INVESTIGATIVE REPORT

How Robust are the Dermatology Life Quality Index and Other Self-reported Subjective Symptom Scores when Exposed to a Range of Experimental Biases?

Caroline S. MURRAY and Jonathan L. REES
Department of Dermatology, University of Edinburgh, Edinburgh, United Kingdom

**Subjective-symptom tools used in dermatology have rarely been experimentally tested for cognitive "focus" and "framing" biases. We investigated the effects of affective biases on the Dermatology Life Quality Index (DLQI), the Global Health Question and visual analogue scores. Two experiments tested the response to affect-eliciting words and film. We demonstrated no significant difference in median DLQI scores for subjects exposed to negative vs. neutral words (medians 8.5 and 9.5, respectively), or negative vs. positive words (medians 6.0 and 9.0, respectively, overall *p*=0.41.) Median DLQI scores were similar for groups who had (8.0), or had not (9.0), seen a video clip about a severe skin condition (*p*=0.34). Finally, we compared an <u>A</u>mended DLQI (ADLQI), the DLQI re-worded into neutral "frames", with the standard DLQI. ADLQI median scores were higher (ADLQI 8.25, DLQI 6.75), but not significantly so (*p*=0.47). We have been unable to demonstrate any effects of the biases studied, but the statistical power of our study is modest.
*Key words: dermatology; bias; quality of life.***

Caroline Siân Murray, Department of Dermatology, Room 4.018, First Floor, The Lauriston Building, Lauriston Place, Edinburgh, EH3 9HA, UK. E-mail: carolinesianmurray@ ed.ac.uk

It is widely accepted that objective measures of disease based on patho-biological variables are insufficient to measure the personal impact of disease. There are at least two reasons for this. First, we do not have objective correlates of many states, such as pain or itch (1, 2), and secondly, how the disease process affects an individual person will depend on a range of individual and contextual factors. For instance, the visibility of extensive psoriasis may be a greater burden to an individual who likes to go swimming, than to an individual who never goes swimming. Another example is that a person who previously had very severe disease, for instance bad childhood eczema, will use that as a comparator for their present state: their current disease state might be viewed differently if there was no previous history of skin disease (3).

In recent years, a number of tools have been developed with the goal of measuring the functional burden of disease as experienced by the patient. Examples include the Global Health Questionnaire (GHQ), which has been used to provide an overall assessment of patient-perceived health (4), whilst another generic tool, the visual analogue scale (VAS), has been used as a measure of a range of symptoms such as itch or pain (5–7). For skin disease, one of the most widely used tools is the Dermatology Life Quality Index (DLQI) (8, 9). The designers of the tool had practicality in mind – it was considered a priority that the tool should be short and quick to answer, in order to aid assessment during clinic visits, and thus this 10-question tool was designed. In a medical and economic climate where resources are scarce, the assessment of quality of life and how it may be influenced by medical intervention has become a major research programme. In England, the British Association of Dermatologists and the National Institute of Health and Clinical Excellence (NICE) has advocated the use of the DLQI as a disease assessment tool for patients with psoriasis, to determine whether they should receive certain expensive biological therapies (10) (http://www.nice.org.uk/Guidance/TA134).

Work over the last 25 years, in cognitive psychology and especially in the field of happiness research, has revealed a number of problems surrounding the measurement of subjective states, quality of life and utility (for reviews, see Kahneman et al. (11)). First, and strange as it may initially appear, individuals may not be able to access their own feelings (12, 13), and the way in which information is gathered may alter, or influence, the patient's own perception of their own feelings (3, 14–16). Secondly, a number of cognitive limitations may limit the value of subjective knowledge: patients may not be able to remember changes in their functional status, nor predict the effects of particular interventions or change in state (12, 13, 17–19).

In the present paper, we set out to explore the effects of contextual or "framing" biases on commonly used subjective measures including the DLQI in dermatology. We use the term "framing bias" widely to include bias that stems from how the feeling, emotion or symptom is enquired about. Questions can be "framed" in language or presented in a context that may elicit a stereotyped

answer; for instance, by implying that an aspect of disease should be considered as a negative phenomenon, leading a respondent to consider this aspect as negative where they did not before (3). Secondly, the immediate context may alter how individuals perceive their own symptoms. For example, patients frequently anchor or skew their own assessment of disease by reference to others who they think are less or more fortunate (3, 5).

We therefore designed three experiments in which either the wording of the DLQI was altered, or the immediate context in which individuals completed the DLQI, GHQ or VAS for common symptoms was manipulated. The manipulation was performed using video, listing of negative or neutral words, and alterations in the actual wording of the DLQI.

