Diagnosis and Management of Depression in Three Countries: Results from a Clinical Vignette Factorial Experiment

Carol L. Link, PhD
Theodore A. Stern, MD
Rebecca S. Piccolo, ScM
Lisa D. Marceau, MPH
Sara Arber, PhD
Ann Adams, PhD
Johannes Siegrist, PhD
Olaf von dem Knesebeck, PhD
John B. McKinlay, PhD

1 New England Research Institutes, Watertown, MA, USA
2 Massachusetts General Hospital, Boston, MA, USA
3 Centre for Research on Aging and Gender, University of Surrey, Guildford, Surrey, UK
4 Warwick Medical School, University of Warwick, Coventry, UK
5 Department of Medical Sociology, Heinrich Heine-University of Dusseldorf, Dusseldorf, Germany
6 Department of Medical Sociology and Health Economics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author
Carol L. Link
New England Research Institutes
9 Galen Street
Watertown, MA 02472 USA
+1 (617) 972-3205
Fax: +1 (617) 926-8246
clink@neriscience.com

Reprint Requests
John B. McKinlay
New England Research Institutes
9 Galen Street
Watertown, MA 02472 USA
+1 (617) 972-3012
Fax: +1 (617) 926-8246
jmckinlay@neriscience.com

Published in: Primary Care Companion CNS Disorders, 2011, 13 (5)

Funding Source: This study was funded by the National Institutes of Health, National Institute on Aging, AG16747, Bethesda, MD, USA. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

Running Head: Diagnosis and Management of Depression
Abstract

Objective: International differences in disease prevalence rates are often reported and thought to reflect different lifestyles, genetics, or cultural differences in care-seeking behavior. However, they may also be produced by differences among health care systems. We sought to investigate variation in the diagnosis and management of a “patient,” with exactly the same symptoms indicative of depression, in three different health care systems (Germany, the United Kingdom, and the United States).

Method: A factorial experiment was conducted in which 384 randomly selected primary care physicians viewed a video vignette of a “patient” presenting with symptoms suggestive of depression.

Results: Most physicians listed depression as one of their diagnoses, but German physicians were more likely to diagnose depression in women, while British and American physicians were more likely to diagnose depression in men (p=.0251). American physicians were almost twice as likely to prescribe an anti-depressant (p=.0241) as British physicians. German physicians were significantly more likely to refer the patient to a mental health professional (p<.0001) than British or American physicians. German physicians wanted to see the patient in follow-up sooner (p<.0001) than British or American physicians.

Conclusions: Primary care physicians in different countries diagnose exactly same symptoms of depression differently, depending on the patient’s gender. There are also significant differences between countries in the management of a “patient” with symptoms suggestive of depression. International differences in prevalence rates for
depression, and perhaps other diseases, may in part result from differences among health care systems in different countries.

Keywords: primary care, international comparison,
**Introduction**

International differences in the diagnosis and management of diseases are of increasing interest to epidemiologists and health services researchers. Reported country differences in disease prevalence rates are assumed to be real, and the ensuing search for underlying reasons has focused on family history and genetics, and on cultural differences in life styles and risk behaviors. However, it is equally possible that the same signs and symptoms of disease are diagnosed and managed differently in different countries (or healthcare systems), hence the international differences in disease prevalence rates (Table 1)\(^1\)\(^-\)\(^4\). Variability in provider behavior is increasingly invoked as an explanation for within-country health disparities; however this explanation is seldom extended to an understanding of international health and medical care variations\(^5\), \(^6\). Furthermore, as healthcare reform is debated in the United States there is interest in lessons to be learned from the way other countries organize and finance their health care systems, and how this affects health outcomes and the costs of care\(^7\)\(^-\)\(^9\). International comparative health services and policy research remains limited by reliance on existing medical records with their well-known deficiencies: they are collected for various administrative and billing purposes; they may not reflect what actually occurs during any clinical encounter (e.g. legal and financial pressures to “code up”), and subtle processes of medical care are seldom captured in the medical record (e.g. the number and types of questions asked). Sophisticated statistical analyses of such data can never produce un-confounded estimates of relative contributors to health care variations. Moreover, differences among patients
and in the severity of symptoms presented for diagnosis and management always bedevil the explanation of health care variations.

