>10%</td>
</tr>
</tbody>
</table>
With a standardised domain score of 10% for the category rigour of development, the rating of the German guideline stands in sharp contrast to the high rating of the guideline from England and Wales (Table 4.13). For the German guideline, appraiser A4 attached to each item the lowest possible score. Appraisers A1 and A2 each rated only one item higher than the most basic score, and only the ratings of appraiser A3 show a more heterogeneous picture.

For the question of whether systematic methods were used to search for evidence, appraiser A3 noted that there is a number on the front page where the guideline is archived, and she speculated about whether it might be possible to get more information. As this is only an archive number, where the guideline can be found, this is definitely not the case. The same appraiser commented also that there are “useful comments on the bottom of some pages”, e.g. pages 9, 10 and 17. However, these footnotes refer to requirements for laboratory or ultrasound examinations, but do not provide any information about how the search for evidence was conducted. As there is no information about how the search for evidence took place, all other appraisers judged this item as not fulfilled by the guideline.

In contrast to the search for evidence, all appraisers agreed that the German guideline does not provide any data about the inclusion or exclusion of evidence. They also agreed that the guideline did not provide any information about the methods used for formulating the recommendations, which implies also that the link between the recommendations and the evidence is not clear. All appraisers judged item 12 accordingly with the lowest possible score. Due to the entire lack of procedural information, the question of whether the guideline had been externally reviewed could not be answered, i.e. was scored as low as possible.

With regard to the question of whether the guideline describes health benefits, side effects and risks in relation to its recommendations, appraisers A2 and A4 decided that no information was provided. However, appraiser A4 made the additional comment that the guideline intends to avert potential dangers for the life and the health of mother and infant. This appraiser stated also that potential side effects and risks are not mentioned, and therefore did not give a higher score. Appraiser A3 comments that neither risks, nor health benefits nor side effects are
mentioned except for two sentences on page 2 of the guideline. Despite this, she ranked the issue one score higher than appraisers A2 and A4.

The ratings of appraisers A1 and A4 refer to the fact that the German guideline gives no specific date for updating it. Appraiser A4 mentioned additionally that the guideline specifies the day from which it is effective, but does not give a date for updating. The above ratings demonstrate a straightforward answer to the question of whether a procedure for updating the guideline is provided. However, appraisers A2 and A3 realised and commented that on the first page of the guideline there is evidence of updates, but that it does not become clear why, how and by whom the updates were made.

Clarity and presentation

The standardised domain score of 90% demonstrates well that the four appraisers were satisfied with the clarity and the presentation of the guideline from England and Wales (Table 4.14). They all concluded that the recommendations of the guideline are specific and unambiguous, and that the key recommendations are easily identifiable. Appraiser A4 added the comment that a separate page and a poster with the clinical algorithm are provided, which makes the guideline easily applicable. Information for consumers and for professionals is given separately in the appendix, which might be also useful. As all these materials are provided by the guideline, it is not clear why appraiser A2 decided that the guideline is not supported with tools for application. In addition to this potential criticism, the practising midwife found that the different options for management of the condition are not clearly presented. Unfortunately, she did not elaborate further on this issue.
Table 4.14: Clarity and presentation of the guideline from England and Wales

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>15. The recommendations are specific and unambiguous</td>
<td>4</td>
</tr>
<tr>
<td>16. The different options for management of the condition are clearly presented</td>
<td>4</td>
</tr>
<tr>
<td>17. Key recommendations are easily identifiable</td>
<td>4</td>
</tr>
<tr>
<td>18. The guideline is supported with tools for application</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>92</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>90%</td>
</tr>
</tbody>
</table>

Although the difference was not as large as for the other domains, the German guideline scored 30% lower than the guideline from England and Wales with regard to its clarity and presentation (Table 4.15). While the professor of gynaecology and obstetrics was more or less satisfied and scored this domain with 15 out of 16 possible points, the scientific assistant from the nursing sciences was not convinced by the guideline and scored clarity and presentation with 5 out of 16. Most interesting is, however, that all appraisers with a background in obstetrics scored this domain significantly higher than the scientific assistant from another discipline. Background knowledge might have helped to judge for example whether the recommendations are specific and unambiguous enough. However, although appraiser A2 found that there were at least partially concrete statements it did not become clear, which of
them are evidence-based. As he obviously wanted this particular information, he rated item 15 lower than the other appraisers. In this case, the appraisers with knowledge of obstetrics might interpret the guideline and judge for themselves whether the recommendations are plausible, or not. In this respect these appraisers might be biased and interpret something into the guideline, which is not written there.

Table 4.15: Clarity and presentation of the German guideline

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>15. The recommendations are specific and unambiguous</td>
<td>4</td>
</tr>
<tr>
<td>16. The different options for management of the condition are clearly presented</td>
<td>4</td>
</tr>
<tr>
<td>17. Key recommendations are easily identifiable</td>
<td>3</td>
</tr>
<tr>
<td>18. The guideline is supported with tools for application</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>92  8  67  75</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>60%</td>
</tr>
</tbody>
</table>

As can be also seen from Table 4.15, the professor of gynaecology and obstetrics was convinced that the different options for management of the condition are clearly presented. The other three raters were not entirely convinced, and e.g. appraiser A3 commented "Are there different options in treatment? It's not written. There is just an option between the one who treats the pregnant women ..." However, appraiser A4 found that different options for management are presented, such as by giving time windows for ultrasound screening, and by
specifying that haemoglobin levels should be re-assessed according to the initial test result. However, there are no real alternative options discussed, which relate to preferences of the pregnant woman, or to any uncertainties with regard to the benefits of a test.

With regard to the clarity of presenting the key recommendations such as to make them easily identifiable, appraiser A3 commented that users would be able to find the most important recommendations. However, she suggested also that the key recommendations would be easier to find if there were more emboldened words. The other two appraisers with knowledge in obstetrics were also not entirely satisfied, and the one without prior knowledge was completely dissatisfied with the presentation.

Most evident was the difference between those with background knowledge in obstetrics and the one without, when the question was asked as to whether the guideline is supported with tools for application. It might be that the three appraisers with prior knowledge thought of the so-called "Mutterpass", which is a concise booklet in which the results of all examinations are recorded, when they answered this question. At least appraiser A4 made reference to this booklet, which is carried by the pregnant woman herself, and in which all test results should be documented. Its format is such as to provide defined spaces for each examination, thus helping to prevent any omissions in care. As it mirrors the recommendations of the guideline, it definitely helps to apply it to all pregnant women. This might have not been known to appraiser A2, as he is not professionally involved in antenatal care.

Applicability

Table 4.16 shows that the guideline from England and Wales scores with a standardised domain score of 69% for its applicability, which was the lowest score. No appraiser was entirely convinced by the properties of the guideline in this domain, although appraiser A3 awarded 11 out of 12 possible points, which was by far the highest rating. This might again be due to her lack of experience with scientific work, and the assessment of guideline quality. Most interesting was the finding that appraisers A2 and A3 both found that the potential
organisational barriers in applying the recommendations had been discussed in the guideline. Appraiser A1, the professor of gynaecology and obstetrics, disagreed without further comments. However, appraiser A4 disagreed even more strongly and added that the potential organisational barriers have not been discussed, and that there is no comment as to how the guideline should be implemented, or what changes are necessary. The same appraiser added also that this might be self-explanatory for health professionals working in the system, but that this is not that obvious for interested parties who are not familiar with the health system.

Table 4.16: Applicability of the guideline from England and Wales

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>19. The potential organisational barriers in applying the recommendations have been discussed</td>
<td>2</td>
</tr>
<tr>
<td>20. The potential cost implications of applying the recommendations have been considered</td>
<td>3</td>
</tr>
<tr>
<td>21. The guideline presents key review criteria for monitoring and/or audit purposes</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>67</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td></td>
</tr>
</tbody>
</table>

The ratings on the question whether potential cost implications of applying the recommendations have been considered were generally high, and appraiser A2 added that they were partially discussed in the guideline. Also high ratings were made as the appraisers found that the guideline presents key criteria for monitoring or audits. However, appraiser A2, who has experience in this field, judged the provided criteria as not necessarily plausible.
Also in the domain ‘applicability’ the German guideline was rated lower than the one from England and Wales. Table 4.17 provides the detailed ratings of the appraisers and thus demonstrates from where the standardised domain score of 36% was derived. First of all, none of the appraisers found sufficient evidence about a discussion of the potential organisational barriers in applying the recommendations. The appraisers' interpretation of whether the potential cost implications of applying the recommendations were considered was interesting. Whereas appraisers A2 and A3 found that these implications were not at all considered, appraisers A1 and A4 were convinced that they were considered. The latter two appraisers were possibly satisfied with the introductory statement about the aims of the guideline. These were to recommend care, which is sufficient, suitable and economical, as noted by appraiser A4. Appraiser A4 also mentioned that this is claimed by the guideline, but that there is no comment about how this has been assessed, or to what degree it is achieved. However, it should be borne in mind that health insurance played a major role in the guideline development process. It is therefore unlikely that the cost implications were not considered.

Table 4.17: Applicability of the German guideline

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>19. The potential organisational barriers in applying the recommendations have been discussed</td>
<td>1</td>
</tr>
<tr>
<td>20. The potential cost implications of applying the recommendations have been considered</td>
<td>4</td>
</tr>
<tr>
<td>21. The guideline presents key review criteria for monitoring and/or audit purposes</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>44</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>36%</td>
</tr>
</tbody>
</table>
For the last question in this domain, i.e. whether the guideline presents key review criteria, there was again disagreement. Appraiser A2 with experience in the field was again not at all satisfied with the criteria provided. From the added statement of appraiser A3, it became clear that she did not understand what was meant by “review criteria for monitoring and/or audit purposes”. This might be explained by her experience as a practising midwife, but lack of experience with academic work and management processes, such as monitoring and auditing. The other two appraisers were indifferent about whether adequate, or the most suitable criteria had been provided.

Editorial independence

All four raters found that the guideline from England and Wales was editorially independent from the funding body. This is demonstrated well by the high ratings, which can be seen in Table 4.18. However, in an additional comment, appraiser A4 added that the reason for the high score was that too many organisations and individuals were involved in the guideline development process, as to be dependent. However, this appraiser noted also that there was no explicit comment with regard to independence from the funding body of the guideline development process. Appraiser A3’s comment again mirrors her lack of experience with academic conventions, as she stated that she could not find any reasons to believe that this guideline was funded by any companies. It is definitely worth considering that dependence from a funding body means more than funding by for example pharmaceutical companies.

As appraiser A3 made a statement that she could not find information about conflicts of interest, questions arise about why she rated this item 23 with a score of 2. This might have been a mistake, as it is not logical. In contrast, the score of 3 for this item as attached by appraiser A2 is plausible, as he expected that possible conflicts of interest were recorded, although he found that this was not explicitly done. However, appraiser A4 found that declarations of interest are reported on page 3, but that they are not in the book.
Table 4.18: Editorial independence of the guideline from England and Wales

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>22. The guideline is editorially independent from the funding body</td>
<td>3</td>
</tr>
<tr>
<td>23. Conflicts of interest of guideline development members have been recorded</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>67 83 67 100</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>79%</td>
</tr>
</tbody>
</table>

Table 4.19: Editorial independence of the German guideline

<table>
<thead>
<tr>
<th>Question</th>
<th>Individual scores by appraiser No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>22. The guideline is editorially independent from the funding body</td>
<td>1</td>
</tr>
<tr>
<td>23. Conflicts of interest of guideline development members have been recorded</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score type</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain score per appraiser (%)</td>
<td>0 50 50 0</td>
</tr>
<tr>
<td>Standardised domain score</td>
<td>25%</td>
</tr>
</tbody>
</table>

Again in this category, the German guideline ended with a low standardised domain score of 25% (Table 4.19). In this case, all four raters held the same opinion that potential conflicts of
interest of guideline development members have not been recorded. Despite this, appraisers A2 and A3 were convinced that the guideline is editorially independent from the funding body. Appraisers A1 and A4 were entirely convinced of the dependence from the funding body. How the first two appraisers came to their opinion is not clear, as they unfortunately did not comment on this item. In contrast to that, appraiser A4 commented that editorial independence was not even intended, as the guideline has been developed and issued by the self-governing body of physicians and health insurance funds. This means that this guideline has been issued by an obviously non-independent committee.

