This is a repository copy of *Family and professionals underestimate quality of life across diverse cultures and health conditions: Systematic review*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/83032/

Version: Accepted Version

Article:

Crocker, TF, Smith, JK and Skevington, SM (2015) Family and professionals underestimate quality of life across diverse cultures and health conditions: Systematic review. Journal of Clinical Epidemiology, 68 (5). pp. 584-595. ISSN 1878-5921

https://doi.org/10.1016/j.jclinepi.2014.12.007

© 2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Family and professionals underestimate quality of life across diverse cultures and health conditions: Systematic review

Authors:

Thomas F Crocker[a,*], Jaime K Smith[a,1], Suzanne M Skevington[b]

[a] Academic Unit of Elderly Care and Rehabilitation, Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Temple Bank House, Bradford Royal Infirmary, Duckworth Lane, Bradford BD9 6RJ, UK. Email: tom.crocker@bthft.nhs.uk

[b] Manchester Centre for Health Psychology, School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK.
Email: suzanne.skevington@manchester.ac.uk

[1] Caudwell Children, Minton Collins Building, Shelton Old Road, Stoke-On-Trent ST4 7RY.

Email: jaime.k.smith@hotmail.co.uk

* Corresponding author: Tom (Thomas) Crocker. Tel: +44 1274 383406;

Fax: +44 1274 382766. Email: tom.crocker@bthft.nhs.uk

Disclosures of financial support and potential conflicts of interest: No grants or specific funding were received to conduct this research. While conducting this research, TC & JS were employed by the Academic Unit of Elderly Care and Rehabilitation, Bradford Institute for Health Research. SS carried out this research at the Manchester Centre for Health Psychology, University of Manchester and directs the World Health Organisation’s Quality of Life research programme in UK at the International Hub for Quality of Life Research (IHQoLR) University of Manchester. The authors have no potential conflicts of interest to declare.

Abstract

Objective:

To examine how accurately proxies evaluate quality of life (QoL) in people they know, using cross-cultural data from the multidimensional, multilingual World Health Organization Quality of Life assessment short-form – The WHOQOL-BREF, and whether accuracy varies by health condition or proxy type (e.g. family/professional).

Study design and setting:

Systematic review with meta-analysis: We searched five databases for reports of proxy-completed WHOQOL-BREF scores and aggregated results using a random-effects model. Minimal clinically important difference values were calculated.

Results:

Analyses included nine studies (1980 dyads) of physical (n=762) or mental (n=604) health conditions, or intellectual disability (n=614), in 10 countries. Mean person-proxy correlations ranged from 0.28 (social QoL) to 0.44 (physical QoL). Proxy measures were underestimates (i.e. significantly lower than persons reported for themselves) for social (mean difference (MD) = 4.7, 95% confidence interval (CI) 1.8 to 7.6), psychological (MD=3.7, 95% CI 0.6 to 6.8), and physical (MD=3.1, 95% CI 0.6 to 5.6) QoL. Underestimates varied significantly between health conditions for social ($p<0.001$), psychological ($p=0.009$) and physical ($p=0.02$) QoL. Family members assessed psychological and environmental QoL better than professionals.

Conclusion:

Proxies tend to be imprecise, underestimating QoL, and should be aware of this tendency. Where health care is decided for others, family members’ views about QoL should be prioritised.

Keywords: quality of life, WHOQOL-BREF, proxy, systematic review, social comparison theory, minimal clinically important difference

Running title: Family and professionals underestimate quality of life

Key findings:

Proxies routinely underestimate social, psychological and physical, but not environmental QoL across cultures, relationships, and types of health condition. Proxies are usually imprecise, with the 95% confidence interval of correlation ranging from 0.24 to 0.6 for the physical QoL domain, which had the highest agreement.

What this adds to what is known:

This systematic review is the first to pool evidence on proxy-person concordance from 10 diverse countries worldwide using a single, comprehensive, validated, generic subjective quality of life (QoL) measure - the WHOQOL-BREF.

What is the implication, what should change now:

If health professionals increased family awareness about underestimation, and facilitated more accurate proxy judgements, this could improve and empower shared decision-making in health care.

