

Cartwright, M., Ogden, J., Grunfeld, EA., Weinman, J. (2011). Can self-report questionnaires create illness cognitions in middle-aged men? Health Psychology. (Epub ahead of print)

Title: Brief Report: Can self-report questionnaires create illness cognitions in middle-aged men?

Authors: Martin Cartwright ^{1,2}, Jane Ogden ³, Elizabeth A. Grunfeld ^{1,4} & John Weinman ¹

Affiliations:

¹ Health Psychology Section, Psychology Dept., Institute of Psychiatry, King's College London, London, UK

² Health Services Research, City University, London, UK

³ Department of Psychology, University of Surrey, Guildford, Surrey, UK

⁴ School of Psychology, University of Birmingham, UK

Corresponding author:

Dr Martin Cartwright
Health Services Research
City University
Northampton Square
London
EC1V 0HB
T: +44 (0)20 7040 0879
F: +44 (0)20 7040 0875
E: martin.cartwright.1@city.ac.uk

Keywords: questionnaire context effects; measurement reactivity; survey cognition

ABSTRACT

Objective: To examine the effect of questionnaire context on self-reported illness cognition.

Design: A single-item measure of the perceived impact of lower urinary tract symptoms (LUTS) was embedded twice in a questionnaire battery completed by community-dwelling middle-aged males (N = 1,790). The impact measure was placed in two distinct questionnaire contexts; at the end of a general somatic symptoms questionnaire, and following an illness-specific symptoms questionnaire. The order of the two questionnaire contexts was counter-balanced in a random sub-sample. **Main Outcome Measures:** An established single-item measure of the perceived impact of LUTS. **Results:** Concordance between the two single-item measures was moderate. Scores on a single-item measure of impact were significantly lower when assessed immediately following the completion of a LUTS-specific questionnaire than when assessed following the completion of a general symptoms questionnaire. There was no evidence of order effects. The observed effect was moderated by the severity of LUTS such that the difference in perceived impact scores between contexts (where general symptoms context > illness-specific context) increased as urinary symptoms increased.

Conclusion: Questionnaire context systematically influenced responses on self-report measures of illness impact. The magnitude of the context effect was largest in the highest quintile of LUTS severity, a difference of >0.5 on a scale with a range of 3. These findings may have implications for situations where patient reported outcome measures are used to evaluate healthcare interventions or inform treatment decisions.

INTRODUCTION

Given that illness cognitions are staple constructs in health psychology research and mostly assessed using self-report questionnaires, it is surprising that more attention has not been focused on the potential influence of the structure and format of these instruments on participants' responses (French & Sutton, 2010). In other fields, notably market research and political science, decades of evidence from studies of survey cognition has demonstrated that subtle differences in the presentation of survey items or response formats can have systematic effects on participants' responses (Tourangeau & Rasinski, 1988). Aspects of the survey context that have been shown to affect responses include the question format (e.g. open-ended vs. closed questions; positive vs. negative message framing), available response options (e.g. low frequency vs. high frequency response options), the format of rating scales (e.g. unipolar vs. bi-polar) and respondents' own answers to preceding or even subsequent items (Schwarz, 1999).

Despite a dearth of survey cognition studies focusing specifically on illness cognitions, evidence from the literature on symptom perception suggests that illness cognitions are highly malleable. For example, Pennebaker and colleagues conducted a classic series of studies demonstrating that specific illness expectations, or schema, can be made salient by subtle changes to the standardised instructions participants received prior to completing a questionnaire (Pennebaker, 1982). Once activated, experimentally-induced illness schema strongly influence the reporting of somatic symptoms through a schema-consistent selective search mechanism and reinterpretation of neutral stimuli in line with the salient schema (Anderson & Pennebaker, 1980; Pennebaker & Skelton, 1981). Supporting evidence comes from information-processing studies showing that priming of illness schemas (e.g. common cold) creates an attentional bias towards schema-specific information but not to other non-primed illness schemas (Henderson, Hagger, & Orbell, 2007).