Overall assessment

In Tables 4.20 and 4.21 the appraisals of all four raters have been compiled to give an overview of the ratings in each domain for both guidelines. This has been done for three reasons. Firstly, this helps to get an overview about the particular strengths and weaknesses of each guideline. Secondly, the guidelines can be compared to each other. Finally, this strategy helps to compare the actual rating of the individual appraiser to the overall assessment, i.e. the final judgement about the quality of the guidelines they made.
Table 4.20: In-depth appraisal of the guideline from England and Wales by all four raters

<table>
<thead>
<tr>
<th>Domain</th>
<th>Question No.</th>
<th>Individual scores by appraiser No.</th>
<th>Sum</th>
<th>Standardised domain score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope + purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>48</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4</td>
<td>48</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>48</td>
<td>100</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>58</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td>58</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>4</td>
<td>58</td>
<td>88</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Rigour of development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Clarity + presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Applicability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Editorial independence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 4.21: In-depth appraisal of the German guideline by all four raters

<table>
<thead>
<tr>
<th>Domain</th>
<th>Question No.</th>
<th>Individual scores by appraiser No.</th>
<th>Sum</th>
<th>Standardised domain score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
</tr>
<tr>
<td>Scope + purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rigour of development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Clarity + presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Applicability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Editorial independence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
From the comparison of the in-depth appraisals it becomes clear that both guidelines have the highest score in the domain 'scope and purpose'. From the previous detailed analysis, it is known that both the appraiser without prior knowledge about antenatal care and the specialists were entirely satisfied with the description provided by the guideline from England and Wales. Although all appraisers had an idea about the scope and the purpose of the German guideline, it was difficult for them to extract the information. Moreover, the appraisers did not all draw the same conclusions from what they found in different sections of the guideline. These problems are reflected in the lower standardised domain score of 69% for the German guideline.

The domain 'clarity and presentation' was ranked with the second highest standardised domain score for both guidelines. Also in this domain, the guideline from England and Wales was superior to the German guideline. The appraisers found that the recommendations of the guideline from England and Wales are specific and unambiguous, and that the key recommendations are easily identifiable. Especially the separate page and the poster with the clinical algorithm were found helpful. When addressing the German guideline, the most interesting finding was that all appraisers with a background in obstetrics scored this domain significantly higher than the scientific assistant from another discipline. Background knowledge obviously helped to find the way through the guideline with a lack of structured presentation.

When addressing the two domains that relate to the selection of the guideline development group, the involvement of patients as well as to the scientific quality of the guideline, great differences are observable between the in-depth appraisals of the two guidelines. In the domains 'stakeholder involvement' and 'rigour of development', the guideline from England and Wales has particular strengths. Despite some criticism about the lack of piloting and an appropriate procedure for updating the guideline, the standardised domain scores are high for both domains with a rating of 88% each.

In contrast to that, the German guideline received its lowest scores in the domains 'stakeholder involvement' and 'rigour of development'. The perceived lack of rigour in guideline development was especially criticised. All appraisers with knowledge and experience about
academic work and about guideline development could not find the relevant information about the procedures used for identifying, selecting and using scientific evidence for setting up the recommendations. For all other questions which referred to the rigour of the guideline development process, the ratings were similarly low.

In addition, the appraisers were also not satisfied with the involvement of stakeholders in the guideline development process, which led to a standardised domain score of 25%. The ratings of the appraisers indicate an extremely critical view on the fact that patients' views were obviously not sought and that the guideline does not contain any hint as to whether it was piloted. A heterogeneous picture was evident for the question of whether all relevant professions were represented in the guideline development group. Whereas the obstetrician and the practising midwife were satisfied with the selection of the panel, the other two appraisers were not. This might indicate that the two professional groups involved with providing antenatal care see obstetricians as the outstanding experts for antenatal care. However, it became clear that representatives of other disciplines hold another view.

As can be seen in Table 4.20, the guideline from England and Wales has a standardised domain score of 69%, its weakest point, in the domain 'applicability'. It was found that no appraiser was entirely convinced by the properties of the guideline in this domain, although it became clear that the appraiser with a specialisation in monitoring and audit was more critical than the other appraisers. However, although the guideline from England and Wales could be improved in this domain, the standardised domain score is high enough to still regard it as satisfactory. Compared to the rating of the guideline from England and Wales, the standardised domain score of 36% of the German guideline is again disappointing. However, it has to be noted that this is not the weakest part of the German guideline itself.

The standardised domain score of 79% indicates that the appraisers were quite positive about the editorial independence from funding bodies of the guideline from England and Wales. Nevertheless there was some critique that declarations of possible conflicts of interest are reported, but that they are not directly available in the guideline. It is therefore impossible to
make a judgement about the factors which might have had an influence. The standardised
domain score of 25% for the editorial independence of the German guideline was a finding that
could be explained. As the German guideline has been developed and issued by the Federal
Joint Committee of the self-governing body of physicians and health insurance funds, editorial
independence from the funding body was obviously not even intended.

Does the overall assessment support the ratings of the individual appraisers?

To make a final statement about whether the appraisers would recommend the guideline they
assessed, the AGREE-instrument offers four categories: strongly recommend, recommend
(with provisos or alterations), would not recommend and unsure.

For the guideline from England and Wales, the overall assessment was ‘strongly recommend’
by all four appraisers. Appraiser A3 made the final comment that the guideline means a lot to
read, but that it is very clear and specific, and should be strongly recommended. Appraiser A2
also commented that he strongly recommends the guideline. The only criticism from this
appraiser was that it did not become clear from his point of view, how the implementation of
the guideline was planned, and how it should be evaluated. In addition, appraiser A2 describes
the guideline from England and Wales as being extensive and that the transfer into practice
appeared to be difficult at first sight. However, when he assessed the guideline in depth, he
found that the recommendations are easily comprehensible and finally not that difficult to
transfer into practice. Appraiser A2 also commented that it was not that simple to appraise
both guidelines as the terminology is difficult to understand for a nurse without any background
in obstetrics.

In accordance with the standardised domain scores of 79% and above, it can be concluded on
the basis of this small modelling exercise that the overall assessment of the appraisers to
strongly recommend the guideline from England and Wales is justified.

In contrast to the straightforward recommendation of the guideline from England and Wales,
appraisers A1, A3 and A4 came to the overall assessment that they would recommend the
German guideline with provisos or alterations. Appraiser A2 would not recommend the German guideline at all. This might be due to the fact that appraiser A2 has never used the guideline in practice, and has also no knowledge about its achievements. This is supported by the fact that appraiser A2 scored the domain applicability significantly lower than the other appraisers, which can be seen from Table 4.21. However, it might also be true that the other three raters are used to working with this guideline and have no idea about how antenatal care might be performed otherwise. However, appraiser A2 named the reasons for not recommending the guideline. These were that the development of the guideline was not clear, that there was no comment on the scientific basis of the recommendations, that the guideline is not patient-oriented and that it is mono-professional as well. He also found that the German guideline is not clearly arranged and therefore difficult to appraise. Although appraiser A2 has most likely not calculated the domain scores of his rating, his concluding remarks go in line with the results of his in-depth appraisal using the AGREE-instrument. As can be seen from Table 4.22, especially the domains stakeholder involvement, rigour of development and applicability were scored extremely low by this appraiser. With a score of only 8%, the domain clarity and presentation followed suit.

Table 4.22: Individual appraisers' and standardised domain scores of the German guideline

<table>
<thead>
<tr>
<th>Domain</th>
<th>Appraiser's individual scores (%)</th>
<th>Standardised domain score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
<td>A2</td>
</tr>
<tr>
<td>Scope and purpose</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>Rigour of development</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Clarity and presentation</td>
<td>92</td>
<td>8</td>
</tr>
<tr>
<td>Applicability</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Editorial independence</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>

Appraiser A3 with practical experience as midwife, but without prior knowledge about dealing with the theoretical background of guidelines finally stated she liked the guideline because it is
easy to read, very clear and not too long. As can be seen from Table 4.22, this is partially reflected by her ratings. Appraiser A3 rated the questions around the domain clarity and presentation higher than the standardised domain score. The other domains relating to her final comment, i.e. scope and purpose and applicability were scored lower than the standardised domain score, but obviously high enough to still recommend the guideline with provisos or alterations. In addition, she commented that she did not even notice the hints about additional texts when she was reading the guideline without AGREE-instrument, but would be interested in reading them. She also mentioned that she would be interested in the discussions that took place during the guideline development process and also in the opinions and experiences of patients. Moreover, appraiser A3 would like to know about the costs of what she called “special examinations”. However, despite the wish for such procedural information, appraiser A3 attached the highest rating of all appraisers to the domain rigour of development, which therefore lies also by far higher than the standardised domain score.

The overall assessment of appraiser A4 was that the German guideline recommends a traditional model of care, which has been used for ‘ages’ and is well known to practitioners as well as to pregnant women and their relatives. This has its own benefits. However, there is an obvious need for more documentation about the individuals involved in setting up the guideline, the development process and the underlying evidence. This becomes especially evident from the scores of 0% in the domains rigour of development and editorial independence, as well as from the score of 25% for stakeholder involvement. In addition to the previously mentioned comments, appraiser A4 stated that a framework for regularly updating the guideline is also needed. Moreover, this appraiser stated also that if there was not only a lack of documentation about who contributed to the guideline, the involvement of other relevant parties should be considered, at least that of consumers and midwives.

From the above findings it can be concluded that the questions from the in-depth appraisal with the AGREE-instrument guided the appraisers’ thoughts towards the crucial issues of guideline quality. This was reflected in the rating of the appraisers when using the AGREE-
instrument, but also in the final comments without having calculated the domain scores, and also without having the rating of other appraisers at hand to compare their views against.

### 4.2.2 Comparison of the recommendations of the two national guidelines

It is the aim of this part of the study to add knowledge about the link between the overall quality of a guideline, and the quality of its recommendations. As the AGREE-Collaboration claims that there is a relation between the overall quality of a guideline and the quality of its recommendations, two hypotheses were generated on the basis of the findings from the previous in-depth appraisal of the two guidelines. From this appraisal it was found that the overall quality of the guideline from England and Wales is higher than that of the German guideline. The difference was large enough to lead the appraisers to strongly recommend the guideline from England and Wales as it is, but to recommend the German guideline only with provisos or alterations. One of the appraisers even did not recommend the German guideline at all. On the basis of these findings, the hypotheses were formulated as follows:

**Hypothesis 1:** There are major differences between the recommendations of the guideline from England and Wales and the recommendations of the guideline from Germany.

**Hypothesis 2:** In case of different recommendations, the evidence-base suggests following the recommendation of the guideline from England and Wales.

For the comparison of the individual recommendations, the original guidelines were consulted. It is a clear advantage that both guidelines date from 2003, which implies that the available evidence was the same for both guidelines. As a consequence it is assumed that only the conclusions drawn from the evidence are different. Whenever only one of the guidelines recommended a test, the evidence-base is consulted to clarify what course of action the currently available knowledge suggests, or what kind of evidence is missing to make a conclusive decision. For this, the level of the underlying evidence for the test is extracted from the guideline from England and Wales, and especially the grade of the recommendation is
examined. The grading system of the guideline from England and Wales has been discussed extensively in the chapter on the evaluation of recommendations, and for a clear picture of it, the reader is referred to Tables 2.7 and 2.8 of the thesis.

Another important issue to consider was to compare not only the tests recommended by both guidelines for routine antenatal care, but also what both do not recommend. Also these tests are counted as identical ‘recommendations’ of both guidelines. The tables of the survey questionnaire, which made reference to all tests, which are recommended throughout the EU, were used as the basis for the comparison. As in the original tables, the tests are divided up into the categories ‘physical tests’, ‘technical tests’ and ‘laboratory tests’. All the tests are finally listed in Table 4.23, irrespective of when and how often they are recommended to be performed during pregnancy. The findings from this comparison are described in the following.
Table 4.23: Similarities and differences in the recommendations for routine antenatal care of the guidelines from England/Wales (E/W) and Germany (GER)

<table>
<thead>
<tr>
<th>Category of tests</th>
<th>Test</th>
<th>Recommended by</th>
<th>None*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical</strong></td>
<td>Blood pressure</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Body Mass Index</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Breast examination</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Foetal movement count</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Foetal position</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Formal risk scoring</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Full physical examination</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Fundal height</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maternal weight</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vaginal examination</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Technical</strong></td>
<td>Auscultation of foetal heart</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cardio-tocography</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Doppler Ultrasound</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ultrasound, abdominal</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ultrasound, transvaginal</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Laboratory</strong></td>
<td>Alpha-Feto-Proteine or Triple</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Atypical red cell antibodies</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blood group</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlamydia trachomatis</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Foetal fibronectin</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gestational diabetes OGTT</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gonorrhoea</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haemoglobin</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haemoglobinopathies</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hepatitis B virus</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hepatitis C virus</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HIV</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lues</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Papanicolaou smear</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Placental hormones</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhesus factor determination</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rubella titer</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Streptococcus group B</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Toxoplasmosis</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urinalysis / Bacteria</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urinalysis / Glucose</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urinalysis / Protein</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

* The category none refers to all routine tests identified from the critical review of the literature, which were also used in the survey questionnaire.
Physical tests

In the category 'physical tests', in 6 out of 10 cases both guidelines are identical. The guidelines from England and Wales and Germany recommend the measurement of maternal blood pressure. When assessing the recommendations in greater depth, it was found that both guidelines recommend this examination for each visit. Also recommended by both guidelines are determining the foetal position, measuring fundal height and weighing the mother. Not recommended are breast examinations and foetal movement count.