METHODS

Participants

An opportunistic sample of 215 patients was recruited. Because of the absence of similar prior work, formal power calculations were not performed. For each study, consecutive patients who agreed to take part were enrolled from the Royal Infirmary's Department of Phototherapy in Edinburgh. Details of specific diagnoses were not sought, the usual throughput of the phototherapy department would suggest that the majority (70%) of the patients had psoriasis, a minority (approximately 10%) had eczema and 20% other conditions (for instance generalized pruritus). Ethics committee approval was granted by the Lothian Ethics Committee (LREC reference: 06/S1104/56).

All study procedures and patient interactions were conducted using a consistent, written script. The interaction script and subject information sheet explained that the studies were to determine which sort of questionnaire or score was most accurate in assessing symptoms. The interventions were described in general terms ("You will be given a list of words to memorise" or "If you are randomised into a certain group you may see a film broadcast on terrestrial television") in order to minimize unintentional "unblinding".

All participants completed the GHQ ("In general, for someone of your age, would you say that your health is excellent, very good, good or poor?"), DLQI and VAS of disease extent, itch and insomnia, always in the same order. The DLQI is a 10-question tool, the score of which is acquired by summing the score for each question. The higher the DLQI score, the more severely their quality of life is affected (maximum score 30.) Most participants scored in the region of 6–10, which equates to a "moderate" effect of the skin condition on the quality of life.

Experiment 1

Our hypothesis was that, if subjects were exposed to certain mood-eliciting words, they would affect the subjective score accordingly, for instance, if they had read negative words, their subjective scores would suggest worse disease.

Forty patients were randomized into two groups. Group 1 were asked to read 10 negative words, had one minute to memorize them and were then asked to write them out. After this, they completed the GHQ, DLQI and the VAS of disease extent for both itch and insomnia. Group 2 went through an identical process, but the participants were given a list of 10 neutral words (certain fields of psychology research have identified and use words that elicit certain affective states.) The words for this part of the experiment

are listed in Table I and were taken from the "Balanced Affective Word List" (http://www.sci.sdsu.edu/CAL/wordlist/origwordlist. html). The words were matched with respect to total character and syllable length. A further 41 patients were randomized into two groups. Group 3 were asked to read 10 negative words and then write them out (without the necessity of memorizing them). The participants then completed GHQ, DLQI and VAS or disease extent, itching and insomnia. Group 4 went through an identical process, except participants were given a list of 10 positive words. These words were taken from the University of Florida's NIMH Centre for the Study of Emotion and Attention (http://csea.phhp.ufl.edu/Media.html#bottommedia) and are listed in Table I. The source of affect-eliciting words was altered as this afforded a larger scope of words with more recent and more extensive validation. Again, the words were matched for total character and syllable length.

Experiment 2

Our hypothesis was that if a subject saw a film highlighting the negative aspects of having a skin disease, then this would make them focus on the negative aspects of living with their skin disease and so their subjective scores would imply that they had worse disease.

Fifty-four patients were randomized into two groups, with Group 1 completing GHQ then VAS for disease extent, itch and insomnia, after having watched a 10-min clip from a terrestrial television broadcast ("Real Families: My Skin Could Kill Me", which was broadcast before the "watershed" on ITV1 in October 2005) about living with the severe skin condition, Harlequin ichthyosis. Group 2 just completed the subjective tools without having watched the television clip. All subjects were questioned in the same way and in the same experimental room, whether or not they had watched the film. The randomization result (to watch the film or not) was included in the questionnaire envelope and was opened, with the interviewer present in the study room, the interviewer then adopted the appropriate script (for whether or not the participant was to watch the film) from that point.

Table I. *Experiment 1: words presented to each intervention group*

Negative words (Group 1)	Neutral words (Group 2)	Positive words (Group 3)	Negative words (Group 4)
worry	wagon	angel	abuse
ashamed	aluminium	birthday	bankrupt
gloom	green	beauty	betray
bad	bus	caress	cancer
sick	scan	cheer	cruel
suffering	submarine	freedom	funeral
unhappy	vitamin	glory	gloom
itch	iron	humour	hatred
misery	margin	home	hurt
rejected	resident	joke	jail
		mother	misery
		pretty	poison
		passion	pollute
		reward	rabies
		romantic	rejected
		sun	sad
		sexy	sick
		snuggle	suicide
		treasure	terrible
		triumph	tragedy