In contrast to earlier observational research we conducted an experiment to simultaneously estimate the un-confounded effect of patient attributes (gender, age, race, and socioeconomic status), physician characteristics (gender and years of clinical experience) and country setting (health care system) on the diagnosis and management of a “patient” presenting with signs and symptoms suggestive of depression. Our factorial experiment sought to address the following questions:

1. Do primary care doctors in Germany, the UK and the US diagnose patients differently when exactly the same symptoms suggestive of depression are present?

2. Is the clinical management of depression different between the three countries and if so, what may explain any such differences?

**Methods**

Our factorial experiment was conducted between 2001 and 2006 in three countries with very different healthcare systems: Germany with decentralized care administered by social security agencies; the United Kingdom (UK) with its government supported, tax-based National Health Service; and the United States (US) with a largely employment-based private insurance system. Randomly sampled primary care doctors from each of
these three countries viewed a clinically authentic videotape of a “patient” presenting 
signs and symptoms strongly indicative of depression.

Under the supervision of experienced clinicians, professional actors were trained to 
realistically portray a patient who presents with the signs and symptoms of depression. 
Vignette content was developed with input from experienced clinicians. Patients 
presented with seven symptoms of depression (sleep disturbance, decreased interest, 
guilt, diminished energy, impaired concentration, poor appetite, and psychomotor 
agitation or retardation); suicidal ideation was omitted as being too indicative of 
depression (Table 2)\textsuperscript{10}. Video presentations permit the inclusion of often subtle non-
verbal clues (e.g. a dejected affect and anxiety). Since patients seldom present as 
textbook cases, several other symptoms (e.g. diffuse pain, dyspnea) were also included, 
to increase the authenticity of the case. Patients depicted in the vignette spoke directly to 
the camera and to an unseen physician. The vignettes varied five factors: (1) patient age 
(55 or 75), (2) patient gender, (3) patient socioeconomic status (depicted by their current 
or former occupation as a janitor (cleaner) or school teacher), (4) patient race (black or 
white) in the UK and US, but not in Germany because black patients are infrequently 
encountered and (5) unseen physician gender. In the UK and US, the design was a half 
replicate of a $2^5$ factorial with 16 vignettes. In Germany, the design was a full replicate 
of a $2^4$ factorial, also 16 vignettes. For the vignettes viewed in the UK, patients and the 
unseen physician had British accents and vocabulary, while an American accent and 
vocabulary was portrayed by the same actors in the vignettes viewed in the US. The 
vignettes were dubbed into German for vignettes viewed in Germany. The physicians are
asked open-ended questions such as: would you refer the patient, would you prescribe any medication, or would you recommend any lifestyle change? Additional details concerning this experimental approach and the use of clinical vignettes are presented elsewhere\textsuperscript{5, 6}.

The physician subjects were randomly selected from country specific lists of physicians. The physicians were told that the study was investigating clinical decision making. One of the two video vignettes presented a patient with symptoms suggestive of depression and the other presented a patient with symptoms suggestive of coronary heart disease (not included here). In Germany, generalist or family physicians practicing in the North Rhine – Westphalia region were sampled. In the UK, general practitioners were sampled from health authority lists in the Midlands or Surrey (half of the interviews were in each area). In the US, internists and family practitioners were sampled from physicians practicing in Massachusetts. The physicians were recruited into four strata representing gender by two levels of clinical experience. Experience was defined by the year of their graduation from medical school in the UK and US, and the year of licensure in Germany. Less experienced physicians were recognized as having graduated from medical school (in the UK or US) between 1989 and 1996 (or who were licensed in Germany between 1998 and 2004) and more experienced physicians graduated from medical school (in the UK or US) between 1965 and 1979 (or who were licensed in Germany between 1974 and 1988). Given our interest in comparing physician behavior in the three countries, subjects were only eligible if they had completed their medical training in the country in which they practiced. The balanced design required 128 physician subjects from each
country; 32 of each gender by experience strata with two to view each of the 16 vignettes; there were a total of 384 physicians. Phone calls were made to randomly-selected physicians to screen for eligibility. Interviews took place in the physician’s office, generally “sandwiched” between appointments with real patients. Physicians received a modest financial acknowledgment for their participation (€100 in Germany, £50 in the UK, and $100 in the US). All physicians signed informed consent prior to their participation. The study was approved by New England Research Institutes’ Institutional Review Board and the ethics committees in Germany and the UK.