In contrast to the above measures, calculating the body mass index and formal risk scoring are recommended by the guideline from England and Wales only. However, instead of calculating the body mass index, the German guideline recommends measuring maternal height and weight. With regard to formal risk scoring, both guidelines contain catalogues of what risk pregnancies are, and in which cases more than routine care should be provided. However, in the German guideline this is not called formal risk scoring. In contrast to the guideline from England and Wales, the German guideline recommends a full physical examination in addition to taking the woman's history, and also provides catalogues of which conditions might cause risks. Taking these factors together, it is obvious that both guidelines lead the practitioners to determine the health status of a woman at the beginning of pregnancy.

The discussion about vaginal examinations during pregnancy is interesting. The guideline from England and Wales does not recommend this measure to predict preterm birth. This guideline makes a Grade A recommendation, which implies that the underlying scientific evidence is strong. In contrast to that, the German guideline recommends a 'gynaecological' examination at the first visit, which normally refers to a vaginal examination. However, this initial examination is not intended to assess the risk for preterm birth. Also in this case, the recommendations are nearly identical with regard to their practical implications. Drawing the above arguments together, it can be concluded that the recommendations of both guidelines are nearly identical with regard to the physical examinations.
Technical tests

In the category 'technical tests', the two guidelines give the same recommendations for four out of five tests considered. From the second part of Table 4.23 it is evident that both guidelines only recommend abdominal ultrasound for routine care. While the German guideline recommends three abdominal ultrasound examinations for normal pregnancies, the guideline from England and Wales recommends only two. In contrast to the German guideline, which recommends that the third examination should take place between the beginning of the 29th and the end of the 32nd gestational week, its English counterpart makes the Grade A recommendation that no routine use should be made of ultrasound scanning after 24 weeks. However, both guidelines agree again that cardio-tocography should only be performed if there is an indication to do so. Moreover, both guidelines do not recommend the use of Doppler ultrasound on a routine basis. The same applies to transvaginal ultrasound examinations.

Although it was found that the German guideline recommends the regular auscultation of the foetal heart rate, the England and Wales guideline reports good reasons for not doing so. This guideline explains that the auscultation may confirm that the foetus is alive, but that it is unlikely to have any predictive value. As this is a Grade D recommendation, there might still be some positive aspects to be discovered in the future. However, with the constant improvement of technological equipment, such as ultrasound scanners, this is unlikely. As the low predictive value of auscultation is likely to be also known to developers of the German guideline, they might have put more weight onto the second aspect of this examination, which is the reassurance of the mother. This was named as a valid reason to perform an auscultation by the England and Wales guideline.

Laboratory tests

For 19 out of 22 tests, the two guidelines make the same recommendations. As the overlap is that large, only those tests which need further comments are discussed in greater detail. The results for all other tests are provided in Table 4.23.
The first test, i.e. screening for Alpha-Fetoproteine, or a triple test mainly relates to screening for structural anomalies of the foetus. According to the guideline from England and Wales, screening for Down's syndrome should be offered to every pregnant woman. However, the triple test should be only offered in combination with other tests to provide the highest possible detection rate and in order to limit the number of false positive results. No reference is made to other structural anomalies, such as spina bifida. Also in the German guideline, reference is made to additional tests to screen for foetal genetic aberrations. However, in contrast to the policy to offer such testing to every pregnant woman, the German guideline only recommends testing if a risk for genetic aberrations was identified otherwise.

On routine screening for Chlamydia trachomatis, the guideline from England and Wales states that there is insufficient evidence on its effectiveness and cost effectiveness. However, the guideline states also that this policy is likely to change in the future with the introduction of the national opportunistic screening programme for all men and women under the age of 25 years. This can be interpreted as a sign of the guidelines becoming identical on this point. In contrast to that, both guidelines already recommend offering HIV-testing to pregnant women early in pregnancy. However, both guidelines state that the pregnant woman should make an informed choice about whether she wants to be tested or not.

A real difference between the two guidelines is the recommendation of the German guideline to perform urine dipstick tests to screen for glucose at each visit. According to the guideline from England and Wales, urine testing for glucose has a low sensitivity ranging from 7% to 46%, and a high rate of false positives. Therefore the conclusion was that it is not useful as a screening test. However, according to the discussion of the evidence in the guideline from England and Wales, urine testing for glucose has a high specificity ranging from 84 to 99%. The specificity is also reported to be higher than that of the 50g glucose challenge test. The policy suggested in the German guideline is to use dipsticks at every visit, but to reduce the number of false positives by performing a 75g oral glucose challenge test in case of positive results. Although this does not solve the problem of the low sensitivity of the test, it is an
acceptable decision to apply this cheap and non-invasive urine dipstick test for universal screening. This can be at least justified in health systems, where this investment does not prevent the introduction of more useful tests into the guideline. The decision of the German guideline developers might have been that it is better to pick up only a small proportion of the women with gestational diabetes, rather than providing no screening at all. Unfortunately, neither the guideline, nor any additional documents comment on this, and this leaves the users of the guideline to speculate about the reasons for this decision.

However, the decision of the guideline developers for the guideline from England and Wales to currently not recommend any screening for gestational diabetes at all can also be justified. The guideline reports that to date there is no consensus on the definition, management or the treatment of gestational diabetes mellitus. As an example, the definition of impaired glucose tolerance during pregnancy of the World Health Organization was criticised for its use of cut-off levels of non-pregnant women. In addition to the problem of clearly identifying women at risk, there is a lack of evidence that any of the currently available treatment options of the condition are effective in reducing the number of adverse outcomes. The guideline from England and Wales therefore draws the conclusion that the results of currently ongoing studies have to be awaited before any sort of screening should be introduced. It represents another, though also acceptable, strategy that the developers of the German guideline have decided in favour of cheap screening and an attempt to treat the condition, as long as more conclusive evidence suggests another course of action.

Summary

Drawing the findings from the comparison of the recommendations of both guidelines together it was found that 29 out of 37 (78%) of the tests are viewed identically by the guidelines from England and Wales and Germany. 15 tests are recommended by both guidelines, and 14 tests are not recommended by both guidelines. In addition to those identical recommendations, the guideline from England and Wales recommends 3 tests, which are not recommended by its German counterpart. The German guideline recommends another 5 tests, which are not
included in the guideline from England and Wales.

However, when examining the differences in greater depth it became clear that only very few real differences exist. Both guidelines recommend assessing maternal height and weight at the first visit. The guideline from England and Wales adds with calculating the body mass index only a mathematical model of interpretation. Similarly, both guidelines provide catalogues to identify pregnancies which carry additional risks. Only the German guideline does not call this formal risk scoring, and backs it up with a full physical and gynaecological examination. With regard to the auscultation of the foetal health rate, both guidelines use the same evidence, but draw different conclusions. While the German guideline focuses on the aspect of reassurance to the mother, the England and Wales guideline leaves the decision of whether she wants this examination up to the mother. However, it is most likely that the limited predictive value of the examination was known to both guideline development groups.

For the two other examinations recommended by the German guideline only, explanations could also be found, which relate to the currently limited evidence on these tests and different strategies of dealing with this. However, it is expected that these differences will vanish in the near future. As long as the underlying evidence is weak or ambiguous, different recommendations have to be accepted without necessarily judging one as more appropriate than the other. Drawing these findings together, it is concluded that as true differences, the following remain: the German guideline recommends a full physical examination and a gynaecological examination at the first visit, and urine dipstick tests for glucose. The guideline from England and Wales recommends the screening for Down's syndrome on a routine basis. As a consequence, the number of similar or identical recommendations is 33 out of 37 (89%).

Conclusion to the comparison of recommendations

Relating the findings back to the hypotheses, the results of the comparison of the guidelines' recommendations suggest rejecting both of them.

Hypothesis 1 is rejected because it was found that there are only minor differences between
the recommendations of the guideline from England and Wales and Germany. 89% of the recommendations were found to be similar or identical.

Also hypothesis 2 is rejected, as

in case of vaginal examination, the guideline from England and Wales refers to evidence about the predictive value of vaginal examinations to assess the risk for preterm birth. However, this is not the reason, for which the examination is recommended in the German guideline. This guideline recommends it as a surrogate to a routine preventive gynaecological check-up, as it is recommended for all women in Germany once a year.

the guideline from England and Wales does not provide any evidence about the benefits of a full physical examination at the beginning of pregnancy.

the guideline from England and Wales states that universal screening for chlamydia trachomatis is very likely to be offered in the near future for all women on a population basis. It will then interpret the evidence as the German guideline already does.

the evidence on screening for gestational diabetes can be regarded as inconsistent. Therefore it is acceptable to recommend either one of the possible policies, until more conclusive evidence is available.

How the above described findings might be interpreted will be discussed in the following chapter.
Chapter 5 - Discussion and conclusion to the study

5.1 Introduction

As the findings from the review of the state of the art and the related aspects have been critically discussed alongside the review process, and especially in sections 3.5 and 3.10 of the thesis, the final discussion is focused on the findings from the survey and the critical in-depth analysis and comparison of the two national guidelines. Despite this, it should become transparent, how the focal theory developed and how the different parts of the study contribute to reach the overall aim of the study. Therefore, in this concluding chapter the discussion is focused on how to integrate the findings of the study into theory and into what was previously known. It is aimed at demonstrating what can be learned from the study, and what the thesis adds to current knowledge, although a critical stance is taken regarding its potential weaknesses and limitations. Special care is taken to highlight implications for practice, as well as to make suggestions, what might better be backed up by further investigations.

5.2 The link between the methodological quality of guidelines and the quality of their recommendations

When critically analysing the current state of the art in assessing guideline quality, as was described in Part 1 of Chapter 3, the most important conclusion was that there is still the link missing between the overall quality of a guideline and the quality of its recommendations. To narrow this gap in knowledge, the methodological quality of the guidelines from England and Wales and Germany were appraised in-depth and their recommendations were compared, using a refined application of the AGREE-instrument (The AGREE Collaboration 2001).

From the literature review and the practical application of the AGREE-instrument it was found that this tool has introduced for the first time a measure for assessing the level, to which criteria indicating the methodological quality of a guideline are fulfilled. By this, specific weight
is added to the individual strengths and weaknesses of a guideline. This goes far beyond the simple recognition of the presence or absence of a criterion, which was common in earlier instruments to assess guideline quality. However, the dominating point of critique on the AGREE-instrument is that its calculations lead to a presentation of results, which seem to contradict basic arithmetic principles. The AGREE-instrument offers 1 as the lowest possible score for each question. As a consequence, for a domain consisting of three questions, the minimum score is 3. This is unfortunate, as it can also be argued that when adding a score of 1 out of 4 to an item, 25% of the points towards the rating for ‘strongly agree’ are already given, which can in the best case be called misleading. It would be more appropriate to provide a score of 0 for such cases, in which an appraiser wants to ‘strongly disagree’. This would be especially appropriate, as the category ‘strongly disagree’ also should be used if the relevant information is not provided by a guideline. Reducing this to the essence, it could be said: no information – no credits. Everything else is misleading, as a score of 1 could be interpreted that even in case of complete disagreement a rudimentary level of acceptance is still present. It is therefore suggested that this slightly illogical trait of the instrument should be remedied.

When using the provided formula for calculating a domain score with three items (The AGREE Collaboration 2001), a score of 12 equals 100%. The problem starts with the baseline value being 3, which equals to 0%. When using the formula for a score of 6, this equals to 33%. However, 6 out of 12 are normally 50%. However, the result is only arithmetically correct when using the formula together with the knowledge that 3 is the ‘background noise’. This is, however, out of the realm of investigation. It is especially regarded critical that it hampers the presentation of results such as to establish credibility. This is even more so, as the level of this ‘background noise’ is always different, depending on the number of questions asked in a domain. Due to this, variations of the problem occur, which are even more difficult to explain to a reader without deeper knowledge of the procedures behind the AGREE-instrument. It is therefore impossible to demonstrate results without explanations, as at first sight the domain scores seem to contradict basic arithmetic principles. By simply introducing 0 as the lowest score, the principles would not be violated, while the results of the appraisal are not distorted.
Yet another question arising from the use of the instrument for appraising the two guidelines is, whether the order, in which the guidelines were appraised, might have had an influence on results. This question arose as one of the appraisers wrote that he recommends the guideline from England and Wales even more, as it represents the exact opposite of the methodologically weak German guideline. It might therefore be problematic, when an appraiser appraises more than one guideline on the same topic. Effects, such as getting more or less rigorous in scoring might be observable. It might happen that an appraiser assessed a good guideline first, and appraises a weaker one more harshly than without having appraised the better first. Also the opposite is possible, if after appraising a weak guideline, the positive properties of a better one are overestimated. Another issue that needs clarification is whether differences in reporting and presentation have an influence on ratings. A neat presentation might be able to impress the appraisers, leading to a more generous scoring. As the AGREE-Collaboration provides no information about such effects, these are questions to investigate further.