Introduction

Proxies, such as friends, family members and health professionals, can provide important information about another person's QoL, especially where the other person has little or no means of communication. However, the growing trend to use subjective evaluations in health care through patient-reported outcome measures (PROMs) raises questions about how well a proxy can judge another person's subjective QoL. Where patients and proxies rate the patient, discrepancies are commonly reported and correlations low [1] and it is not clear whether these assessments provide inaccurate or biased information, as commonly assumed [2]. As an important aim of contemporary health care is to improve QoL for those with chronic diseases [3], empirical information from PROMs is essential to delivering high quality services. Our work is underpinned by the World Health Organisation's definition of QoL, which focuses on the subjective experience: *"An individual's perceptions of their position in life, in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards and concerns"* [4]. This definition implies that QoL is decided by making comparisons at personal, interpersonal and societal levels, in processes consistent with social comparison theory [5,6].

When completing QoL measures it is important to ascertain the similarity of a proxy answer to how the person would have replied had they responded directly for themselves. Close family or friends act as proxies for patients, but where this information is absent, health and social care professionals decide on care, so it is important to ascertain the accuracy of these assessments [7]. Subjective QoL is difficult for others to judge because information about it is largely invisible to observers, so accuracy is necessarily limited. There is growing consensus in the QoL field that *"no one can judge [QoL] better than the patient"* [8]. Even when asked, accuracy depends on a person’s ability to communicate about their QoL, and for the receiver to be listening and comprehending. These explanations have been offered for why person-proxy ratings are so low [1,9]. Sneeuw, Sprangers and Aaronson [10] found greater person-proxy agreement on health domains that were more observable. It is not clear whether proxy judgements of different QoL dimensions are the same for all important domains. While this has been tested for physical and psychological domains [10], a recent international generic instrument - The WHOQOL-BREF (The World Health Organization Quality of Life Assessment short-form) also measures environmental and social QoL domains as they are cross-culturally important to evaluating QoL in health care [11,12]. Testing person-proxy concordance across a profile of different domains can be valuable where the dimensions assembled within one measure apply the same metric.

As many investigations of proxy judgements of QoL have been conducted in Western countries, data from other cultures is scarce, impeding access to firm conclusions about whether low concordance could be a ‘universal’ characteristic of such judgements. A cross-cultural analysis is important to understanding the multiplicity of ways that people from diverse cultures judge their QoL and that of others. Another impediment to such investigations has been that developers of popular international generic health-related QoL assessments were slow to make non-English language versions available. Where translations exist, much needed cultural adaptation has been largely overlooked until recently. Conventional cross-cultural methodologies constrain the semantic and conceptual equivalence between language versions, subsequently limiting metric equivalence. A novel fast-track methodology devised to develop the WHOQOL-BREF [13,14] simultaneously created multiple language versions with higher equivalence levels than before [15]. Greater compatibility therefore enables us to combine QoL data from a single instrument; namely the WHOQOL-BREF.

Consequently, there is a strong case for completing a systematic review when sufficient evidence becomes available from studies using the same instrument, as this enables us to aggregate compatible data from the same domains, improving the quality of results. A systematic review that incorporates information about multiple health conditions and diverse cultures could provide insight into the concordance levels that might be expected from person-proxy information derived from a particular instrument, providing guidance about interpretation to health and social care professionals who use it.

There is a growing literature on using the WHOQOL-BREF for proxy assessment. We conducted a systematic review to aggregate WHOQOL-BREF findings, as this has not been done previously. The WHOQOL-BREF contains 25 internationally agreed important facets of subjective QoL scored within physical, psychological, social or environment domains. Reliable, valid [11,12], and sensitive to change [16,17], the WHOQOL-BREF is feasible and acceptable to use with sick and healthy adults [17].