Skelton, Loveland & Yeagley (1996) demonstrated that the salience and accessibility of symptom schema can be increased through detailed recall of schema-relevant information. In a separate study participants who, without any priming, reported relatively high levels of general symptoms (e.g. headache, upset stomach) or relatively low levels of global health status on standardised questionnaire measures subsequently demonstrated attentional bias for general illness-related words relative to control words (Williams, Wasserman, & Lotto, 2003). The implication here is that participants who are highly symptomatic or in poor health have greater access to a higher-order (general) illness schema which accounts for the attentional bias.

Collectively these studies demonstrate that (i.) specific illness schemas can be easily activated through implicit suggestion, priming or recall of conceptually related information, (ii.) once activated illness schemas direct attention to schema-relevant and schema-consistent information, and (iii.) neutral information is reinterpreted in line with the activated schema. As the process of completing a standard health or illness questionnaire involves serial recall of conceptually related symptom information, the studies highlighted raise the possibility that this process alone may be sufficient to activate an illness schema which, through the mechanisms described, could then influence responses to subsequent items in the questionnaire. Analogous processes are thought to underlie the observed *question-behaviour* or *mere measurement effects* (Spratt, Sprangenberg, Block, Fitzsimons, Morwitz, & Williams, 2006) whereby merely completing a questionnaire raises the salience of particular cognitions (e.g. implementation intentions towards a specified behaviour) which in turn increases the likelihood that the respondent will act in accordance with the valence of that cognition (Morwitz & Fitzsimons, 2004). Mere measurement effects have been found for several health-related behaviours (Godin, et al., 2010; Sandberg & Conner, 2009).

Ogden (2003) provocatively asked whether the structure and format of questionnaires widely used in Health Psychology are creating the constructs they purport to measure. The studies reviewed above provide indirect support for this idea but to date there appears to be little direct evidence for the effect of questionnaires themselves on self-reported illness cognitions (Conner & Norman, 2005, p. 353; French & Sutton, 2010). To address this gap in the literature we conducted a proof-of-principle study to establish whether a measure of illness cognition, the perceived impact of lower urinary tract symptoms (LUTS), could be affected by the preceding survey items.

Method

Participants

Community-dwelling men aged 50-79 years were identified from five GP Practices in central London. Exclusion criteria included cognitive impairment, severe health issues and lack of English language sufficiency. Potential participants (N = 4,164) were sent the *Health Matters for Men* survey with a reply-paid envelop. Non-responders received a reminder after 5-6 weeks.

Questionnaires were returned by 1,790 participants (response rate = 43%). The mean age was 62.8 years (SD = 8.13), 69% were either married or living with a partner and the sample was more deprived than England as a whole (mean Index of Multiple Deprivation for sample = 27.27 (8.46), mean IMD for England = 21.67 (SD = 15.74); $z = 15.05$, $p < .001$, two-tailed, $r = 0.36$) (DCLG, 2009). Self-reported ethnic group membership showed that 85.1% classified themselves as White, 8.8% as Black, 1.9% as South Asian, 2.2% as East Asian and 2.1% as Mixed or Other ethnicity.

Design

A single-item measure of the perceived impact of LUTS taken from the ICSmaleSF was embedded twice in a health survey in two different questionnaire contexts. At one point the

impact measure was presented at the end of a general symptoms questionnaire taken from the IPQ-R. In this context the measure is referred to as I-Gen (Impact – general symptom context). At another point in the survey the impact measure was presented immediately following the ICS_{male}SF urinary symptom scale and is referred to as I-Spec (Impact – illness specific context). The wording of the two impact measures, the timeframe (i.e. previous month) and the response format were identical for I-Gen and I-Spec.