The internal consistency of the domain ‘clarity and presentation’ of the AGREE-instrument

In contrast to that, a question that has arisen from the literature review, and which was not answered by the AGREE-Collaboration, can now be answered to some degree. In Chapter 3.3.4, the question was asked of whether the rating of the appraisers is sufficiently reliable, or whether it might be subjective. It was discussed there that the developers of the instrument calculated the Cronbach $\alpha$ coefficient for each quality domain to demonstrate the internal consistency (The AGREE Collaboration 2003). Intra-class correlations were calculated to assess the reliability within each domain, and it was found that the internal consistency ranged between 0.64 and 0.88 and was therefore acceptable for most domains. Nevertheless, the domain ‘clarity and presentation’ scored low with a Cronbach $\alpha$ of 0.69 and an intra-class correlation of 0.57. The effect was not explained by the AGREE-Collaboration. However, the findings from the in-depth appraisal of the two guidelines on antenatal care shed light on this. For the domain ‘clarity and presentation’ it was found that the appraisers with background
knowledge on the topic scored significantly higher than the appraiser from another discipline. Background knowledge might therefore help to decide whether the recommendations are specific and unambiguous. However, it might also be the case that appraisers with background knowledge interpret something into the appraised guideline, which is not written there. The use of the domain ‘clarity and presentation’ is therefore questionable, due to a lack of reliability. This is definitely an issue that has to be worked on by the AGREE-Collaboration, as it implies that the appraisal of guidelines is subjective on this issue, although an instrument is used.

Despite the above harsh criticism, the AGREE-instrument has its strengths in appraising the methodological quality of a guideline. It is especially useful to indicate to the appraiser the strengths and the weaknesses of a guideline, dissecting the individual domains in a structured manner. However, this is only the case, when a guideline and its development process are well documented, as has been stated by the developers of the instrument themselves (The AGREE Collaboration 2003). This would mean that the instrument of the AGREE-Collaboration has a lack of discriminating properties between guidelines of minor quality, which do not fulfil the set criteria, and guidelines with a lack of documentation. It can not be excluded that a guideline makes good recommendations, but that a lack of documentation leads to categorisation of a guideline as of minor quality. This is exactly what happened with the German guideline. Only the practical experience of the benefits of this guideline prevented three of the four appraisers from judging it as not at all recommendable. The appraiser without background knowledge would have withdrawn this guideline from practice.

Good guidelines = good recommendations?

As a consequence of the review of the literature and on the basis of the above critique on the AGREE-instrument, the question arose of whether good guidelines make good recommendations. To critically address the potential lack of discriminating properties between guidelines of poor quality and guidelines with a lack of documentation, the content of the guidelines from England and Wales and Germany were compared. After the in-depth appraisal of the two guidelines, this question of whether good guidelines make good recommendations
can be answered to some degree. From the in-depth appraisal it was found that both assessed
guidelines make equally good recommendations. The guideline from England and Wales,
which scored high in all domains and was therefore judged to be a good guideline – makes
good recommendations. However, the German guideline, with doubtful quality according to the
appraisal with the AGREE-instrument – makes equally good recommendations. These findings
can be interpreted as demonstrating that the methodological quality of guidelines and the
quality of their recommendations are not related in a manner, stable enough to permit
conclusions to be drawn about the quality of one in relation to the other.

However, the above findings have to be treated with caution, as they are derived from a
comparison of two guidelines at a national level. It can therefore be only claimed that at a
national level at least, the recommendations emanating from guidelines can be expected to be
of the highest quality, though their presentation and documentation might not necessarily
suggest this. Withdrawing the German guideline from practice would have been completely
inappropriate, as the comparison of the recommendations within the two national guidelines
clearly demonstrated that the recommendations of the German guideline are as good as those
of the guideline from England and Wales. It is therefore concluded that the methodological
quality of a guideline as assessed by the AGREE-instrument is not linked to the quality of its
recommendations. On the basis of this finding it is suggested that the speculations of the
AGREE-Collaboration about the potential link between them should be changed so as to warn
the users of the instrument not to draw preliminary and unjustified conclusions.

The warning not to draw conclusions about the content of guidelines by appraising their
methodological quality cannot be overestimated, as the AGREE-Collaboration recommends its
instrument as a tool for policy makers to help them decide which guidelines could be
recommended for use in practice (The AGREE Collaboration 2001). After the detection of the
weaknesses of the instrument, the additional comment of the AGREE-Collaboration is
regarded as a vital amendment. It states that the instrument should be part of a formal
assessment process in such an instance (The AGREE Collaboration 2001). However, this
amendment should be placed in a more prominent position. Moreover, it should be highlighted that the appraisal of the quality of a guideline's recommendations should form the core of such an appraisal, rather than being just surplus to the findings from the appraisal with the AGREE-instrument. Relating this back to the practical example as provided in this thesis, the German guideline would have been rejected if the recommendations were not assessed additionally, or if the guideline was appraised by individuals without experience in using the guideline only. As the guideline is universally implemented in Germany and makes recommendations which are as good as the evidence-based guideline from England and Wales, this would mean a wrong decision.

However, if the AGREE-instrument was used by the guideline developers, it could have helped to improve the guideline's structure, and would have therefore enhanced its credibility. The AGREE-instrument might therefore be more useful to stimulate good documentation of the guideline development process, rather than for the appraisal of a guideline's true quality, or for comparing two or more guidelines. This means that not only the instrument should be criticised for its methodological problems, but also the German guideline for not providing important background information. It is no longer appropriate to keep the underlying evidence used to set up a guideline in a cloud of mystery, claiming that experts have “done the thing right”. The movement for evidence-based care and quality in medicine might lead to a trend that guidelines with a lack of documentation and transparency will be replaced by more sophisticated ones in the future. However, it is also expected that this does not necessarily mean an improvement of the quality of the future guidelines' recommendations.

5.3 Limitations of the study and suggestions for future research

5.3.1 The categories ‘evidence-based guidelines’ and ‘expert-opinion-based guidelines’ are not appropriate at a national level

One of the questions to be answered from the in-depth appraisal of the guidelines from
England and Wales and Germany was how an evidence-based guideline on antenatal care and a guideline based on expert opinion on the same topic relate to each other. In contrast to the initial idea of the comparison, it was found from the critical in-depth appraisal that the two guidelines under scrutiny could not be put into these categories. It was found that the German guideline mainly suffers from a lack of documentation about how the developers of the guidelines reached their recommendations, e.g. how they searched for, selected and used the evidence. After the comparison of the recommendations of the two guidelines, it seems that the German guideline is not only based on the opinion of experts, but on scientific evidence. Instead, the issuing body covers the real developers of the guideline, as well as the procedures used. Hence, it contributes to speculation about the lack of standards behind the guideline development process. The lack of documentation does not mean that such activities have not taken place, and it can therefore not be concluded that the German guideline is based on expert opinion, rather than on scientific evidence.

Similarly to the problems in categorising the German guideline the label 'purely evidence-based', it is also not correct for the guideline from England and Wales. Although this guideline follows the principles of evidence-based health care, and provides a detailed track record of the underlying evidence, not all recommendations are based on conclusive evidence. This can be seen from the grades attached to the recommendations to demonstrate how strong the evidence-base is for each of them. In case of a recommendation for or against the screening for gestational diabetes, it was found that the current evidence allows both decisions, depending on what is expected from future research.

The study did therefore not contribute as much to the initial idea to critically appraise the relation between evidence-based guidelines and guidelines based on expert opinion, as was initially expected. However, this was outweighed by the critical theoretical discussion of guidelines and their quality, as well as by the discussion of decisions and their underlying evidence. In particular the discourse on how to weight the evidence to set up guidelines brought up important aspects. As a deviation from the initial idea it is now hypothesised that
guidelines at a national level are of good standard, irrespective of what is reported of the guideline development process. However, the critical in-depth comparison of two national guidelines is too small an investigation to draw valid conclusions about whether this applies to national guidelines generally, or whether this applies to the appraised two guidelines on antenatal care. Further research is therefore needed to test this hypothesis.

5.3.2 The link between the methodological quality of guidelines and their recommendations

From the previous discussion of the properties of the AGREE-instrument on the basis of the critical in-depth appraisal of the national guidelines from England and Wales and Germany, it was concluded that the methodological quality of a guideline as assessed by the AGREE-instrument is not linked to the quality of its recommendations. However, this conclusion might not be entirely justified. From the appraisal it was found that both national guidelines made equally good recommendations, irrespective of what was suggested by the appraisal with the AGREE-instrument. This means that in fact the methodological quality of the guideline from England and Wales and the quality of its recommendations are congruent. Only for the German guideline it was found that its poor methodological quality as assessed by the AGREE-instrument did not indicate the minor quality of the guidelines content.

As a consequence of these findings it can only be said that a guideline of poor methodological quality does not necessarily make inappropriate recommendations. The findings from the in-depth appraisal of the two guidelines therefore highlight the need for further research into the relationship between the methodological quality of guidelines and the quality of their recommendations. As the relationship was tested for two national guidelines only, comparisons of the recommendations of more guidelines and on other topics are needed to establish

- whether methodologically good guidelines always make good recommendations, or

- whether the quality of a guideline's recommendations is independent from the methodological
quality of the guideline itself.

Despite the need for further research, the innovative approach of extending the properties of the AGREE-instrument by using a purposefully selected sample of appraisers together with the critical comparison of the appraised guidelines' recommendations enabled insights into the relation between the methodological quality of guidelines and the quality of their recommendations, which have not been commented on in previously published work.

5.3.3 Guidelines are only an instrument for improving antenatal care

Within this study, a narrow operational definition of antenatal care was used. Only the baseline clinical care of healthy women with uncomplicated singleton pregnancies was considered. This routine antenatal care was further limited to care up to the estimated date of delivery. Within this narrow definition, the focus was additionally put onto screening tests. Preventive measures, such as the administration of anti-D to Rhesus negative women to prevent maternal iso-immunisation against foetal blood cells, or the prescription of folic acid before and in early pregnancy to prevent neural tube defects, were also not considered.

Despite this narrow focus on the diagnostic part of antenatal care, it is acknowledged that antenatal care is a complex intervention, also including the preparation for childbirth and parenthood, preventive measures and should also address maternal expectations from care as well as psycho-social needs. It is also perceived necessary to acknowledge that the living conditions of pregnant women as well as young families, such as housing, nutrition, physical exercise, their income and their general health status play a major role in ensuring the health of the future citizens of the European Union. However, addressing all the factors with a potential influence on the outcome of pregnancies would have led too far from what the study wanted to achieve, which was to make cross-border antenatal care safer and to contribute to a true European approach to antenatal care in the future. It is nevertheless recommended to further investigate into the other components of antenatal care.
Another potential critique could be that the focus was put on healthy women experiencing uncomplicated pregnancies, who are most likely to experience a good outcome of pregnancy without any intervention. This means that the study did not address conditions with the likely effect of lowering maternal and infant morbidity and mortality, such as gestational diabetes or premature birth. However, the focus of the study can on the one hand be seen as a first step to establishing co-operation at the European level with all the Member States, by not discussing controversial issues in the first instance. On the other hand, the focus of the study was to improve awareness of cross-border antenatal care, to highlight the importance of basic safety for the majority of pregnant women, before addressing specific aspects of care.

The same deviation from the aim of the study applies to the potential criticism that the focus has been put on guidelines, neglecting that on the one hand, the existence of guidelines does not necessarily mean that they are used, and on the other hand does not necessarily guarantee that the quality of the provided care is adequate. Guidelines are not an end in themselves, but a means to improve care. To develop their potential, guidelines need to be implemented. Although it is acknowledged that guidelines are not a stand-alone instrument for guaranteeing the quality of care, the investigator took the stance that it must be ensured that guidelines are of the highest possible quality, before they are implemented. The study was therefore directed at the gaps in knowledge about guideline quality, before setting off to address the implementation stage. However, the implementation of the developed minimum guideline is addressed in the conclusion to the thesis, and first steps have already been taken.