The aim of the current systematic review of WHOQOL-BREF research was to examine proxies’ judgements of the QoL of people they know for specific QoL domains and to compare three broad groups of health conditions. An additional aim was to aggregate results from diverse cultures measured by one instrument (the WHOQOL-BREF), to take advantage of the instrument’s advanced design and multiple equivalent language versions [18]. This information enables us to better examine a case for the ‘universality’ of proxy behaviour. An additional aim was to examine the accuracy of proxy judgments and compare it for family and health professionals to find out who makes the most reliable judgements. This information has implications for delivering care to those unable to express their wishes.

2 Methods

We conducted a systematic review of the literature to examine the accuracy of estimates obtained from proxy administration of the WHOQOL-BREF. We refer to the individual whose QoL was judged by another as the *person,* and the individual who made that judgement as the *proxy*. Inclusion criteria were:

(i) The WHOQOL-BREF was completed by the person and a proxy;

(ii) The QoL of the person was judged by a proxy—completed by someone other than the person using his or her judgement about the other person's quality of life. (For clarity, we excluded proxy's judgements of their own QoL.)

(iii) All participants were adult, as defined by their culture. Although the WHOQOL-BREF has been used in proxy assessments of children and adolescents, it is only fully standardised for adult use so younger populations were excluded.

2.1 Search strategy

We developed the search strategy in collaboration with a librarian. Five electronic databases were searched (7th January, 2013) for articles published after 1990, without restriction to

Box 1: Example of search strategy used for PsycInfo:

 # term (number of results)

1 whoqol*.mp (678)

2 ((who or who's or whos) adj quality of life).mp. (147)

3 ((who or who's or whos) adj qol*).mp. (59)

4 world health organi#ation* quality of life.mp. (1082)

5 world health organi#ation* qol*.mp. (33)

6 "quality of life"/ and world health organi#ation*.mp. (898)

7 or/1-6 [WHOQOL] (1539)

8 (quality of life prox* or qol* prox*).mp. (9)

9 (proxy or proxies).mp. (4294)

10 (informant* or surrogate*).mp. (11,362)

11 (relative* or famil* or wife or wive* or husband* or partner* or spouse* or parent* or sibling* or sister* or brother* or mother* or father* or daughter* or son or sons or friend* or carer* or caregiver*).mp. (725,950)

12 (observer* or significant other*).mp. (29,709)

13 exp family members/ or wives/ or spouses/ or husbands/ or siblings/ or brothers/ or sisters/ or caregivers/ (120,287)

14 or/9-13 [Proxy] (759,419)

15 7 and 14 [WHOQOL and Proxy] (412)

16 8 or 15 [(Qol proxy) OR (WHOQOL and Proxy)] (420)

17 limit 16 to yr="1990 -Current" (420)

---

mp = title, abstract, heading word, table of contents, key concepts, original title, tests & measures

/ = subject heading, exp = explode, adj = adjacent language: PsycInfo, Medline, Science Direct, Embase and CINAHL. An example of the search strategy used for PsycInfo is presented in Box 1.[1] The overall strategy was: proxy AND (whoqol OR (who AND qol)) using various key words with appropriate stemming, wildcards and subject headings as applicable.

2.2 Data collection

Two authors (TC & JS) independently screened the title and abstract of references for eligibility. Only studies excluded by both authors were removed at this stage. At phase two, we sought full text articles and assessed them for eligibility. Disagreements were resolved through a consensus meeting with the third author (SS). Finally, the third author independently assessed the penultimate pool of papers.

We extracted data on the setting and characteristics of persons and proxies. We also extracted descriptive statistics (e.g. mean, standard deviation), and comparisons between (e.g. mean difference, correlation), person and proxy completion of the WHOQOL-BREF. One researcher extracted data, which the second then checked. We discussed and resolved discrepancies by consensus. Additional data and corrections to potential errors in reporting were resolved with study authors where appropriate and possible.

2.3 Data analysis

To examine the quality of these studies we developed a relevant assessment based on the Cochrane Collaboration’s risk of bias tool for interventions [19]. Risk was assessed against sets of criteria in four domains: participant selection, independent completion of the WHOQOL-BREF by person and proxy, reporting bias, and missing outcome data. For each

---

[1] Full details of search strategies for each database are available on request.

domain in every study we judged the risk as high, low, or unclear if insufficient information was available.