Table II. *Experiment 1: subject characteristics, median scores and p-values of Kruskal-Wallis analysis of variance (ANOVA)*

	Age, years mean (range)	M/F	Median DLQI (interquartile range)	Median GHQ (interquartile range)	Median VAS (interquartile range)		
					Extent	Itch	Insomnia
Group 1 (Neg) n=20	39.8 (17–71)	12/8	8.50 (5.00–11.75)	2.00 (1.00–2.00)	3.60 (2.80–5.60)	2.15 (0.90–4.33)	3.60 (1.50–7.20)
Group 2 (Neut) n=20	41.4 (18–68)	9/11	9.50 (5.75–14.25)	2.00 (2.00–2.00)	3.70 (1.80–5.30)	2.15 (1.08–4.40)	4.80 (2.85–5.73)
Group 3 (Neg) n=19	39.0 (20–72)	10/9	6.00 (2.00–10.50)	2.00 (2.00–3.00)	2.50 (0.60–5.00)	1.50 (0.55–3.95)	1.90 (0.70–5.75)
Group 4 (Pos) n=22	37.4 (16–74)	11/11	9.00 (2.25–12.50)	2.00 (2.00–3.00)	4.80 (2.50–7.50)	1.30 (0.40–3.00)	3.40 (1.30–6.00)
p=Kruskal-Wallis			0.41	0.44	0.35	0.46	0.27

DLQI: Dermatology Life Quality Index; GHQ: Global Health Question; VAS: visual analogue scale.

Experiment 3

In this study, our hypothesis was that if the DLQI focused on negative aspects of disease, then re-framing it into "neutral" frames should result in scores implying a better quality of life. We also hypothesized that if the DLQI focused on the negative, then this may negatively affect the responses to other subjective symptom scores.

Eighty patients were randomized into two groups and each of these two groups further split into two sub-groups, giving a total of four sub-groups. Half the subjects answered the GHQ and standard DLQI, whilst the other half answered an altered DLQI (ADLQI) and the standard GHQ. The ADLQI mirrored the standard DLQI, but an attempt was made for each question to be re-written in a neutral frame, thereby, minimizing the possibility of a positive or negative framing and potentially reducing the possibility of a stereotyped answer. The ADLQI is shown in the electronic appendix (http://adv.medicaljournals. se/article/abstract/10.2340.00015555-0768/app1). Division of the two groups allowed the ordering of the examination to be manipulated, with half the subjects receiving the GHQ first, and then either the DLQI or the ADLQI, with the other half receiving the GHQ second.

Demographic variables including age and sex, together with the results, were de-identified and recorded in Excel. Statistical analyses were undertaken using R-software (http://www.R-project.org (20)).

RESULTS

Examination of raw data, not surprisingly, showed that the majority of variables were non-normally distributed. Medians were therefore compared using the Kruskal-Wallis (KW) analysis of variance (ANOVA), or for count data, Fisher's exact test for $r \times c$ contingency tables. Formal significance was taken at $p < 0.05$.

Because of the limited range of the GHQ questionnaire, results were also examined using Fisher's exact test, but this did not alter any of the conclusions and is not presented.

Experiment 1

The impact of affect-eliciting words. A total of 81 subjects were studied and their characteristics are shown in Table II. There were four intervention groups, numbered 1–4, as mentioned above. There were no significant differences in the sex allocation (Fishers test, $p = 0.81$) nor median ages (Kruskal-Wallis, $p = 0.70$) between the four groups. Median scores for the four groups and p-values using the Kruskal-Wallis ANOVA are shown in Table II. As can be seen there are no significant differences evident.

Experiment 2

The impact of watching a film about living with a severe skin condition. A total of 54 subjects were studied. Their characteristics and the median group-scores and Kruskal-Wallis ANOVA are shown in Table III.

There was no significant sex difference between the two groups, those who were shown the video and those who were not (Fisher, $p = 0.27$). The median age of those shown the video was 52 compared with those who were not shown the video of 33; a difference that is highly significant (KW, $p = 0.001$). However, scatter plots did not show any obvious correlation between

Table III. *Experiment 2: subject characteristics, median scores and p-values for Kruskal-Wallis analysis of variance (ANOVA)*

	Age, years mean (range)	Sex M/F	Median DLQI (interquartile range)	Median GHQ (interquartile range)	Median VAS (interquartile range)		
					Extent	Itch	Insomnia
Group 1 (video) n=27	50.3 (17–77)	14/13	8.00 (4.50–9.00	3.00 (2.00–3.00)	4.50 (2.95–6.15)	2.00 (0.85–3.90)	3.90 (2.10–6.35)
Group 2 (no video) n=27	35.8 (16–74)	9/18	9.00 (4.50–15.00)	2.00 (1.50–3.00)	3.20 (1.65–6.40)	2.70 (0.95–5.15)	2.20 (1.45–5.05)
p=Kruskal-Wallis			0.34	0.20	0.50	0.11	0.21

DLQI: Dermatology Life Quality Index; GHQ: Global Health Question; VAS: visual analogue scale.