5.4 Contributions made by this study

A recent investigation into the definition of quality in social policy research found that for experts in the field, it is of central importance that research makes a contribution to policy and practice (Becker et al. 2006). However, for approximately half of the studied population it was found to be equally important that research contributes to the advancement of knowledge. An
effective synthesis of theory and knowledge was also seen as to be of critical importance, although a contribution to the advancement of theory was regarded as much less important compared to contributions to policy and practice and the advancement of knowledge. Despite this ranking in importance, it is aimed in the following sections to address the contributions, the thesis makes in all three areas.

5.4.1 Contributions to knowledge and understanding

Suggestions to improve the guidelines from England and Wales and Germany

With regard to the methodological quality of the guidelines from England and Wales and Germany it was found that according to the latest state of the art, the German guideline is no longer appropriate. It should be revised in order to enhance the identification of key messages and should be supported by a work-up set out in a graph of what has to be done at what time during pregnancy. However, the most impressive problem related to clarity and presentation of the German guideline was the complete absence of a clearly defined and stated aim. This means that there was no way of assessing whether the guideline had achieved its planned outcomes. Clarity and presentation have to be improved to ensure that all interested parties can get the message of the guideline. This is of special importance to newly qualified health professionals from this discipline, but also to ‘outsiders’ from agencies dealing with guideline quality, patients, and those who have to decide about policies and funding. Understanding the guideline as it is has proven to be difficult for those who are not experienced in using it. Especially against the background of increasing patient mobility and mobility of health professionals in the European Union, transparency of national guidelines becomes a key aspect for ensuring patient safety. In addition to improving clarity and presentation of the guideline's content, background information is needed to shed light onto what the underlying evidence of the guideline is, and how it is used. To ensure transparency, this information should at least be provided in an additional document. Also here, the German guideline can be improved substantially.
Despite its outstanding methodological quality compared to its German counterpart, the guideline from England and Wales can also be improved. As an example, the guideline urgently needs to fix a date for review. Otherwise, there might be the danger of adhering to a guideline, which is based on outdated evidence for too long. As the guideline from England and Wales acknowledged, substantially new knowledge is awaited for e.g. the diagnosis and the treatment of gestational diabetes. This is the main critique on the guideline from England and Wales. It is published as a book, containing all the background information as well as a clinical algorithm. However, the book is extensive, expensive and appears as if it were designed for eternity. It is unlikely that it will be thrown away by practitioners every time a new one is published. However, amending it by loose sheets of paper is also not a foolproof measure to ensure that all practitioners are kept up to date. The guideline from England and Wales therefore holds the danger of being used for much longer than it should be with regard to the rapidly growing evidence-base. Also for this guideline it might be useful to publish the clinical guideline separately from the procedural information and the underlying evidence.

A complete overview of guidelines on antenatal care in all EU member states

The aim of guidelines is to assure and maintain a high level of health care, in order to continually improve it to avoid outdated and possibly harmful interventions, which may also lead to unnecessary costs. Another major aim of guidelines is to keep up to date about the latest state of the art of their respective subject (The European Commission (n.d.) a). This need for information attains even more importance since patients as well as health care professionals make use of the right of free movement throughout the EU. Health care workers practising in a country other than that in which they were trained need guidelines to familiarise and integrate themselves into the respective national practice. Moreover, health care professionals need information when a non-resident of their country seeks care. Also patients, politicians and institutions financing health care need information about what care is provided in the other member states of the EU. This study provides such an overview for guidelines on antenatal care in the member states of the European Union for the first time. As it was seen as
important to make this contribution widely available, for the above reasons, the findings were published immediately in well respected scientific journals (Bernloehr et al. 2005, Bernloehr et al. 2007).

Before the commencement of the survey, a review of the literature revealed that there is a significant lack of knowledge about antenatal care in the member states of the EU. Moreover, only five publications were found, which compare antenatal care in Europe (Blondel et al. 1985, Hemminki & Blondel 2001, Heringa & Huisjes 1988, Langer et al. 1999, World Health Organization 1987). However, all of them were focused on the organisation of care, but not on the content of care, or the guidelines according to which care should be provided. None of these studies included all current member states of the Union.

The most comprehensive study found was that of the World Health Organization’s (WHO) study group on perinatal care in the WHO European region (World Health Organization 1987). Between 1981 and 1982, the group conducted a postal survey on routine antenatal care in 31 of the 33 countries of the European region. Unfortunately, the names of the countries are not given in relation to the findings, and the countries of the European region are not exactly the same as the current member states of the EU. Therefore, only general and limited information could be extracted. For example, one finding of the WHO study was that 18 out of the 23 countries have official guidelines on routine examinations that are recommended during pregnancy. The findings from the survey for this study were similar, in that 20 of the 25 member states had official guidelines. The WHO survey found that 12 of the 18 countries have guidelines issued by the state. The survey brought up that 13 member states of the EU have such a guideline. Six countries were reported by the WHO to have recommendations issued by major universities. In the survey, no such guideline was mentioned. In contrast to that, the survey found that four countries have guidelines published by the national society of obstetricians, and three countries having a guideline published by a governmental institution in co-operation with the society of obstetricians. However, most relevant to our study was the WHO's conclusion that even after their major survey very little information was available about
the actual content of antenatal care. This gap is now narrowed by this study as it provides the most comprehensive data set about the content of guidelines on antenatal care in the EU published to date. The publications are provided as Appendix 3 to the thesis (Bernloehr et al. 2005, Bernloehr et al. 2007). Making the current state of the art within the EU transparent is seen as a means to make decision-makers, health professionals as well as the recipients of antenatal care, aware of how their own care scheme relates to the current state of the art in the European Union. Moreover, potential gaps in care for pregnant women seeking antenatal care in different member states can be identified and eliminated effectively, which fundamentally contributes to the aim of the study, which is to help to make cross-border antenatal care safer.

However, the study also contributes towards the aims of the European Union. Making national guidelines on a specific health issue transparent and additionally developing a common minimum guideline on the basis of them stimulates co-operation and co-ordination amongst the member states of the EU. This is important, as co-operation and co-ordination were named as two major aims of the European Union in the health sector, which was discussed extensively in Chapters 3.7 and 3.8 of the thesis.

**Identification of outdated and of widely recommended measures**

Although guidelines are being widely used and accepted, not all recommendations for antenatal care are based on sound scientific evidence. This is well demonstrated in the guideline for England and Wales, in which recommendations are graded according to the strength of the evidence on which they are based (National Collaborating Centre for Women’s and Children’s Health 2003). In this publication, several recommendations are made with a clear statement that more information is needed and that the recommendation is either likely to be changed soon, or is provisional. Screening for Chlamydia trachomatis and the routine screening for group B streptococcus can be named as examples. Both are currently not recommended, but evidence is classified as insufficient. This might lead to a change in recommendations as more conclusive evidence is coming up. Another example is the
recommendation to take a urine sample on every occasion at which the blood pressure is measured. This is supported by limited evidence only, but it is included in the guideline. The publication of the findings from the survey, and especially the clear presentation of the measures recommended by more than half of the member states and for more than half of the EU population help national decision-makers to identify widely recommended, but also the outdated measures for antenatal care. It was also attempted to present the results in such a manner as to enable the reader to easily identify what each member state is recommending, and how this compares to the recommendations of the other member states.

As has been discussed extensively in Chapter 3.4 of the thesis, expert advice is vital to complement limited or contradictory evidence for making decisions in health care. This becomes even clearer when finding that one author claims that to date only 20% of all processes in medicine are evidence-based (Kreienberg & Berg 2002). The process of setting up a common minimum guideline on the basis of double majorities of what the national guidelines recommend played another important role to reach the aim of improving the starting conditions for the Union’s future citizens. With this, not only the widely recommended methods were established, but also outdated and possibly harmful interventions were identified. The issuing bodies of the current national guidelines are therefore encouraged to scrutinise their guidelines for measures which are not recommended by double majorities, and especially for those recommended by very few member states only.

5.4.2 Implications for practice

Recommendation of a common minimum guideline on antenatal care for the EU

In the Treaties of Maastricht and Amsterdam it is clearly stated that Community action shall complement national health policies in order to improve public health, prevent human illness and diseases and to obviate sources of danger to human health (Außen- und Finanzminister der Europäischen Gemeinschaft 1992, Generaldirektion Gesundheit und Verbraucherschutz
Co-operation between the member states was highlighted as very important, although organisation and delivery of health services remains entirely the responsibility of the member states (Paton et al. 2002). Although this prevents the superimposition of a single health care system to the member states, it nevertheless allows a common approach or guideline for a clearly defined public health problem.

Therefore, countries without a guideline might consider setting up a guideline similar to that of their European neighbours in addition to just comparing their approach to antenatal care to those of the other countries. Although it is acknowledged that the absence of a national guideline does not necessarily imply less quality of care, some advantages of having a national guideline can be important. Local protocols for antenatal care can be highly efficient for the local population, but they do not necessarily cover the needs of a rapidly changing population within the framework of the EU. In addition, transparency for care seekers as well as care providers is limited. Moreover, setting up and revising an unlimited number of similar guidelines continuously consumes more resources than revising one European guideline.

From the findings of this study it became clear that money is not a factor militating against the establishment of a common European guideline on antenatal care, as currently the less wealthy member states recommend more tests per pregnancy than the wealthier ones. This was even evident when the GNP per capita was adapted to the respective national price level according to the purchasing power parity. Potential reasons for this might be the fact that labour is cheaper in those member states, but also bureaucratic reasons, such as routines which are based on outdated protocols. However, the results of the study also provided evidence about the use of more expensive tests, rather than simply multiplying the cheaper ones. Despite this extensive antenatal care, which might be similar to the extent of care for other health problems, the overall level of resources invested in the health sector is in the new member states still much lower than in the ones of before May 2004. In recent years, the new member states invested on average around 4.5% of their GNP into health care, compared to 8.5% of the GNP in the EU-15 States (Commission of the European Communities 2004).
Another reason for the more extensive care might be that the movement for evidenced-based medicine is more advanced in the ‘older’ member states, which led to a reduction of examinations and tests with doubtful effectiveness. This mechanism might work hand in hand with a well established bureaucracy in the new member states. It might be that low efforts into evidence-based medicine, together with a highly developed bureaucracy and routine in meeting bureaucratic targets lead to an unjustified extension of antenatal care. In this respect, a common European guideline for antenatal care might help the less wealthy member states to save money without compromising the outcome of pregnancies in their countries.

Acknowledging all these facts, an EU-wide guideline on the minimum requirements for antenatal care in uncomplicated singleton pregnancies is recommended. This neither goes in the direction of the unwanted harmonisation of health systems, nor does it pose insurmountable financial hardships to the less wealthy member states of the Union. Introducing a common minimum guideline based on the current recommendations of the national guidelines would potentially mean lower costs for the less wealthy member states. The suggested common minimum guideline consists of the 23 tests, which are recommended by the national guidelines of at least 50% of the member states and which additionally apply to at least 50% of inhabitants of the Union. The thus identified tests are shown in Tables 5.1 to 5.3, again grouped into physical, technical and laboratory tests.

Table 5.1: Physical tests recommended for the EU-guideline

<table>
<thead>
<tr>
<th>Blood pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Mass Index</td>
</tr>
<tr>
<td>Foetal position</td>
</tr>
<tr>
<td>Formal risk scoring</td>
</tr>
<tr>
<td>Fundal height</td>
</tr>
<tr>
<td>Maternal weight</td>
</tr>
<tr>
<td>Vaginal examination</td>
</tr>
</tbody>
</table>
Table 5.2: Technical tests recommended for the EU-guideline

<table>
<thead>
<tr>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auscultation (foetal heart)</td>
</tr>
<tr>
<td>Ultrasound, abdominal</td>
</tr>
<tr>
<td>Ultrasound, transvaginal</td>
</tr>
</tbody>
</table>

Table 5.3: Laboratory tests recommended for the EU-guideline

<table>
<thead>
<tr>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-Feto-Proteine/Triple</td>
</tr>
<tr>
<td>Atypical red cell antibodies</td>
</tr>
<tr>
<td>Blood group</td>
</tr>
<tr>
<td>Gestational diabetes OGTT</td>
</tr>
<tr>
<td>Haemoglobin</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
</tr>
<tr>
<td>HIV</td>
</tr>
<tr>
<td>Lues</td>
</tr>
<tr>
<td>Rhesus factor</td>
</tr>
<tr>
<td>Rubella titer</td>
</tr>
<tr>
<td>Urinalysis / Bacteria</td>
</tr>
<tr>
<td>Urinalysis / Glucose</td>
</tr>
<tr>
<td>Urinalysis / Protein</td>
</tr>
</tbody>
</table>

As all but four of these tests are additionally supported by the most up to date scientific evidence, the above recommended minimum guideline for the EU has a sound basis grounded on the opinion of experts from all member states of the EU as well as on evidence (National Collaborating Centre for Women’s and Children’s Health 2003). For the four tests for which the evidence is not clear or does not support the routine use of a test to date, but which are recommended by double majorities, further investigation and expert discussion is required. These tests are vaginal examinations to predict a premature ripening of the cervix, the auscultation of the foetal heart rate, an oral glucose tolerance test to screen for gestational diabetes, and performing urinalyses for glucose.