To assess the accuracy of proxy responses we assumed that the person's response was a true score of their QoL for each item and treated this as the reference value, so a 'perfect' proxy response would be identical to the relevant person. Therefore, we examined the correlation between proxy and person responses, which indicates the extent that proxies match persons who rate their QoL higher or lower than average. We also examined the mean difference that indicates the amount by which proxies systematically underestimate or overestimate the responses of the person overall. In summary, correlation assesses relative imprecision or the degree of randomness in proxy reports of QoL, while mean difference assesses bias [20].

A minimal clinically important difference (MCID) for the WHOQOL-BREF has not been established to the best of our knowledge. Since MCIDs are usually calculated for specific conditions [21] it is unclear how meaningful a generic MCID would be, but we present reference values here to assist interpretation. The WHOQOL-BREF is scored on a standardised (0-100) scale, where zero equals worst QoL and 100 equals best QoL. We calculated a 'small' difference (0.2 standard deviations) [22] from a broad sample (n=11,380) [12] as 3.6, 3.5, 4 and 3.2 for the physical, psychological, social and environment QoL domains, respectively. From the same study, differences in scores between well and sick samples were calculated as 14.4, 6.9, 5 and 1.9 for the physical, psychological, social and environment domains, respectively [12].

Although we planned to examine differences between pre- and post-intervention measures of the proxy-administered WHOQOL-BREF, this was not possible because it was infrequently reported, and person and proxy responses were usually amalgamated. We planned to pool multiple measures of correlation such as Kendall's tau and the intraclass correlation coefficient, but this was not possible, as they were infrequently reported.

We transformed correlation coefficients (Pearson's r) to a normal distribution and stable variance using Fisher's transformation (z), and calculated standard error from sample size. As various diagnostic groups were included (e.g. Alzheimer's, cancer), and different proxy types (e.g. spouse, health professional), we anticipated heterogeneity, so measures of agreement (z; mean difference) were combined using a random-effects model [23].

We conducted subgroup analyses based on participant types and agreed to group persons by broad health or disability condition: mental health, physical health, and intellectual disability. Proxy roles related to the person were grouped as close family or health care professional. As our studies did not report individual level data, we did not conduct subgroup analyses to examine age or gender.

3 Results

3.1 Study selection

INSERT FIGURE 1 HERE

We identified 1311 unique records through database searches and included 26 articles in a full-text assessment (kappa for eligibility screening = 0.66, good agreement [24]). Eleven studies [25-35] reported in 15 [25-39] of the 26 articles met the eligibility criteria, and were included in this review (kappa for full-text eligibility assessment = 0.62, good agreement [24]). Eight of the 11 studies [25,26,27,30-32,34,35] reported sufficient information to be included in the quantitative synthesis (see figure 1).

Of the 11 studies, one was a published abstract [29], and one used the WHOQOL-BREF to validate the Brazilian QOL-AD [33], so comparisons between WHOQOL-BREF responses were not the focal interest. Although meeting the inclusion criteria, we were not able to obtain relevant data, so these studies are not described further.

3.2 Characteristics of included studies

3.2.1 Study design

All included studies used a cross-sectional design. Person and proxy completed the WHOQOL-BREF (the proxy completed for the other) and both QoL reports were compared to examine concordance.

Additional objectives of included studies were to compare QoL measures for patient and carer groups with the general population and examine factors that predicted a person's QoL.

3.2.2 Participants

The studies contained 1980 dyads or triads in total. In seven studies each person was paired with one proxy, but in Schmidt 2010 [35] some persons had two proxies, and these analyses were presented separately. Our analysis took the responses of proxy one only, as this data set was more complete (n=614 (proxy one) vs. n=257 (proxy two)). For infertile couples in Chachamovich 2010 [30], both women and men were proxies for each other. After tossing a coin to select one pairing, we designated men as persons and women as proxies. For full details see table 1.

The studies contained results from 10 countries across five continents: Australia [28,31], Brazil [30,34,35], Czech Republic [35], Germany [35], Republic of Korea [32], Kuwait [25], Spain [35], Sudan [26,27], Turkey [35], and UK [35].