Table IV. *Experiment 3: subject characteristics, median scores and p-values of Kruskal-Wallis analysis of variance (ANOVA)*

	Age, years mean (range)	Sex M/F/Un known	Median QI (interquartile range)	Median GHQ (interquartile range)
Group 1 (DLQI then GHQ) $n=21$	45.1 (28–68)	14/6/1	6.50 (2.75–10.75)	3.00 (2.00–3.00)
Group 2 (GHQ then DLQI) $n=21$	42.8 (16–79)	9/11/1	7.00 (4.50–11.00)	2.00 (2.00–3.00)
Group 3 (ADLQI then GHQ) $n=19$	42.3 (19–81)	9/9/1	8.00 (6.75–10.50)	2.00 (1.00–3.00)
Group 4 (GHQ then ADLQI) $n=19$	44.0 (17–76)	8/11/	8.50 (6.00–13.50)	2.00 (1.75–3.00)
$p=$Kruskal-Wallis			0.47	0.76

DLQI: Dermatology Life Quality Index; ADLQI: Amended Dermatology Life Quality Index; GHQ: Global Health Question; QI: Quality of Life Index score.

age and the outcome measures, so this difference was ignored. Median scores and *p*-values for the Kruskal-Wallis ANOVA are shown in Table IV. As can be seen, there are no significant differences evident.

Experiment 3

Re-wording DLQI into neutral frames. A total of 80 subjects were studied and their characteristics are summarized in Table IV. GHQ and quality of life scores (QI) were examined following four "treatments". The first "treatment", the DLQI, was compared with the second "treatment", ADLQI and, following this, the ordering of GHQ and DLQI/ADLQI were studied (hence treatment groups were numbered as follows: 1, 2 (DLQI) and 3, 4 (ADLQI.)

There were no significant differences in sex (Fisher, $p=0.17$) or age (KW, $p=0.92$) between the four groups. Median scores and Kruskal-Wallis ANOVA across the four groups for QI (DLQI and ADLQI) and GHQ are listed in Table IV. These differences are not significant (KW for QI $p=0.47$ and GHQ $p=0.76$). The ordering had no effect on GHQ ($p=0.60$) or QI ($p=0.5$) scores and therefore groups 1 and 2, and 3 and 4 were combined. Medians for the ADLQI and the DLQI for these combined groups were 8.5 and 7, respectively, a difference that was close to statistical significance with a *p*-value of 0.07 (Kruskal-Wallis test).

DISCUSSION

The results presented are, essentially, negative and, in that sense, they can be viewed as reassuring. Using the criteria of statistical significance we were unable to significantly alter the scores with the various attempts at manipulation of the context or wording of the questionnaires of VAS. There are a number of limitations to the work we present.

Although we studied 215 subjects, we did so in the absence of formal power calculations and a type II error is always possible. Whereas, if the effect of any biases had been major, then we may have detected it, more modest effects will probably have gone undetected. We cannot rule out clinically relevant effects, although our data provide the effect estimates for future studies.

Secondly, even within the experimental paradigm we adopted, there were limitations to the way the experiments were carried out. For instance, although we used a video of a child affected by skin disease, we did not find a suitable video that we thought was meaningful to use as a control. We also found it extremely difficult to alter the wording of the DLQI without producing a caricature of it. The differences seen between the altered DLQI and the genuine DLQI approach significance, but of course, interpretation of these differences is not straightforward. The fact that a different questionnaire produces a different median score is not unexpected and, even if the difference had been significant, it does not invalidate, in any way, the use of the DLQI. Another facet of this experiment is that it demonstrates that the DLQI itself would not appear to bias answering of other scores: the GHQ scores were similar whether the participants had been exposed to DLQI or to the supposedly neutral-framed ADLQI.