To explore the possibility that any observed differences between I-Gen and I-Spec were the result of order effects rather than context effects, a counter-balanced between-subjects design was incorporated into the study. Using a computer generated random number procedure 3,914 potential participants were randomised to Condition 1 and received a version of the survey in which the I-Gen item was presented earlier, while 250 were randomised to Condition 2 and received a version of the survey that presented I-Spec earlier than I-Gen. Assuming a response rate of 40% this would afford power of 0.77 for a two-tailed parametric t-test assuming an α -level of 0.05 and a small effect size ($d = .2$)¹. The randomisation process was subsequently tested by comparing groups on a broad range of demographic (age, deprivation), health (self-reported health status, general somatic symptoms, urinary symptoms) and psychological variables (health value, health anxiety, somatosensory amplification, catastrophising, negative affect). The effect size r was used to describe the magnitude of differences between conditions and ranged from 0.01 to 0.06. Since all effect size estimates fell below Cohen's criteria for a *small* effect we concluded that the randomisation was successful.

Measures

¹ Given the large sample, this asymmetrical split provided sufficient power to detect even small order effects in a direct comparison of means; however we acknowledge that a more equal split across counter-balanced conditions would have increased power further which may have been preferable but is highly unlikely to have changed the findings substantively.

General somatic symptoms. The Illness Identity scale from the Revised Illness Perception Questionnaire (IPQ-R) (Moss-Morris, Weinman, Petrie, Horne, Cameron, & Buick, 2002) was used as a general measure of somatic symptoms. This scale presents 14 commonly experienced symptoms (e.g. sore throat, breathlessness, stiff joints, headaches) with the instruction “*please indicate how much each symptom has interfered with your life over the last month*”. Responses are made using a 4-point Likert scale (<Not at all> <A little> <Somewhat> <A lot>).

Lower urinary tract symptoms. LUTS were assessed using the International Continence Society’s short-form male questionnaire (ICS*male*SF) (Donovan, Peters, Abrams, Brookes, Rosette, & Schafer, 2000). This instrument asks about the frequency of 13 widely recognised urinary symptoms including voiding symptoms (e.g. hesitancy, weak stream), storage symptoms (e.g. urgency, frequency) and incontinence symptoms (e.g. stress incontinence, post-void dribbling).

Impact of illness (single-item). A single-item measure of the perceived impact of LUTS was taken from the ICS*male*SF (Donovan et al, 2000). This item asks ‘overall how much do your urinary symptoms interfere with your life?’ and presents a 4-point response scale (<Not at all> <A little> <Somewhat> <A lot>).

RESULTS

There was less than 2% missing data on the variables in the analysis. Missing values were imputed in SPSS using the MCMC approach. All reported results are based on the imputed dataset but the pattern of results is replicated when the analyses are repeated on the raw data without imputed values.

Table 1 presents cross-tabulations of responses on I-Gen and I-Spec for each counterbalanced condition. The diagonal represents concordance between I-Gen and I-Spec. In Condition 1 (I-Gen precedes I-Spec) discordant responses were 4.4 times more likely to be in the direction I-

Gen > I-Spec than I-Gen < I-Spec (409 vs. 92 cases respectively). In Condition 2 (I-Spec precedes I-Gen) discordant responses were 2.8 times more likely to be in the direction I-Gen > I-Spec than I-Gen < I-Spec (22 vs. 8 cases respectively). Once chance had been accounted for, agreement between the two measures was 41% for Condition 1 ($K = .413$, $p < .001$) and 44% for Condition 2 ($K = .437$, $p < .001$) which represent only a 'moderate' degree of concordance (Landis & Koch, 1977).

A mixed ANOVA with counterbalancing as a between-subjects factor (Condition 1: I-Gen precedes I-Spec vs. Condition 2: I-Spec precedes I-Gen) and questionnaire context as a within-subjects factor (general symptoms context vs. illness-specific symptoms context) was used to compare perceived impact scores. The mean (SD) impact scores in Condition 1 were I-Gen = 1.60 (.88) and I-Spec = 1.37 (.68), and in Condition 2 I-Gen = 1.53 (.81) and I-Spec = 1.36 (.65). The analysis showed no effect of counterbalancing ($F(1, 1788) = .38$, $p = .54$, $\eta^2 = .000$), a significant effect of questionnaire context ($F(1, 1788) = 41.30$, $p < .001$, $\eta^2 = .020$) and no interaction between counterbalancing and questionnaire context ($F(1, 1788) = .77$, $p = .38$, $\eta^2 = .000$).