INSERT TABLE 1 ABOUT HERE

3.2.2.1 Persons

We grouped persons into the three broad categories of (a) **Mental health** (n=604): schizophrenia, major affective disorder, neurosis, bipolar disorder, psychosis [26,31,32]; (b) **Intellectual disability** (n=614): mild learning disability [35]; (c) **Physical health** (n=762): cancer, type 2 diabetes, multiple sclerosis, infertility [25,27,28,30,34]. As Awadalla 2005 [26] reported three groups of persons with mental health problems, we analysed these separately.

3.2.2.2 Proxies

In six studies, proxies were close family members: cohabiting partners and spouses [30,34], family members living-in or with regular contact [25-27]), and unspecified family members ([32]. In two studies, proxies were healthcare professionals: nurses [28] and case managers [30]. In one study, proxies were family, friends and professionals combined [35] so they were not included in these subgroup analyses.

3.2.3 Comparisons

The nine studies used to compare person and proxy reports from the WHOQOL-BREF employed various analytical techniques. Seven reported means and standard deviations in each WHOQOL-BREF domain for proxy and person [25,26,30-32,34,35]. Six reported associations between person and proxy scores for each WHOQOL-BREF domain using the Pearson correlation coefficient (r) [25,26,28,31,34,35]. These were the measures used in the quantitative analyses. In addition, two studies reported the intra-class correlation coefficient (ICC) [32,35]. There was insufficient data from one study [27] to include it in any quantitative analyses.

At an individual item level, only one study reported means, while two reported correlation coefficients, one using Pearson's r and ICC [35] and one Kendall's tau [25]. Two other studies examined item level agreements using Pearson's rho and kappa, but clustered results without reporting individual values [26,27].

3.3 Risk of bias

We typically judged risk of bias to be low or unclear. Only three studies were judged to have high risk of bias (see table 2). We judged risk of bias due to methods of participant selection to be low in seven studies, but unclear in two. We considered that the possibility of proxy and person knowing how each other had answered posed a low risk of bias in five studies, a high risk in three, and was unclear in one. Selective reporting of favourable results represented an unclear risk in all cases because, as expected, included studies had not published pre-study protocols (required for RCTs). We judged risk of bias due to incomplete outcome data as low in seven studies, but for two it was unclear.

INSERT TABLE 2 HERE

3.4 Results of studies and synthesis

3.4.1 Data handling

We transformed all mean domain scores to a scale from 0-100 commensurate with the WHOQOL-BREF protocol if they were reported in a different format. The social domain mean presented for persons with major affective disorders in Awadalla 2005 [26] are from data corrected by the authors, following contact.

3.4.2 Mean difference

Forest plots are presented in figure 2; full details are presented in tables A.1 and A.2 in appendix A. For the physical domain, the mean difference between person and proxy scores (0 to 100 scale) was estimated as 3.1 (95% CI: 0.6 to 5.6, p=0.02). We found significant between-study heterogeneity (I^2=74%, Chi^2=46 on 10 degrees of freedom (df), p<0.001).

The psychological domain mean difference within dyads was estimated as 3.7 (95% CI: 0.6 to 6.8, p=0.02); between-study heterogeneity was significant (I^2=85%, Chi^2=67 on 10 df, p<0.001).

The mean person-proxy difference for the social relationships domain was 4.7 (95% CI: 1.8 to 7.6, p=0.001), and there was between-study heterogeneity (I^2=77%, Chi^2=43 on 10 df, p<0.001).

The environment domain within-dyad mean difference was 1.2 (95% CI: -1.4 to 3.7, p=0.4), and between-study heterogeneity was also significant (I^2=81%, Chi^2=54 on 10 df, p<0.001).

INSERT FIGURE 2 HERE

3.4.3 Correlation

The Pearson correlation coefficient between person and proxy scores was estimated for each domain. The physical domain correlation was r=0.44 (95% CI: 0.24 to 0.60, p<0.001). This showed significant between-study heterogeneity (I^2=89%, Chi^2=27 on 3 df, p<0.001).