To preserve the orthogonal (independent) associations of patient factors, physician factors, and the country of care (or health care system) factor from our balanced factorial design, we used an analysis of variance (ANOVA) for both dichotomous and continuous outcomes. The error term was the residual after specifying the full factorial model (pure error due to replication) with 192 degrees of freedom. While logistic regression might have been more appropriate (technically) for dichotomous outcomes, we used ANOVA for several reasons: (1) the full factorial model could not be specified using logistic regression due to data sparseness (with only two observations in each of the 16x3 cells there were likely to be many cells with no data when considering a dichotomous outcome 2x16x3 = 364 cells (our sample size); (2) this Fisherian regression was equivalent to discriminant analysis; (3) the models were equivalent due to the Central Limit Theorem; (4) comparison of p values from a full factorial ANOVA model, an ANOVA model with main effects and two-way interactions, and a logistic regression model with main effects and two-way interactions, we found similar values; and (5) ANOVA results
appeared to be more easily understood. Since we were primarily interested in the main effect due to country of care (or health care system) and the two-way interactions of country and patient/physician factors and since there were no significant main effects due to patient race for the two country analysis (UK and US), we considered a full factorial model including patient age, gender, and socioeconomic status, audio gender of the unseen physician, physician gender and experience, and country of care. The only exception was when we considered two-way interactions of country of care and patient race, in which we restricted the analysis to two countries using the full factorial model specified above with the additional of a factor for patient race. Finding a statistically significant (p<0.05) effect due to country, we employed Tukey’s studentized range multiple comparisons\textsuperscript{12} to determine which countries were significantly different. Given our sample size of 384, we had 80% power to detect an absolute difference of 17% between the means of three groups (e.g. 20% of physicians from country A versus 37% of physicians from country B would ask questions about alcohol).

**Results**

The response rates were similar in the three countries: 65.0% in Germany, 64.9% in the US, and 59.6% in the UK. Table 3 (information gathering and diagnosis) and Table 4 (management) summarize the main effect due to country of care (or health care system). The p value from the ANOVA is given; along with Tukey’s multiple comparisons (means with a common letter superscript were not significantly different at the .05 level).
Significant interactions are indicated by a letter in the last column of Tables 3 and 4 and presented graphically in Figure 1.

1. Information Gathering (Table 3)

American physicians asked more questions than their European counterparts. The significant two-way interaction of country of care and patient gender (p=.0342, Figure 1A) was due to the propensity of American physicians to ask more questions of women, while European physicians asked more questions of men. Since American physicians were likely to ask the most questions, they were also the most likely to ask about pathology, medical history, pain, alcohol, the patient’s psychological state, their social situation, and more general questions. German physicians asked the fewest questions and were least likely to inquire about these items. The main effect due to country of care related to questions about alcohol use was modified by a two-way interaction of country and patient race (p=.0476, Figure 1B) where British and American physicians were equally likely to ask white patients about alcohol; if the patient was black, British physicians were less likely to ask them about alcohol. British physicians were least likely to want to perform a complete physical (11.7%) than were American (39.0%) or German (46.9%) physicians.

2. Diagnosis (Table 3)

The vast majority of physicians (89.6%) gave depression as one of their diagnoses. While there was no main effect due to country (p=.2361), there was a significant two-way interaction of country of care and patient gender (p=.0251, Figure 1C), German
physicians were more likely to diagnose depression in women, but British and American physicians were more likely to diagnose depression in men. American physicians were the most certain of their depression diagnosis and British general practitioners were the least certain of their depression diagnosis.

3. Clinical Management (Table 4)

British physicians were least likely to prescribe an anti-depressant or to refer the patient to a mental health professional. In addition to variations in referrals to a mental health professional by country of care, there was also variation in the type of professional referral. In Germany, 19 of the 37 (51%) physician referrals to a mental health professional were to a neurologist, 14 (38%) were to a psychiatrist, and 4 (11%) were to a therapist. In the UK, of the 7 referrals to a mental health professional, 4 (57%) went to a therapist and 3 (43%) were to a psychiatric practice nurse. In the US, of the 23 referrals to a mental health professional, 13 (57%) were to a therapist, 5 (22%) were to a psychiatrist, and 5 (22%) were to a psychologist.