To investigate whether the observed context effect was moderated by the severity of urinary symptoms, a variable representing total LUTS score (based on the ICSmaleSF) was included in the model as a covariate. In this analysis there was no effect of counterbalancing ($F(1, 1787) = 2.06$, $p = .151$, $\eta^2 = .000$) and no interaction between counterbalancing and questionnaire context ($F(1, 1787) = 1.11$, $p = .293$, $\eta^2 = .000$) but a significant main effect of context ($F(1, 1787) = 45.95$, $p < .001$, $\eta^2 = .023$), a significant main effect of LUTS ($F(1, 1787) = 2114.8$, $p < .001$, $\eta^2 = .533$) and a significant interaction between context and LUTS ($F(1, 1787) = 142.63$, $p < .001$, $\eta^2 = .072$). Table 2 shows the direction and magnitude of differences in perceived impact scores between questionnaire contexts across quintiles of LUTS.

DISCUSSION

This study provides evidence that questionnaire context, operationalised as the immediately preceding survey items, influences responses on a measure of illness cognition. Scores on a single-item measure of LUTS impact were significantly lower when assessed immediately following the completion of a LUTS-specific questionnaire (I-Spec) than when assessed following the completion of a general symptoms questionnaire (I-Gen). There was no evidence of order effects. The observed context effect was moderated by the severity of LUTS symptoms such that the difference in perceived impact scores (where I-Gen > I-Spec) increased as urinary symptoms increased.

To explain the main effect of questionnaire context we posit the *facilitated recall hypothesis*. When completing a single-item measure of LUTS impact as part of a general somatic symptoms questionnaire, participants read the question then search their memory for confirmatory evidence of cardinal urinary symptoms and concomitant impact. This heuristic process is in line with research showing that schemas are easily activated, consist of archetypal symptom clusters and generate biased attentional processes as part of a schema-consistent selective search mechanism (Anderson & Pennebaker, 1980; Bishop & Converse, 1986; Williams, Wasserman, & Lotto, 2003; Pennebaker & Skelton, 1981). In contrast, completing an illness-specific questionnaire facilitates recall of both confirmatory evidence (memories of urinary symptoms that they have experienced) and disconfirmatory evidence (an absence of memories of other urinary symptoms that they have not experienced or memories of urinary symptoms that they have experienced only infrequently or weakly). Exposure to both confirmatory and disconfirmatory evidence overrides the default heuristic process and moderates reports of perceived impact when measured immediately after an illness-specific symptoms questionnaire.

Facilitated recall also explains the moderating effect of LUTS. At low levels of LUTS the difference in perceived impact between the contexts is small (Table 2). Few urinary symptoms are available to be recalled therefore there is little confirmatory evidence in either the general symptoms context or the illness-specific context and disconfirmatory evidence in the illness-specific context has a minimal moderating effect on judgements of impact. At higher levels of LUTS, the illness-specific context will continue to generate some disconfirmatory evidence since few people experience the full range of urinary symptoms maximally, whereas the general symptoms context provides only increasing levels of confirmatory evidence. As LUTS increase the different mental processes generated by the questionnaire contexts are reflected in increasing differences in impact scores.

Counterbalancing was employed to control for potential order effects in a repeated measures design but can also be interpreted as a measure of the duration of the observed context effects. The failure to find a significant main or interaction effect of counterbalancing in any analysis suggests that there was little carryover from the general symptom context to the illness-specific context or vice versa which implies that the context effect is relatively transient.

There are some potential limitations of the study. The response rate was 43% and the sample was drawn from a population of middle-aged males, therefore it cannot be assumed that the findings generalise to non-responders, younger men or women. However, the large sample size and minimal exclusion criteria afford greater confidence that the effect is robust amongst middle-aged men and evidence from related areas, such as schema activation and attentional bias, demonstrate effects in men and women of various ages. The observed effects are therefore unlikely to be specific to middle-aged men. A further potential limitation of the study is that we investigated the effect of context on a single-item measure of impact. Critics may argue that single-item measures are inherently less reliable than scale measures. While

this may be true, the observed difference in impact scores between contexts was not merely a reflection of random variation or unreliability but of systematic bias. Moreover, scale measures are comprised of individual items and, partly due to their greater reliability, context effects may actually be easier to detect at the aggregated scale level.