Another important factor to consider is the timing and frequency, with which the individual tests should be performed during pregnancy. In addition to their presentation in Tables 5.1 to 5.3, all tests representing the mathematical consensus of double majorities were shown in Table 4.4 of the thesis. In this table, the wide variations with regard to the national recommendations are
shown. As an example, the maternal blood pressure is recommended to be measured about eight times during each pregnancy as a mean of all national guidelines. However, the individual recommendations for measuring the blood pressure vary between five and fourteen times during pregnancy. This indicates that although convincing, the suggested minimum guideline marks the beginning of a process, which might culminate in a consensus conference at which national representatives of the relevant institutions as well as individual health professionals can find a consensus, which can be finally accepted by all member states.

Another important fact to consider is that despite the potential benefits of the suggested minimum guideline, it is clearly a minimum guideline for healthy women with uneventful pregnancies. Specific health problems of individual member states of even regions must be additionally addressed by e.g. the national governments. States or regions with a high prevalence of haemoglobinopathies might set up screening programmes, which pick up such conditions. The same applies to states or regions with a high prevalence of sexually transmitted diseases, which are not covered by the minimum guideline. The issuing bodies of the current national guidelines are therefore advised to supplement the suggested common minimum guideline with the tests important for their respective populations.

5.4.3 Contributions made to further developing methodology

Suggestion of a new model for setting up guidelines on EU-level

From the reviews of the literature and of the state of the art it was concluded that evidence does not translate into guidelines by itself, although guidelines should be evidence-based. Moreover, it was reinforced that producing guidelines requires sound evidence, but also decisions. To gain insights into this act of policy, the literature was explored to greater depth. From the critical analysis of the literature on decision-making in health policy, it was found that the cornerstone in making sensible decisions is the way of dealing with inadequate, incomplete, and ambiguous evidence, as well as with the factors which were subsumed under
the header of ‘colloquial’ evidence as explained in the section on how to weight the evidence to set up guidelines. It was found that although scientific evidence should always be prioritised over other decision-modulating factors, the development of guidelines never takes place independent from values, habits and tradition, judgement, political factors, resources, pragmatics and contingencies, lobbyists and pressure groups, as well as experience and expertise of the decision-makers. Despite suggestions for using and weighing up different kinds of evidence to produce guidance in health care, such as the deliberative process as recommended by the Canadian Health Services Research Foundation (Canadian Health Services Research Foundation 2005 + 2006a), it was found that up to the present time there is no evidence to back up any model of decision-making, and that as yet there is no tool available for the production of guidelines at an international level.

Based on the findings from the reviews, it was perceived necessary to use the existing national guidelines on antenatal care from the member states as the basis to develop a model to set up a common minimum guideline for the EU. These guidelines represent the national decisions about what is wanted and acceptable for the respective populations, and what was decided to be appropriate in the light of the need to distribute limited resources. It would not be appropriate to disregard such important decisions that have been made at a member state level of the European Union. This applies at least as long as decisions cannot be kept value-free by using research-based strategies for weighing up the different kinds of evidence. Hence, a model is needed, by which national ideas about good antenatal care will be respected, but which also enables the utilisation of the potential benefits of a common minimum guideline.

To find out whether there are any models for decision-making at an EU-level, which are used to reach common decisions for the Union, the literature review was extended to decision-making in the Council of Europe. From this it was hoped to be able to develop a framework according to which the existing national guidelines on antenatal care could be used to efficiently set up a common minimum guideline, without compromising its quality. The findings of this review were demonstrated and discussed in detail in Chapter 3 of the thesis, but led to
the following result: A voting system by qualified majorities is the basic principle for decision-making in the Council of Europe. However, the current system is complicated and often judged as suffering from a lack of transparency and justice (SCADPlus (n.d.), Treaty of Nice 2001). It was therefore not suitable for the purposes of the study.

However, the proposed Constitution for Europe already seeks to improve the functioning of decision-making by reforming the system of qualified majority voting, while still considering the population sizes of the member states (Berbalk et al. 2005, SCADPlus (n.d.), The European Union Constitution 2004). The suggested system was called 'double majorities' and means that the majority of the member states must vote in favour of a decision and that they must represent the majority of the population of the EU. However, instead of mathematically defining a majority as more than 50%, the Constitution goes only half way and sets slightly different percentages as thresholds. This was judged to be a step in the right direction, but not sufficiently clear and easy to apply.

On the basis of the above findings, finally an alternative model was set up as a variation of official EU policy. However, it was decided that for the model of decision-making in the thesis, basic mathematic principles should not to be violated. Otherwise, it might neither be easily applicable, nor acceptable on a broad basis. In general terms, the model arose from the EU policy of double majority voting, attempts were made to eliminate the disadvantages of the EU model, making it more transparent and easier to apply. The suggested model for decision-making to set up European guidelines can therefore be regarded as a variation of official EU policy, developed from critically appraising the current state of the art in decision-making in the EU together with what was found from the comparison of the survey and the critical in-depth appraisal of the two national guidelines. In accordance with basic mathematics, the model of 'double majorities' for setting up European guidelines was defined as follows:

more than 50% of the member states need to vote in favour of a decision, and the decision must apply to more than 50% of the inhabitants of the EU.
The proposed new model of ‘double majorities’ is not only just, with an equal vote being given to every member state. It also acknowledges the specific problems of the countries, which have to serve a large population. That this model of decision-making might be acceptable on a broad basis was reinforced by a recent finding. During his presidency of the Council of Europe, Wolfgang Schüssel pledged in June 2006 to hold a plebiscite throughout all member states of the EU on the same day to pass the common constitution for Europe. In this context, Schüssel specified that the constitution should be accepted, if the majority of the European population and the majority of member states voted in favour of it (Falksohn et al. 2006, Heil 2006). The Austrian Chancellor’s proposal is covered fully by the findings of this study, which were already formulated at the time of his proposal.

The proposed model of true double majorities works for bringing national guidelines on antenatal care together into a common European approach, without compromising quality. Especially the finding that all but four measures, which are suggested by the national guidelines with double majorities are identical to a guideline based on the highest currently available level of scientific evidence supported the decision-making model. This is even more so as the evidence for the four remaining measures is not (yet) sufficiently conclusive to decide finally for or against the introduction of them into a common guideline. It is therefore acceptable to recommend them until more evidence is available, which is done in the suggested common minimum guideline on antenatal care as proposed by this thesis.

Drawing the theoretical considerations and the findings from all parts of the study together the above model of ‘double majorities’ is suggested for setting up common guidelines at EU-level. However, it has to be kept in mind that the model was developed on the basis of national guidelines, which were found to be of sufficient quality for such a process. One is therefore warned not to use the model uncritically on guidelines from sub-national level, without critically assessing the quality of the guidelines’ recommendations before. In addition, it is recommended to test the model on guidelines from other disciplines. However, although these limitations of the study need to be considered, it has to be acknowledged that for the
suggested model for setting up common minimum guidelines in the EU a range of evidence was triangulated with an extensive critical review of the state of the art. The suggested model is therefore well grounded in the current theoretical knowledge of how to develop guidelines of the highest quality, backed up with what was found from the investigation.

This approach of triangulation and mixing methods was taken on purpose in order to specifically tailor the methods to answer the research questions, rather than on the basis of the perception that a mixed methods approach is superior in itself. Using one of the methods alone would not have led to a complete picture of the problem. In this study, each source of data, i.e. the survey, the critical review of the state of the art and the in-depth appraisal of the two national guidelines contributes its part to develop the theoretical background for policy-making on antenatal care in the European Union. However, to use the potential of such a triangulation of methods, care was taken to achieve a real integration of findings. The findings are therefore initially presented separately, but were then used to make sense of the quantitative findings of the survey. This approach of one part of the study informing the other, enabled to gain more from the findings than would have been possible when using them separately.

Using this approach ensured the originality, i.e. the development of new theoretical and practical insights and concepts, as it not only brought together what the national guidelines recommend for antenatal care, but also how this compares to the published evidence. Moreover, the reasoning behind the guidelines was explored, especially acknowledging the decision-making necessary for setting up guidelines at a national level.

**Suggestions to substantially modify the AGREE-instrument**

AGREE, which stands for Appraisal of Guidelines for Research and Evaluation in Europe, is an international collaboration of experts from 13 European countries that was funded under the BIOMED-2 Programme of the European Commission (The AGREE Collaboration 2003). This collaboration developed a framework and an instrument for assessing the quality of clinical practice guidelines, by which a global assessment of an entire guideline can be achieved.
The AGREE-instrument has, for the first time, introduced a measure for assessing the level, to which criteria to assess the methodological quality of a guideline are fulfilled, and therefore adds a specific weight to the individual strengths and weaknesses of a guideline. This goes far beyond the simple recognition of the presence or absence of a criterion, which was common in the earlier instruments to assess guideline quality.

However, the AGREE-instrument has flaws, which were discussed extensively in Chapter 5.2 of the thesis. The most prominent point of critique is that the calculations of the domain scores lead to a presentation of results, which seem to contradict basic arithmetic principles. The AGREE-instrument offers 1 as the lowest possible score for each question. As a consequence, for a domain consisting of three questions, the minimum score is 3. This is unfortunate, as it can also be argued that when adding a score of 1 out of 4 to an item, 25% of the points towards the rating for ‘strongly agree’ are already given, which can be called misleading in the best case. It would be more appropriate to provide a score of 0 for such cases, in which an appraiser wants to 'strongly disagree'. This would be correct, as the category ‘strongly disagree’ also should be used if the relevant information is not provided by a guideline. It is therefore suggested that this illogical trait of the instrument should be remedied by simply introducing 0 as the lowest possible score. This would ensure that the principles would not be violated, while the results of the appraisal are not distorted.

Another weakness of the AGREE-instrument is the lack of information about potential effects when an appraiser is appraising more than one guideline on the same topic. Effects, such as getting more or less rigorous in scoring might be observable. Another issue needing clarification is whether differences in reporting and presentation of guidelines have an influence on ratings. It might be that a neat presentation is able to impress the appraisers, leading to a more generous scoring. It is desirable that the AGREE-Collaboration provides such information in the future. The same applies to the observed differences in the ratings of appraisers with and without background knowledge of the appraised guidelines' subject. It might be useful in this context to recommend a purposive selection of appraisers, rather than
simply recommending the number of appraisers to enhance the reliability of the appraisal. As could be demonstrated in section 4.2.1 of the thesis, the different rating patterns and comments of the appraisers with purposefully chosen backgrounds provide important additional insights into the properties of an appraised guideline.

Despite the above harsh criticism, the AGREE-instrument has its strengths in appraising the methodological quality of a guideline. It is especially useful to hint the appraiser towards the strengths and the weaknesses of a guideline, dissecting the individual domains in a structured manner. However, this is only the case, when a guideline and its development process are well documented, as has been stated by the developers of the instrument themselves (The AGREE Collaboration 2003). This is obviously the most serious point of critique.

By additionally comparing the recommendations of the two national guidelines for the thesis it could be demonstrated that the instrument of the AGREE-Collaboration has a lack of discriminating properties between guidelines of minor quality, which do not fulfil the set criteria, and guidelines with a lack of documentation. It was found that the German national guideline makes recommendations which are as good as those of the guideline from England and Wales, but that a lack of documentation leads to categorisation as a guideline of minor quality. It is therefore concluded that the methodological quality of a guideline as assessed by the AGREE-instrument is not linked to the quality of the recommendations made by the appraised guideline. On the basis of this finding it is suggested that the speculations of the AGREE-Collaboration about the potential link between them are changed into a profound warning so that the users of the instrument will not draw preliminary and unjustified conclusions.

The warning about drawing conclusions about the content of guidelines by appraising their methodological quality can not be overestimated, as the AGREE-Collaboration recommends its instrument as a tool for policy makers to help them decide which guidelines could be recommended for use in practice (The AGREE Collaboration 2001). After the detection of the weaknesses of the instrument, the additional comment of the AGREE-Collaboration is seen as a vital amendment. It states that the instrument should be part of a formal assessment process.
in such an instance (The AGREE Collaboration 2001). However, this amendment should be
placed in a more prominent position. Moreover, it should be highlighted that the appraisal of
the quality of a guideline’s recommendations should form the core of such an appraisal, rather
than being just a surplus to the findings from the appraisal with the AGREE-instrument. Also
the idea of using a purposefully selected panel of appraisers could be considered.