For the psychological domain, the person-proxy correlation was r=0.34 (95% CI: 0.21 to 0.45, p<0.001), and between-study heterogeneity was not significant (I^2=0%, Chi^2=10 on 3 df, p=0.4).

The social domain dyadic correlation was r=0.28 (95% CI: 0.18 to 0.38, p<0.001) but heterogeneity was significant (I^2=53%, Chi^2=6.4 on 3 df, p=0.09).

The environment domain correlation was r=0.32 (95% CI: 0.26 to 0.38, p<0.001) and heterogeneity was not significant (I^2=0%, Chi^2=2.2 on 3 df, p=0.5).

3.4.4 Per item correlation

On each item, proxy-person correlations were reported using different statistics in different studies so could not be quantitatively combined, but common trends are presented. Across the six groups where item correlations were reported, three facets had particularly small, and in some cases negative correlations: dependence on medication or treatment, pain and discomfort, and positive feelings. Five facets on energy and fatigue, sex-life, financial resources, mobility, and transport showed the closest proxy-person associations.

3.4.5 Subgroup analyses

Investigating possible sources of heterogeneity was limited by the small number of studies, so the following results deserve cautious interpretation. Details for table 3 and table 4 are summarised here. We found some evidence that the person’s health condition affected the size of the mean difference between person and proxy in the physical (p=0.02), psychological (p=0.009) and social (p<0.001) domains, but not environment (p=0.5). Results for the physical domain showed significantly greater underestimation of QoL by proxies for persons with mental health conditions than intellectual disabilities.

In the psychological and social domains we found a significantly smaller underestimation of QoL by proxies for persons with mental health conditions than for intellectual disabilities. Proxies of persons with physical health conditions also underestimated QoL in the social domain significantly less than proxies of persons with intellectual disabilities did.

We found no evidence that underestimating QoL was significantly greater for family members than health professionals in the physical (p=0.5), psychological (p=0.5) or social (p=0.1) domains (Table 3). However, there was a significant difference with respect to a person's environmental QoL (p<0.001), which health care professionals significantly underestimated while close family members did not.

Evidence supported the view that the person's condition affected the degree of person-proxy correlation in the physical domain (p<0.001), where the correlation was significantly smaller for persons with intellectual disabilities compared to physical or mental health conditions. This pattern was not found in the psychological (p=0.4), social (p=0.7) and environment (p=0.6) domains. While person-proxy correlations tended to be higher in each domain when the proxy was close family, family were only significantly more precise proxies than professionals, in the psychological domain (p=0.01).

Participants from different countries tended to underestimate QoL in all four domains. Exceptionally the Sudanese samples tended to overestimate environmental QoL, but these three samples came from the same study [26]. Although the sample in Chachamovich [30] overestimated environmental QoL, the other Brazilian sample did not [34], in line with other cultures.

INSERT TABLE 3 HERE

INSERT TABLE 4 HERE

4 Discussion

The subjective quality of life of a person is largely invisible, and therefore difficult for others to know. We wanted to examine whether there would be systematic differences in person-proxy judgements about important aspects of subjective QoL. The WHOQOL-BREF profile enables us to do this by providing a high quality internationally standardised instrument containing physical, psychological, social and a unique environment domain. Because multiple language versions are available, we could identify studies conducted in diverse cultures worldwide, where the WHOQOL-BREF was used to obtain proxy judgements about people with physical and mental health conditions and intellectual disability, and to compare them with the person's own report. Rarely have researchers been able to satisfactorily investigate whether these properties of judgement approach 'universality', by using an equivalent metric. This equivalence in measurement provided by multiple language versions of the WHOQOL-BREF enables us to take one important step in this direction, despite sample heterogeneity.

Proxy reports of QoL are only weakly or moderately related to those provided by the persons themselves. There was greater overall agreement about QoL in the physical domain than in the psychological, social or environment domains. On average, close family gave more precise judgements of QoL than professionals, although this was only significant in the psychological domain. In contrast, proxies for people with intellectual disabilities were more imprecise than for other types of health condition, although this was only statistically significant in the physical domain.