French & Sutton (2010) called for more research on measurement reactivity and the current study contributes to the evidence base and has implications for practice and research. The finding that questionnaire context significantly affects responses on a measure of illness cognition challenges the general implicit assumption that self-report measures are neutral, unbiased instruments and supports the notion that measuring changes the measure (Knowles, 1988; Ogden, 2003). The overall magnitude of the context effect was *small* but the interaction between context and LUTS was *medium* and in the highest LUTS quintile (which most closely reflects the levels of LUTS seen within the healthcare system) the mean difference of > 0.5 on a scale with a maximum range of three (i.e. 1 to 4) was statistically *large* (Clark-Carter, 1997). In an era where patient reported outcome measures are increasingly used to evaluate new drugs and healthcare interventions and guide treatment decisions (e.g. American Urological Association, 2010), systematic bias created by measurement instruments can have serious consequences for clinical practice. There are implications for research too. There might be a case, for example, to recommend that in illness-specific versions of the IPQ (see French & Weinman, 2008; Moss Morris et al, 2002) the Identity Scale is either limited to symptoms that are associated with the condition or distinguishes between these and more general symptoms. The effect of these differing approaches on responses to the Identity scale and other IPQ scales requires empirical investigation. More generally, further studies are needed to explore other constructs beyond perceived impact, to investigate context effects on scale measures, and to test the generaliseability of questionnaire context effects to other clinical populations and situations. Lastly, there would be obvious

practical implications from studies that investigate whether particular questionnaire contexts could be said to generate objectively more accurate responses from participants.

References

- Anderson, D., & Pennebaker, J. (1980). Pain and pleasure - Alternative interpretations for identical stimulation. *European Journal of Social Psychology*, *10*, 207-212. doi: 10.1002/ejsp.2420100208
- American Urological Association (2010). *Management of BPH (Revised)*. Retrieved August 15, 2011, from <http://www.auanet.org/content/guidelines-and-quality-care/clinical-guidelines.cfm?sub=bph>
- Bishop, G., & Converse, S. (1986). Illness representations: A prototype approach. *Health Psychology*, *5*, 95-114. doi: 10.1037/0278-6133.5.2.95
- Clark-Carter, D. (1997). *Doing quantitative psychological research from design to report*. Hove, UK: Psychology Press.
- Conner, M., & Norman, P. (2005). *Predicting Health Behaviour*. Maidenhead, England: Open University Press.
- DCLG. (2009). *Communities and neighbourhoods: Indices of deprivation 2004*. Retrieved May 26, 2011, from <http://webarchive.nationalarchives.gov.uk/+http://www.communities.gov.uk/archived/general-content/communities/indicesofdeprivation/216309/>
- Donovan, J., Peters, T., Abrams, P., Brookes, S., Rosette, J., & Schafer, W. (2000). Scoring the short form ICSmaleSF questionnaire. *Journal of Urology*, *164*, 1948-1955.
- French, D., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology*, *15*, 453-468. doi:10.1348/135910710X492341
- French, D., & Weinman, J. (2008) Assessing illness perceptions: beyond the IPQ. *Psychology & Health*, *23*, 5-9. doi: 10.1080/08870440701616714
- Godin, G., Sheeran, P., Conner, M., Delage, G., Germain, M., Belanger-Gravel, A., et al. (2010). Which survey questions change behavior? Randomised controlled trial of mere measurement interventions. *Health Psychology*, *6*, 636-644. doi: 10.1037/a0021131
- Henderson, C., Hagger, M., & Orbell, S. (2007). Does priming a specific illness schema result in an attentional informational-processing bias for specific illnesses? *Health Psychology*, *26*, 165-173. doi: 10.1037/0278-6133.26.2.165

Klimidis, S., Minas, I., & Yamamoto, K. (2001). Impact of Illness Scale: Reliability, validity and cross-cultural utility. *Comprehensive Psychiatry*, *42*, 416-423. doi: 10.1053/comp.2001.26266