There was also considerable variation in the type of antidepressant medication prescribed in the three countries. Of the 33 antidepressant prescriptions in Germany, 11 (33%) were for St. John’s Wort and/or Valerian, 8 (24%) were for a tricyclic antidepressant (TCA), 6 (18%) for a selective serotonin reuptake inhibitor (SSRI) / serotonin-norepinephrine reuptake inhibitor (SNRI), 7 (21%) were unspecified and 1 (3%) was for a benzodiazepine. In the UK, of the 23 antidepressants prescribed, 18 (78%) were for a
SSRI/SNRI, 2 (9%) for a TCA, and 3 (13%) were unspecified. In the US, of the 42 antidepressants prescribed, 38 (90%) were for a SSRI/SNRI and 4 (10%) were for a TCA.

There was a significant two-way interaction of country of care and physician gender when considering referral to another medical (non-mental health) professional (p=.0007, Figure 1D); female German physicians were more likely to refer to another medical professional than were male German physicians; but male British physicians were more likely to refer to another medical professional than were female British physicians, with little variation by gender seen for American physicians. American physicians gave more lifestyle advice (including more advice about exercise).

Regarding follow-up, one British and one American physician reported no need to see the patient again. For those suggesting a follow-up visit, the German physicians would, on average, see the patient again in about a week while their American counterparts were content to wait more than two weeks.

**Discussion**

We found considerable variation by country (and health care system) in the diagnosis and management of a patient with symptoms strongly suggestive of depression. German physicians were more likely to diagnose depression in women, than were British and American physicians, who were more likely to diagnose depression in men. This may be
attributable to differences in the content of medical education among the three countries. That depression may be as common in men as in women and frequently under-diagnosed has been a recent emphasis in medical education\textsuperscript{14}. The certainty of a depression diagnosis, which was highest in the United States, may be important, as we have found that certainty drives test ordering, prescriptions, and referrals\textsuperscript{15,16}. Physicians in the US are significantly more likely to prescribe anti-depressants than are British physicians, which may be a response to several influences: direct-to-consumer advertising, defensive medical practices, or rewards for doing more, in contrast to general practitioners in the UK who are rewarded more for doing less. Between-country differences related to which professional the patient would be referred, may reflect variation in the availability of different types of practitioners in the countries studied, Table 1. American physicians may ask more questions of patients secondary to their need to explore more options to avoid downstream complications. German physicians asked fewer questions and offered less lifestyle advice, possibly because of the shorter average time allocated for appointments: 5.5 minutes was allocated for a routine appointment in Germany compared to 9.7 minutes in the UK and 18.1 minutes in the US\textsuperscript{17}. Although German physicians were also the most likely to do a complete physical, the definition of a complete physical varied by country (as German physicians have 12.5 minutes allocated for a complete physical versus 19.8 minutes in the UK and 36.0 minutes in the US\textsuperscript{17}. Variability in when patients would be seen for a follow-up visit may also be a function of the average time allocated for a visit – having a shorter time allocated for a visit in Germany may result in more frequent visits.
Our research has both strengths and limitations. Well-designed and carefully conducted experiments can provide un-confounded estimates and permit cause and effect statements. Use of a random community-based sample, rather than a convenience sample at a local medical institution or professional conference, enhances the generalizability of our results. Because all physician subjects encountered exactly the same patient and constellation of symptoms of depression, any differences among the three countries cannot be attributed to variation among patients or the seriousness of the symptoms. While an experiment may enjoy excellent internal validity, there is always a threat to external validity (e.g. a physician may act differently with the vignette patient than with real patients seen in their everyday practice). We deliberately attempted to strengthen external validity by: (1) specifically instructing the physician subjects to respond as they would with a real patient in their practice or surgery; (2) having physicians view the vignettes in the context of their clinic practice rather than at home or at a professional meeting where it is likely that they saw real patients before and after the vignettes; (3) conducting role-play sessions and field-testing with experienced clinicians to ensure that the vignettes were clinically authentic. When asked how typical the vignette patient was compared to their own patients, 81.2%, 80.3%, and 85.2% of German, British, and American physicians thought that the presentation was either very or reasonably typical. There is evidence of a secular decline in response rates for social surveys, especially those attempting to recruit physicians, which now run around 50%. Diligent efforts were made to achieve the response rates for our study (60-65%). However, there is always a possibility that the generalizability of our results may be threatened by selection bias. Generalizability to national samples may be poor if the selected sampling areas do not
reflect the nation as a whole. Also, it is impossible to separate any cultural variation from country or health care system variation.