Greater agreement found for physical QoL concurs with findings by Sneeuw, Sprangers and Aaronson [10]. Previously this was explained as proxies relying on the more visible signs of factors that they believe underpin QoL. According to this perspective, social and psychological domains are more difficult for proxies to judge, even where communications channels are relatively normal. Communication with people who have intellectual disabilities is widely acknowledged as problematic [e.g. 40]. This may explain the poorer estimates of QoL by comparison with other health conditions, as proxies find it harder to understand the perspectives of those with intellectual disability. However, this argument appears to contradict the reason given by Sneeuw, Sprangers and Aaronson for greater overall agreement on the physical domain (that it is more observable) [10]. The lowest average correlation of any subgroup for any domain was for intellectual disabilities in the physical domain. If proxies relied on physical signs rather than communication, we might expect greater disparity between the correlations for the physical domain and those for social or psychological domains, in a group with less ability to communicate. Yet average correlations in these domains are similar where the person has intellectual disability, unlike physical and mental health problems.

Proxies tended to underestimate the subjective QoL of others using the WHOQOL-BREF and we found negligible evidence of systematic overestimates. Family members and health professionals usually believe that the subjective QoL of those they care for is worse than the person reports themselves, especially in the social, psychological and physical domains. In these domains the type of health condition made an important difference to the extent to which proxies underestimate QoL. In particular, people with intellectual disability have much better psychological and social QoL than is assumed by their proxies. There was also widespread bias in estimates of physical and psychological QoL for persons with physical health conditions, suggesting that for some of the conditions we included there are large underestimates, while for others there may be none.

Although both informal and formal carers tended to underestimate the QoL of those they care for, we found that family members were significantly better at estimating the person's environmental QoL than professionals (nurses and case managers). Family members were unbiased, while professionals systematically underestimated environmental QoL. This new discovery was made possible by using the WHOQOL BREF, which is unique among international generic measures in assessing environmental QoL. As family members usually spend more time in a shared environment that professionals have little access to, it is unsurprising that they were able to make better judgements as they have more tangible contextual information about the person's environmental QoL. However, they were not significantly more precise, just unbiased. It may be that professionals often hold less positive opinions of the environments their patients inhabit than those who share it with them. As more data becomes available from a broader pool of proxies (e.g. doctors, professional carers) these results should be revisited and differences among professionals and family relations explored.

In most countries, proxies tended to underestimate QoL in all four domains of the WHOQOL-BREF. However, environmental QoL tended to be overestimated in two low- and middle-income countries (Brazil and Sudan) where greater awareness of material, 'objective' conditions may have had a different impact on judging subjective QoL, than in high-income countries. However, these studies also investigated very different health conditions, the same result was not repeated in a second Brazilian study, and other factors (e.g. youth, intimacy) might influence overestimation of another's QoL. Without new published data we are unable to conclude when proxies will underestimate QoL, although it seems to be a tendency.

The response shift model in which social comparison processes play a part could explain underestimation of QoL by proxies, as this model supports the view that patients adapt their

QoL following the onset of a chronic or life threatening condition. During this process the event alters their internal values, standards and/or conceptualisation of QoL. As these concepts are similar to those found in the WHO definition of QoL, namely "…goals, expectations, standards and concerns", this feature indicates a degree of theoretical consensus. People assess QoL differently after a life-changing event when they have adapted to their situation [41]. For carers, understanding the process whereby response shift works, namely how person and proxy evaluations of QoL may change over time, would enable them to interpret WHOQOL-BREF scores more meaningfully for this purpose.