Knowles, E. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55*, 312-320.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Morwitz, V., & Fitzsimons, G. (2004). The mere-measurement effect: Why does measuring intentions change actual behaviour? *Journal of Consumer Psychology*, *14*, 64-74. doi: 10.1207/s15327663jcp1401&2_8

Moss-Morris, R., Weinman, J., Petrie, K., Horne, R., Cameron, L., & Buick, D. (2002). The revised Illness Perception Questionnaire (IPQ-R). *Psychology & Health*, *17*, 1-16. doi: 10.1080/08870440290001494

Ogden, J. (2003). Some problems with social cognition models: A pragmatic and conceptual analysis. *Health Psychology*, *22*, 424-428. doi: 10.1037/0278-6133.22.4.424

Pennebaker, J. W. (1982). *The Psychology of Physical Symptoms*. New York: Heidelberg.

Pennebaker, J., & Skelton, J. (1981). Selective monitoring of physical sensations. *Journal of Personality and Social Psychology*, *41*, 213-223.

Sandberg, T., & Conner, M. (2009). A mere measurement effect for anticipated regret: Impacts on cervical screening attendance. *British Journal of Social Psychology*, *48*, 221-236. doi: 10.1348/014466608X347001

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.

Skelton, J., Loveland, J., & Yeagley, J. (1996). Recalling symptom episodes affects reports of immediately-experienced symptoms: Inducing symptom suggestibility. *Psychology and Health*, *11*, 183-201. doi:10.1080/08870449608400252

Sprott, D., Sprangenberg, E., Block, L., Fitzsimons, G., Morwitz, V., & Williams, P. (2006). The question-behaviour effect: What we know and where we go from here? *Social Influence*, *1*, 128-137. doi: 10.1080/15534510600685409

Tourangeau, R., & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*, 299-314.

Williams, P., Wasserman, M., & Lotto, A. (2003). Individual differences in self-assessed health: An information-processing investigation of health and illness cognition. *Health Psychology*, *22*, 3-11. doi: 10.1037/0278-6133.22.1.3

Table 1 Cross-tabulation of I-Gen and I-Spec

		Condition 1 (I-Gen precedes I-Spec)				Condition 2 (I-Spec precedes I-Gen)			
		I-Spec				I-Spec			
		Not at all	A little	Somewhat	A lot	Not at all	A little	Somewhat	A lot
I-Gen	Not at all	56.5 (950)	4.0 (68)	0.2 (3)	0.1 (1)	56.9 (62)	5.5 (6)	0.0 (0)	0.0 (0)
	A little	12.9 (217)	9.8 (165)	0.7 (11)	0.0 (0)	10.1 (11)	14.7 (16)	0.0 (0)	1.8 (2)
	Somewhat	2.0 (33)	5.9 (99)	2.1 (36)	0.5 (9)	2.8 (3)	3.7 (4)	0.0 (0)	0.0 (0)
	A lot	0.5 (8)	1.5 (25)	1.6 (27)	1.7 (29)	0.9 (1)	1.8 (2)	0.9 (1)	0.9 (1)

The first figure in each cell represents the percentage of participants in that cell, the absolute numbers are included in parentheses.

Table 2 Moderation of the questionnaire context effect by number of LUTS

LUTS Quintile	N	Mean (SD)		Ranks			Z [†]	Effect size r
		I-Gen	I-Spec	I-Gen > I-Spec	I-Gen = I-Spec	I-Gen < I-Spec		
1 (Lowest)	378	1.07(.30)	1.01 (.13)	23	353	2	-3.74***	0.19
2	308	1.15(.39)	1.08(.31)	35	261	12	-3.15**	0.18
3	398	1.31(.53)	1.17(.41)	76	298	24	-5.27***	0.26
4	327	1.80(.83)	1.47(.63)	121	176	30	-7.60***	0.42
5 (Highest)	379	2.61(.95)	2.10(.89)	176	171	32	-9.69***	0.50

[†] Wilcoxon signed-rank test; ** p < 0.01, *** p < 0.001