However, if the AGREE-instrument had been used by guideline developers, it could have
helped to improve the future guideline’s structure, and to therefore enhance its credibility. The
AGREE-instrument might therefore be more useful to stimulate a good documentation of the
guideline development process, rather than for the appraisal of a guideline’s true quality, or for
comparing two or more guidelines. It is therefore expected that guidelines with a lack of
documentation and transparency will be replaced by more sophisticated ones in the near
future. However, it is also expected that this does not necessarily mean an improvement of the
quality of the guidelines’ recommendations.

5.4.4 Contributions made to theory and theoretical understanding

A central ambition of this study was to further develop the theoretical understanding of
antenatal care at the European level. The theoretical background was intended to be
developed for making health policy on antenatal care in the European Union. This aim was
mainly tackled in Part 1 of Chapter 3 of the thesis. There, the underlying principles of guideline
development were explored and two major gaps in current theory were identified.

The first identified lack in knowledge was that the link between the overall quality of a guideline
and the quality of its recommendations, which is their actual content, is missing. The previous
section on the contributions made to further developing methodology already made reference
to this. Although there the focus was put on improving the AGREE-instrument, it nevertheless
became clear that there is no stable and reliable relationship between the overall, i.e. the
methodological quality of a guideline and the quality of the guideline’s content. The in-depth
appraisal using purposefully selected experts and the subsequent comparison of the recommendations of the guidelines from England and Wales and Germany could clearly demonstrate this.

The second identified lack in knowledge was the question of how to deal best with the different kinds of evidence for making evidence-based health policy. This was exhaustively explored in the review of the state of the art of the concepts within the theoretical framework. Within the framework of evidence-based policy making, the concepts of evidence-based health care and evidence-based policy were explored. Special care was taken to critically analyse the nature and the hierarchy of evidence, highlighting the role of expert opinion. However, from the analysis of the factors with the potential to modify decisions, it became clear that the crucial question in setting up guidelines as well as in making other health policy decisions is how to weight the different kinds of evidence. As in the theoretical discourse, the concepts are presented as they relate to each other, the reader is referred back to Part 1 of Chapter 3.

As a conclusion, from analysing the concepts within the theoretical framework, it became clear that the question of how to deal best with the different kinds of evidence is currently high on the agenda of institutions in health care, such as the Canadian Health Services Research Foundation. However, it became also clear that currently no model or strategy for weighing up the different kinds of evidence to produce guidance in health care has scientifically proven properties. For this it was suggested that it is most appropriate to develop a model for setting up guidelines on the European level using the national guidelines as the basis. This model has been presented under 5.4.3 as a suggestion of a new model for setting up guidelines on the European level.

5.5 Conclusion

In this study, the best available sources with regard to antenatal care within all member states of the European Union were critically analysed. By this, it was possible to establish for the first
time, what guidelines on antenatal care existed in the different member states. For this survey, a tool was developed and tested, with which it is possible to reliably extract the recommendations of guidelines on antenatal care. This tool might be used by other investigators in the future.

As another major step forward, it is now known what the member states of the European Union recommend for antenatal care. Before the survey, such data was not transparent. The detrimental effects of this lack of information could be demonstrated in the section on randomised controlled trials comparing different schedules of care. In this chapter trials were presented testing 'new' models, which were identical to the established care in another country. From the analysis of the current health policy of the EU it was found that antenatal care fits perfectly into the health aims of the EU. It addresses fundamental health determinants and has the potential to prevent ill health in the first place by improving the intrauterine starting conditions of the future citizens of the European Union. Making efforts in antenatal care transparent can also contribute to the alleviation of the current acceptance problem of the EU by addressing a health problem, which lies at the heart of Europe's citizens. Transparency of what care is recommended in the member states of the European Union is therefore seen as crucial for reaching the public health aims of the European Union as well as for making cross-border health care safe for pregnant women.

From the review of the literature and of the state of the art it was found that to date there was no tool available for setting up international guidelines, but the idea was reinforced that it is of critical importance to respect national decisions and values. These guidelines represent national decisions about what is wanted and acceptable for the respective populations, and what was found suitable in the light of the need to distribute limited resources. It was therefore concluded that the existing national guidelines on antenatal care from the member states should be used as the basis to develop a model to set up a common minimum guideline for the EU. From this process, two achievements emanated: a common minimum guideline on antenatal care for the member states and a model for setting up guidelines at European level.
From the findings of the survey it became clear that money is not a factor militating against the establishment of a common European guideline on antenatal care. It could be demonstrated that the member states with a Gross National Product below the EU-average recommend more tests per pregnancy than the more wealthy members. This was even evident when the Gross National Product per capita was adapted to the respective national price level according to the purchasing power parity. Although it is acknowledged that there are complex reasons for more intensive antenatal care recommended by the guidelines in the less wealthy member states, the findings of this study suggest that a common European guideline for antenatal care might help the less wealthy member states to potentially save money without compromising the outcome of pregnancies in their countries.

Acknowledging the findings from all parts of the thesis, an EU-wide guideline on the minimum requirements for antenatal care in uncomplicated singleton pregnancies is recommended. This neither goes in the direction of the unwanted harmonisation of health systems, nor does it pose insurmountable financial hardships to the less wealthy member states of the Union. For initiating discussion amongst the national experts of antenatal care, a common minimum guideline is suggested. It consists of the tests recommended by at least 50% of the member states and which additionally apply to at least 50% of inhabitants of the Union. As all but four tests within this guideline are additionally supported by the currently available scientific evidence, it has a sound basis grounded on the opinion of experts from all member states of the EU as well as on evidence. By these double majorities it was acknowledged that special care needs to be taken not to exclude the less wealthy member states or states with many citizens to care for from the potential benefits of a common minimum guideline. The four tests for which evidence is not clear to date, but which are recommended by double majorities of the national guidelines, require further investigation. Moreover, the timing and frequency of the individual tests need to be considered.

However, before addressing controversial issues, such as screening for gestational diabetes, a trusting relationship and co-operation needs to be established at the European level. A
common minimum guideline, based on double majorities of the national recommendations for antenatal care is an excellent starting point for the development of such processes. By integrating the decisions and values of the member states, the shaping of a common European set of values with regard to antenatal care is encouraged, opening up the discussion for a true European approach in the future. At present, there is a window of opportunity to introduce a common minimum guideline on antenatal care in the EU, which should be used to ensure the best starting conditions for the future citizens of the Community.
References


Allin S et al 2004 Making decisions on public health: a review of eight countries World Health Organization on behalf of the European Observatory on Health Systems and Policies, Brussels


Andersen et al 1989 [Preventive health examinations of pregnant women in Denmark. Structure and organisation] Ugeskrift for Laeger 151(49): 3304-08 [Danish]

Audit Commission of the National Perinatal Epidemiology Unit 1998 First class delivery. Improving maternity services in England and Wales. A national survey of women's views of maternity care Belmont Press, London


ÄZQ Zentralstelle der Deutschen Ärzteschaft zur Qualitätssicherung in der Medizin 1999 Checkliste Methodische Qualität von Leitlinien Zentralstelle der Deutschen Ärzteschaft zur Qualitätssicherung in der Medizin, Köln


Barker DJP 1995 Fetal origins of coronary heart disease British Medical Journal 311(6998): 171-74

Barker DJP et al 1993 Fetal nutrition and cardiovascular disease in adult life Lancet 341: 938-41


Bernloehr A et al 2007 National guidelines on antenatal care: a survey and comparison of the 25 member states of the European Union European Clinics in Obstetrics and Gynaecology Online First, January 05


Burgers J 2004 Personal communication via e-mail on 5 January

Busquin P 2002 Foreword In: Busse R et al (eds.) The European Union and health services: the impact of the single European market on member states IOS Press, Amsterdam pp v

Busse R 2002 Consumer choice of healthcare services across borders In: Busse R et al (eds.) The European Union and health services: the impact of the single European market on member states IOS Press, Amsterdam pp 231-48

Byrne D 2003 The European role in patient mobility HOPE conference on free movement and cross-border co-operation in Europe – the role of hospitals Speech/03/315, Luxembourg 20 June 2003

Canadian Health Services Research Foundation 2005 Final report: Conceptualizing and combining evidence for health system guidance Canadian Health Services Research Foundation, Ottawa

Canadian Health Services Research Foundation 2006a Weighing up the evidence: making evidence-informed guidance accurate, achievable, and acceptable Canadian Health Services Research Foundation, Ottawa


Carrol G et al 2001 WHO systematic review of randomised controlled trials of routine antenatal care The Lancet 357: 1565-70

Cluzeau F 2004 Personal communication via e-mail on 5 January


Commission of the European Communities 1998 Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions on the development of public health policy in the European Community COM/98/0230 final Commission of the European Communities, Brussels


D'Ubaldo C et al 1999 [Frequency and practice of prescribing prenatal HIV tests during pregnancy] [Italian] Minerva Ginecologica 51(9): 323-30 (abstract)

De Groote K et al 1997 Prevention of vertical transmission of hepatitis B virus infection. Is there a standard policy in Flanders (Belgium)? Acta Gastro-Enterologica Belgica 60: 255-58

Den Almindelige Danske Jordemoderforening 1998 personal letter to the author 9th November


Department of Health, Chief Medical Officer 2003 Public health in England Department of Health, London


Deutsches Ärzteblatt 2003 Kassen zahlen Behandlung im EU-Ausland: Europäischer Gerichtshof verwirft Genehmigungspflicht der Krankenkassen Deutsches Ärzteblatt 100(21): B1160

Dolowitz DP, Marsh D 2000 Learning from abroad: the role of policy transfer in contemporary policy-making Governance 13(1): 5-24

Doward J 2005 Parents demand gag on cot death doctor’s lectures: outrage at international acclaim for Meadow Observer 16 January


European Commission Directorate General for Employment, Industrial Relations and Social Affairs 1999 ‘Best practice’: state of the art and perspectives in the EU for improving the effectiveness and efficiency of European health systems Office for Official Publications of the European Communities, Luxembourg


European Health Management Association 2000 *The impact of market forces on health systems: a review of evidence in the 15 European Union member states* European Health Management Association, Dublin


Falksohn R et al 2006 Europa: der müde Erdteil *Der Spiegel* 19: 142-146

Forbes A, Griffiths P 2002 Methodological strategies for the identification and synthesis of ‘evidence’ to support decision-making in relation to complex healthcare systems and practices *Nursing Inquiry* 9(3): 141-55

Garcia J, Saurel-Cubizolles MJ 1983 Maternity care compared *Health and Social Service Journal* 93: 78-79

Generaldirektion Gesundheit und Verbraucherschutz (SANCO) 2000 *Jahresbericht 1999 (SANCO-2000-01153-00-00-DE-TRA-00)* SANCO, Brussels


Graham I 2004 European health policy: working within a paradox *Journal of Allied Health* 33(2): 139-43

Greenhalgh T et al 2003 Transferability of principles of evidence based medicine to improve educational quality: systematic review and case study of an online course in primary health care *British Medical Journal* 326: 142-45


Harrison S 2005, personal communication


Hemminki E, Gissler M 1993 Quantity and targeting of antenatal care in Finland Acta Obstetricia et Gynecologica Scandinavica 72: 24-30


Hunter DJ 1997 Desperately seeking solutions: rationing health care Addison Wesley


Kite M 2004 We can’t reunite thousands of mothers with children wrongly taken from them Sunday Telegraph 18 January


Kreienberg R, Berg D 2002 Disease-Management Mammakarzinom Frauenarzt 43(9): 1010-12


Kunz R et al 2000 Lehrbuch evidenzbasierte Medizin in Klinik und Praxis Deutscher Ärzte-Verlag, Köln


Kyprianou M 2005b Speech/05/400, 30/6/2005 at the Conference of German Health Ministers in Munich The European Commission, Brussels


Lekea-Karanika V et al 1991 Antenatal care received and its association with preterm birth in Greece Paediatric & Perinatal Epidemiology 5(3): 304-19

Lindmark G 1992 Assessing the scientific basis of antenatal care. The case of Sweden International Journal of Technology Assessment in Health Care 8(suppl 1): 2-7


Liu DTY et al 1992 Antenatal care towards the year 2000 Midwives Chronicle and Nursing Notes 12: 388-90


Mead M, Ashton R 1996 Activities, responsibilities and independence of midwives within the European Union, report European midwives-liaison committee, West Hunsbury


Merten M 2006 Leitlinien: unzureichend umgesetzt Deutsches Ärzteblatt 103(4): 143


National Collaborating Centre for Women’s and Children’s Health 2003 Antenatal care: routine care for the healthy pregnant woman ROCG Press, London