**Conclusion**

There are significant differences between countries (and health care systems) in the way primary care physicians diagnose and manage patients with depression, even when the same “patients” present exactly the same symptoms strongly suggestive of the disease. Such variations may result from international differences in: (1) medical education; (2) attitudes towards patients; (3) the availability of other practitioners; and (4) different methods of reimbursement. Whether similar differences in diagnosis and management occur with other illnesses, like arthritis and cancer, remains to be investigated. Some of the reported international differences in disease rates and health care outcomes may be a result of differences in the manner in which physicians in different countries (and health care systems) diagnose and manage the same symptoms of disease.
References


Figure Legends

Figure 1: Significant two-way interactions of country and other design factors. The panel letter (A-D) reflects the corresponding letter from the last column of Tables 3 and 4.
Table 1. Mental health disorders and resources by country (NA = not available)

<table>
<thead>
<tr>
<th></th>
<th>Germany</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-standardized disability-adjusted life years per 100,000 (2004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unipolar depressive disorders</td>
<td>955</td>
<td>961</td>
<td>1455</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>184</td>
<td>184</td>
<td>184</td>
</tr>
<tr>
<td>Lifetime risk at age 75 of DSM-IV disorders (%)</td>
<td>33.0</td>
<td>NA</td>
<td>55.3</td>
</tr>
<tr>
<td>Mental health resources</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total psychiatric beds per 10,000 population</td>
<td>7.5</td>
<td>5.8</td>
<td>7.7</td>
</tr>
<tr>
<td>Psychiatrists per 100,000 population</td>
<td>11.8</td>
<td>11</td>
<td>13.7</td>
</tr>
<tr>
<td>Psychiatric nurses per 100,000 population</td>
<td>52</td>
<td>104</td>
<td>6.5</td>
</tr>
<tr>
<td>Neurologists per 100,000 population</td>
<td>3.4</td>
<td>1</td>
<td>4.5</td>
</tr>
<tr>
<td>Psychologists per 100,000 population</td>
<td>51.5</td>
<td>9</td>
<td>31.1</td>
</tr>
<tr>
<td>Social workers per 100,000 population</td>
<td>477</td>
<td>58</td>
<td>35.3</td>
</tr>
</tbody>
</table>
Table 2. Signs/symptoms of depression (SIGECAPS) embedded (and excluded) and distractions in the clinical vignette presentation

<table>
<thead>
<tr>
<th>Classic Symptoms of Depression that were included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep Disturbance</td>
</tr>
<tr>
<td>Decreased Interest</td>
</tr>
<tr>
<td>Guilt</td>
</tr>
<tr>
<td>Reduced Energy</td>
</tr>
<tr>
<td>Inability to Concentrate</td>
</tr>
<tr>
<td>Poor Appetite</td>
</tr>
<tr>
<td>Psychomotor Retardation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symptoms of Depression that were not included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicidal Ideation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-verbal clues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dejected affect</td>
</tr>
<tr>
<td>Anxiety</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distractions added to vignette</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffuse pain</td>
</tr>
<tr>
<td>Recently stopped smoking</td>
</tr>
<tr>
<td>Shortness of breath had sent patient to emergency room recently</td>
</tr>
<tr>
<td>Headaches</td>
</tr>
</tbody>
</table>
Table 3. Diagnostic strategies (information gathering, percent with a depression diagnosis, and certainty of depression diagnosis) by country (health care system)