Although we have not demonstrated any effects of framing or contextual factors in our study, the study itself was experimental and may not reproduce the sorts of real life factors that will influence the way people respond to questionnaires. For instance, and rather mundanely, a patient whose appointment has been delayed excessively, one can imagine, might be considered more likely to weight his or her own disease more heavily. It would be difficult to capture such influences. Furthermore the use of measures such as DLQI as justification for therapy (or denial of therapy) as in the UK is also much more complex than some appreciate (21). Clinical anecdote suggests that patients are quite capable of "gaming" the system to achieve what they feel are appropriate, and one should remember that quality of life, health status and patients' perception of these measures are distinct (21). It is difficult not to imagine that if patients are meaningfully consented, and the purpose of the DLQI as a justification of clinical need is explained, that patients will not moderate their answers accordingly.

Finally, it was not our purpose to compare different questionnaires, or measures of aspects of diseases. There is already a large literature on this and on the advantages and disadvantages of speciality or disease-specific scoring systems vs. more generic questionnaires, such as

EuroQOL or SF-36, for instance (22–24). We do feel that it is important, however, that in view of the fact that there is an increasing literature on the design, use, limitations of various disease-scoring systems and on cognitive psychology as a whole, that this information is acknowledged and used to continue to validate the subjective tools that we commonly use.

ACKNOWLEDGEMENTS

REFERENCES

1. Rees J, Murray CS. Itching for progress. Clin Exp Dermatol 2005; 30: 471–473
2. Bringhurst C, Waterston K, Schofield O, Benjamin K, Rees JL. Measurement of itch using actigraphy in pediatric and adult populations. J Am Acad Dermatol 2004; 51: 893–898.
3. McColl E, Fayers P. Context effects and proxy assessments. In: Fayers P and Hays R, editors. Assessing quality of life in clinical trials. Second edn. Oxford: Oxford University Press; 2005, p. 131–146.
4. Bjorner JB. Self-rated health. In: Fayers P and Hays R, editors. Assessing quality of life in clinical trials. Second edn. Oxford: Oxford University Press; 2005, p. 309–323.
5. Wahlgren CF. Measurement of itch. Semin Dermatol 1995; 14: 277–284.
6. Wahlgren CF. Itch and atopic dermatitis: clinical and experimental studies. Acta Derm Venereol Suppl 1991; 165: 1–53.
7. Hägermark O, Wahlgren CF. Some methods for evaluating clinical itch and their application for studying pathophysiological mechanisms. J Dermatol Sci 1992; 4: 55–62.
8. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. Clin Exp Dermatol 1994; 19: 210–216.
9. Lewis V, Finlay AY. 10 years experience of the Dermatology Life Quality Index (DLQI). J Invest Dermatol Symp Proc 2004; 9: 169–180.
10. Smith CH, Anstey AV, Barker JN, Burden AD, Chalmers RJ, Chandler D, et al. British Association of Dermatologists guidelines for use of biological interventions in psoriasis 2005. Br J Dermatol 2005; 153: 486–497.
11. Kahneman D, Diener E, Schwarz R. Well-being: the foundations of hedonic psychology. 1st edn. Schwartz R, editor. New York: Russell Sage Foundation Publications; 2003.
12. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996; 66: 3–8.
13. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. Pain 2003; 104: 187–194.
14. Robinson MD, Clore GL. Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. J Pers Soc Psychol 2002; 83: 198–215.
15. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 2002; 128: 934–960.
16. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004; 306: 1776–1780.
17. Fredrickson BL, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. J Pers Soc Psychol 1993; 65: 45–55.
18. Kahneman D, Krueger AB, Schkade D, Schwarz N, Stone AA. Would you be happier if you were richer? A focusing illusion. Science 2006; 312: 1908–1910.
19. Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. Cognitive and emotional perspectives. JAMA 1993; 270: 72–76.
20. Team RDC. A language in environment of statistical computing. Vienna, Austria 2006.
21. Bradley C. Importance of differentiating health status from quality of life. Lancet 2001; 357: 7–8.
22. Both H, Essink-Bot ML, Busschbach J, Nijsten T. Critical review of generic and dermatology-specific health-related quality of life instruments. J Invest Dermatol 2007; 127: 2726–2739.
23. Chren MM, Lasek RJ, Quinn LM, Covinsky KE. Convergent and discriminant validity of a generic and a disease-specific instrument to measure quality of life in patients with skin disease. J Invest Dermatol 1997; 108: 103–107.
24. Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: what do dermatology life quality index scores mean? J Invest Dermatol 2005; 125: 659–664.

Electronic appendix

No.	Title	Available at:
1	Amended Dermatology Life Quality Index	http://adv.medicaljournals.se/article/abstract/10.2340.00015555-0768/appendix