Neuberger J 1992 Ethics and health care: the role of research ethics committees in the United Kingdom King’s Fund Institute, London

NHS Centre for Reviews and Dissemination 2001 Undertaking systematic reviews of research on effectiveness: CRD report 4 The University of York NHS Centre for Reviews and Disseminations, York


Parsons W 1995 Public policy Edward Elgar, Aldershot

Paton C et al 2002 The European Union and health services: summary In: Busse R et al (eds.) The European Union and health services: the impact of the single European market on member states IOS Press, Amsterdam pp 1-13


Polit DF, Hungler BP 1999 Nursing research: principles and methods Lippincott, Philadelphia

Rabbata S 2003 Qualitätssicherung: vorhandene Instrumente nutzen. Ärzte kritisieren geplantes Zentrum für Qualität in der Medizin Deutsches Ärzteblatt 100(9): B449


South East Partners 2003 *EU policy update: public health* South East Partners Brussels Office, Brussels


The AGREE Collaboration 2001 *Appraisal of guidelines for research and evaluation (AGREE) Instrument* St. George's Hospital Medical School, London

The AGREE Collaboration 2003 Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project *Quality & Safety in Health Care* **12**: 18-23


The Danish Ministry of Health 1997 *Health care in Denmark* The Ministry of Health, Copenhagen


The European Commission 2004 Healthcare cooperation: patients to benefit from new Commission proposal (DN: IP/04/508) The European Commission, Brussels


The Ministry of Health 1997 Health care in Denmark The Ministry of Health, Copenhagen


Treaty of Nice 2001 Treaty of Nice amending the treaty on European Union, the treaties establishing the European Communities and certain related acts (2001/C 80/01) Official Journal of the European Communities 10.3.2001

Treaty of Nice 2001 Treaty of Nice amending the treaty on European Union, the treaties establishing the European Communities and certain related acts (2001/C 80/01) Official Journal of the European Communities 10.3.2001

University of Surrey 2001 Ethical guidelines for teaching and research University of Surrey, Guildford

Villar et al 2001 WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care The Lancet 357: 1551-1564


Wismar M et al. 2002 The European Union and health services – the context. In: Busse R et al (eds.) The European Union and health services: the impact of the single European market on member states. IOS Press, Amsterdam pp 17-29

Wismar M et al. 2002 The European Union and health services – the context. In: Busse R et al (eds.) The European Union and health services: the impact of the single European market on member states. IOS Press, Amsterdam pp 261-72


World Health Organization 1987 Wenn ein Kind unterwegs ist... Öffentliches Gesundheitswesen in Europa 26 World Health Organization, Kopenhagen


World Health Organization 2002 WHO randomized trial World Health Organization, Geneva

World Health Organization 2003 What is antenatal (or perinatal) care? What are its boundaries? What is the efficacy/effectiveness of antenatal care? What are the financial and organizational implications of antenatal care? World Health Organization Regional Office for Europe’s Health Evidence Network (HEN), Copenhagen

Appendix 1

Databases used for the review of the literature and of the state of the art
<table>
<thead>
<tr>
<th>Ref. No.</th>
<th>Name of database or search term</th>
<th>Abbreviation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ärztliches Zentrum für Qualität in der Medizin</td>
<td>AZQ</td>
<td><a href="http://www.aezq.de">http://www.aezq.de</a> <a href="http://www.leitlinien.de">http://www.leitlinien.de</a></td>
</tr>
<tr>
<td>3</td>
<td>Centre for evidence-based medicine, Oxford</td>
<td></td>
<td><a href="http://www.cebm.net">http://www.cebm.net</a></td>
</tr>
<tr>
<td>4</td>
<td>Cochrane Collaboration</td>
<td></td>
<td>University of Surrey's gateways <a href="http://www.cochrane.org">http://www.cochrane.org</a></td>
</tr>
<tr>
<td>5</td>
<td>Community Research and Development Information System</td>
<td>CORDIS</td>
<td><a href="http://www.cordis.lu">http://www.cordis.lu</a></td>
</tr>
<tr>
<td>6</td>
<td>Cumulative Index to Nursing and Allied Health Literature</td>
<td>CINAHL</td>
<td>University of Surrey's online databases</td>
</tr>
<tr>
<td>7</td>
<td>Department General for Health and Consumer Protection e-mailing service</td>
<td></td>
<td><a href="mailto:sanco-news@cec.eu.int">sanco-news@cec.eu.int</a></td>
</tr>
<tr>
<td>8</td>
<td>Deutsches Institut für Medizinische Dokumentation und Information, Köln</td>
<td>DIMDI</td>
<td><a href="http://www.dimdi.de">http://www.dimdi.de</a></td>
</tr>
<tr>
<td>9</td>
<td>Evidence Based Medicine</td>
<td>EBM</td>
<td><a href="http://www.evidence.de">http://www.evidence.de</a></td>
</tr>
<tr>
<td>10</td>
<td>International Confederation of Midwives</td>
<td>ICM</td>
<td><a href="http://www.internationalmidwives.org">http://www.internationalmidwives.org</a></td>
</tr>
<tr>
<td>11</td>
<td>Medical Literature On-Line</td>
<td>MEDLINE</td>
<td>University of Surrey's online databases</td>
</tr>
<tr>
<td>12</td>
<td>Midwives Information &amp; Resource Service</td>
<td>MIDIRS</td>
<td><a href="http://www.midirs.org">http://www.midirs.org</a></td>
</tr>
<tr>
<td>13</td>
<td>National electronic Library for Health</td>
<td></td>
<td><a href="http://www.nelh.nhs.uk">http://www.nelh.nhs.uk</a></td>
</tr>
<tr>
<td>15</td>
<td>Official websites of the EU</td>
<td></td>
<td><a href="http://www.europa.eu.int">http://www.europa.eu.int</a></td>
</tr>
<tr>
<td>16</td>
<td>Royal College of Obstetricians and Gynaecologists</td>
<td>RCOG</td>
<td><a href="http://www.rcog.org.uk">http://www.rcog.org.uk</a></td>
</tr>
<tr>
<td>17</td>
<td>Scottish Intercollegiate Guidelines Network</td>
<td>SIGN</td>
<td><a href="http://www.nhshealthquality.org">http://www.nhshealthquality.org</a></td>
</tr>
<tr>
<td>18</td>
<td>Summaries of the Union’s legislation</td>
<td>SCADPlus</td>
<td><a href="http://europa.eu.int/scadplus/scad_en.htm">http://europa.eu.int/scadplus/scad_en.htm</a></td>
</tr>
</tbody>
</table>
Appendix 2

Survey questionnaire, covering letter and consent form
European Study

ANTENATAL CARE IN THE EUROPEAN UNION:
EQUAL CHANCES FOR ALL NEW CITIZENS OF THE COMMUNITY

Dear Sir or Madam,

As a person professionally involved in the health sector of your country, you are well aware of the importance of international studies to address major health issues. Within the framework of a project on antenatal care in the Member States of the European Union, we would therefore like to ask you for your co-operation.

The project aims at making cross-border health care safer for pregnant women. It is based on the premise that despite growing awareness for trans-national health issues there is still a major lack of transparency and knowledge about care in the neighbouring countries. In order to narrow this gap in knowledge, we decided to conduct a survey on guidelines on antenatal care in the Member States of the European Union.

The study has been developed as a PhD project at the University of Surrey in cooperation with the University of Bielefeld. It is closely supervised and is anonymous and confidential. If you have any further questions, please do not hesitate to contact us under the e-mail addresses provided in the header of this letter.

In order to take only a minimum of your limited time, the questionnaire has been designed to need only about 15 minutes to be filled in. Please have a look at the brief instructions and the survey sheet, and contribute to the ambitious aims of our project. Please find also a consent form to be signed. In order to stay within the limited time frame of the project, we would be grateful to receive your answer until 30 May 2004. Many thanks for your co-operation.

Yours sincerely

Annette Bernlöhr, MSc
ANTENATAL CARE IN THE EUROPEAN UNION:
EQUAL CHANCES FOR ALL NEW CITIZENS OF THE COMMUNITY

How to use the questionnaire?
The attached questionnaire is specifically designed to capture the content of national guidelines on antenatal care quickly and easily. If you are not the right person to fill it in, please pass it on to a person who is occupied with antenatal care, or clinical guidelines.

Please make sure that you use your official national guideline on antenatal care for normal pregnancies as the basis to fill in the questionnaire. Using data from local protocols or information about your own practice or views would substantially compromise the study. In case that there is no national guideline available in your country, please fill in questions 1 to 3 of the questionnaire and return it.

If a national guideline is available, you are kindly invited to fill in the entire questionnaire. Please make sure that you use the guideline that specifies care for normal uncomplicated singleton pregnancies in healthy women up to the estimated date of delivery.

Please use the three tables attached in order to specify which measures are recommended in your national guideline. As the time in pregnancy at which the test should be performed is crucial, please tick the measure in the box relating to the respective gestational week. Please use the column "1st visit" for all tests that should be performed at the first visit, or as early as possible during pregnancy. If a test is to be performed at each visit, please tick the boxes according to the routine pattern of visits that is recommended, starting from the earliest week at which the pattern starts. This does not interfere with the tests marked under "1st visit".

Example: If blood pressure measurement is required at week 12 mark "X" in this column. If it is optional at week 18, mark "O" in the respective box. If you want to indicate a range, mark "X - - - X" as shown in the table for weeks 24 to 28.

<table>
<thead>
<tr>
<th>Test</th>
<th>1st visit</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6  8  10</td>
<td>12  14  16  18  20  22  24  26  28  30  32  34  36  38  40</td>
</tr>
<tr>
<td>Blood pressure</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

If there are different guidelines for 1st and subsequent pregnancies, please copy tables 1 to 3 and attach these to the questionnaire.

If there are any problems involved with filling in the questionnaire, if you like to receive it in another language, or if you wish to have more detailed information about the study, please feel free to contact me by e-mail: annette.bermioehr@t-online.de

By filling in and returning the questionnaire, you give consent to use the data according to the aims of the study.

Many thanks for your co-operation!
QUESTIONNAIRE ON NATIONAL GUIDELINES ON ANTENATAL CARE

1. The guideline applies to the following country (please tick)

A  B  DK  SF  F  D  GR  IR  I  LUX  P  E  SW  NL  UK

2. There is a national guideline on antenatal care

☐ No
☐ Yes. Please give the full reference under which the guideline is available, or send a copy of the guideline

3. Your institutional affiliation and the issuing body of the guideline

<table>
<thead>
<tr>
<th>Institution</th>
<th>I work for</th>
<th>Has issued the guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ministry of Health</td>
<td></td>
<td></td>
</tr>
<tr>
<td>National society of obstetricians</td>
<td></td>
<td></td>
</tr>
<tr>
<td>National society of midwives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Physical tests

<table>
<thead>
<tr>
<th>Test</th>
<th>1st visit</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood pressure</td>
<td></td>
<td>6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast examination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fetal movement count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fetal position</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal risk scoring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full physical examination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundal height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal examination</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Technical tests

<table>
<thead>
<tr>
<th>Test</th>
<th>1st visit</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auscultation of fetal heart rate</td>
<td></td>
<td>6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40</td>
</tr>
<tr>
<td>Cardio-tocography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doppler Ultrasound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasound, abdominal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasound, transvaginal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Laboratory tests

<table>
<thead>
<tr>
<th>Test</th>
<th>1st Visit</th>
<th>gestational week at which the test is recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-Feto-Protein or Triple</td>
<td></td>
<td>6  8  10  12  14  16  18  20  22  24  26  28  30  32  34  36  38  40</td>
</tr>
<tr>
<td>Atypical red cell antibodies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia trachomatis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fetal fibronectin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational diabetes OGTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gonorrhoea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobinopathies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis C virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papanicolaou smear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placental hormones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus factor determination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rubella titer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptococcus group B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis / Bacteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis / Glucose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis / Protein</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Consent form

Antenatal care in the European Union: Equal chances for all new citizens of the Community

I the undersigned voluntarily agree for this organisation to take part in this study.

I have read and understood the information provided. I have been given a full written explanation by the investigators of the nature and purpose of the study, as well as of what I will be expected to do. In addition to that, I have been given contact details under which I have the opportunity to ask questions on all aspects of the study.

I understand that the data are strictly anonymous and the information is provided in confidence. I therefore agree that I will not seek to restrict the use of the results of the study on the understanding that my anonymity is preserved.

I confirm that I have read and understood the above and freely consent to participating in the study.

Name: ________________________________

Position: ________________________________

Workplace address: ________________________________

Signed: ________________________________

Date: ________________________________
MATERIAL REDACTED AT REQUEST OF UNIVERSITY