<table>
<thead>
<tr>
<th>Information Gathering</th>
<th>Germany</th>
<th>UK</th>
<th>US</th>
<th>p value</th>
<th>Significant Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of questions</td>
<td>2.6 a</td>
<td>4.5 b</td>
<td>8.0 c</td>
<td>&lt;.0001</td>
<td>A</td>
</tr>
<tr>
<td>Questions about (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pathologies</td>
<td>28.9 a</td>
<td>60.9 b</td>
<td>74.2 c</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>medical history</td>
<td>23.4 a</td>
<td>39.1 b</td>
<td>64.0 c</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>pain</td>
<td>23.4 a</td>
<td>24.2 ab</td>
<td>37.5 b</td>
<td>.0243</td>
<td>B</td>
</tr>
<tr>
<td>alcohol</td>
<td>3.1 a</td>
<td>11.7 a</td>
<td>22.7 b</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>psychological state</td>
<td>25.0 a</td>
<td>45.3 b</td>
<td>52.3 b</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>social situation</td>
<td>37.5 a</td>
<td>47.7 ab</td>
<td>54.7 b</td>
<td>.0209</td>
<td></td>
</tr>
<tr>
<td>general</td>
<td>32.8 a</td>
<td>52.3 b</td>
<td>69.5 c</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Complete physical (%)</td>
<td>46.9 a</td>
<td>11.7 b</td>
<td>39.0 a</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

Diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Germany</th>
<th>UK</th>
<th>US</th>
<th>p value</th>
<th>Significant Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression diagnosis</td>
<td>86.7 a</td>
<td>89.8 a</td>
<td>92.3 a</td>
<td>.2361</td>
<td>C</td>
</tr>
<tr>
<td>Certainty of depression</td>
<td>68.9 ab</td>
<td>65.4 a</td>
<td>74.0 b</td>
<td>.0578</td>
<td></td>
</tr>
</tbody>
</table>

1 Means (by country) with the same superscript letter can not be considered statistically different at the .05 level using Tukey’s multiple comparisons.

2 Significant two-way interactions of country and another design variable that modify the country effect are indicated in the last column, the letter refers to the panel of Figure 1 where the interaction is plotted.
Table 4. Management strategies (prescription, referral, lifestyle advice, and time to follow-up) of a patient with symptoms strongly suggestive of depression by country (health care system)

<table>
<thead>
<tr>
<th>Country</th>
<th>Anti-depressant prescription (%)</th>
<th>Referral (%)</th>
<th>Time to next appointment (days)</th>
<th>Significant Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Germany</td>
<td>UK</td>
<td>US</td>
<td>p value</td>
</tr>
<tr>
<td></td>
<td>25.8 ab</td>
<td>18.0 a</td>
<td>33.6 b</td>
<td>.0241</td>
</tr>
<tr>
<td>Referral (%)</td>
<td>28.9 a</td>
<td>5.5 b</td>
<td>18.0 a</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>28.9 a</td>
<td>7.8 b</td>
<td>7.0 b</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>6.2 a</td>
<td>7.0 a</td>
<td>12.5 a</td>
<td>.1050</td>
</tr>
<tr>
<td></td>
<td>7.8 a</td>
<td>28.9 a</td>
<td>9.4 b</td>
<td>.0321</td>
</tr>
<tr>
<td></td>
<td>6.2 a</td>
<td>7.0 a</td>
<td>12.5 a</td>
<td>.1050</td>
</tr>
<tr>
<td></td>
<td>7.6 a</td>
<td>11.0 b</td>
<td>16.4 c</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

1 Means (by country) with the same superscript letter can not be considered statistically different at the .05 level using Tukey’s multiple comparisons.

2 Significant two-way interactions of country and another design variable that modify the country effect are indicated in the last column, the letter refers to the panel of Figure 1 where the interaction is plotted.
Figure 1:

A. Number of Questions by Gender and Country
   - Germany
   - UK
   - US
   - p = 0.0342

B. Questions about Alcohol by Race and Country
   - Germany
   - UK
   - US
   - p = 0.0476

C. Depression Diagnosis by Gender and Country
   - Germany
   - UK
   - US
   - p = 0.0251

D. Referral to Other Medical Professional by Gender and Country
   - Germany
   - UK
   - US
   - p = 0.0007