Our review draws on social comparison theory, which underpins the WHO definition of QoL and is widely used in health (e.g. [5]). Our theoretical application indicates that, in this context, proxies make downward comparisons, assessing people that they care for as less fortunate than themselves [5]. The evaluation process we have studied has led us to reconsider the WHOQOL definition for use in this context, so as to incorporate the proxy view: *"An individual's perceptions of (other people's) position in life, in the (shared) context and (shared) value systems in which they live…"*. It may also help to explain why family members appear to be more accurate proxies. It seems quite plausible that family members share more *"goals, expectations, standards and concerns"* than they share with professionals, and these issues affect judgements about the other's QoL, as well as consideration of their own. We would therefore expect health and other professionals to have different but overlapping sets of goals, expectations, standards for, and concerns about others, compared with family members, due to their training, professional experience and service aims. This theoretical reanalysis at a societal level may partly explain why they arrive at different QoL conclusions in some domains. New theoretically-driven research is needed in this field. More generally, the process whereby people underestimate the QoL of others deserves further work to understand which diseases and conditions proxies underestimate most, in which conditions QoL is most accurately perceived and why. This information would be valuable in situations where proxies provide the ‘voice’ for vulnerable people who do not have one.

There are some limitations to our research. As the WHOQOL-BREF is only validated for adults we only included adult samples of dyads; the results may differ where parents are proxies for their children, compared to other family relations. Other roles and other types of health professionals should also be investigated in the future. To examine scores given by proxies, we were only able to include samples where the person and the proxy were judged capable of completing the WHOQOL-BREF by the study researchers. We therefore relied on them to exclude from their results those incapable of completing the measure. The selected studies provide reasonable confidence that the questions were understood. Moreover, our results show that correlations for the physical health group were not significantly different from the psychological health or intellectual disabilities groups, except for physical QoL. Some of the proxy - person difference, and heterogeneity between studies, may be accounted for by the way questions were adjusted for proxy completion [42]. Different answers could have been given by proxies who were asked to ‘stand in the others shoes’, compared with those just asked to assess the other during completion. As few studies reported this information, we were unable to disaggregate it. More studies on proxy perspectives are required.

Where feasible, a person must be asked about their own subjective experience of QoL for reasons of accuracy. Although in some conditions and for some domains, close family may be expected to provide reports that correlate moderately with the person’s QoL, this is unlikely to be sufficiently reliable for confident use as a surrogate measure. The low concordance found in this study indicates that efforts to produce new versions of the WHOQOL that facilitate completion by people who have difficulty understanding or communicating [35,43], are well placed.

To improve clinical decision-making when the person’s view is not available, the evidence in this review suggests professionals should defer to the views of close family members, especially when psychological and social QoL are under consideration, and for specific facets of environmental QoL, such as perceived financial resources and transport. However, further research should test this assertion for a greater variety of health care professionals, family members and person conditions. To assist family members we should tell them about the implicit tendency to underestimate another person’s social, psychological and physical QoL, especially for carers of people with intellectual disability, where the gaps in perceptions of social and psychological QoL were the greatest. Such information may provide reassurance that people perceive these dimensions of QoL to be better than many proxies think.

In the light of increasing life expectancy in many parts of the world, we should inform and train health professionals, especially those working with older adults, to take account of the empirical evidence of widespread underestimation and imprecision. This detailed information may help to empower professionals and familial carers to make better shared decisions at critical times in the lifespan of the person for whom they care.

5 Conclusion

Proxies are generally poor judges of a person’s QoL and tend to underestimate the QoL of the individual they care for across a diverse range of contexts. However, assessments by family members tend to be more similar to the person’s own QoL perspective than those by professionals (nurses and case managers). The ability of proxies to assess a person’s QoL is inconsistent across the important internationally agreed QoL dimensions captured by the WHOQOL-BREF, where the physical domain is judged more accurately than other domains. However, details of these patterns vary depending on the person’s type of disability or health condition, and their relationship with the proxy. This information should be recognised and used when family and health professionals share decisions about the type and level of care a person receives on their behalf.

References

[1] Cummins RA. Proxy responding for subjective well-being: A review. Int Rev Res Ment Ret. 2002;25:183-207. doi: 10.1016/S0074-7750(02)80009-X.

[2] Sneeuw KC, Aaronson NK, Osoba D, Muller MJ, Hsu MA, Yung WK, Brada M, Newlands ES. The use of significant others as proxy raters of the quality of life of patients with brain cancer. Med Care. 1997 May;35(5):490-506. PubMed PMID: 9140337.

[3] Department of Health. Guidance: Collection and reporting of patient reported outcome measures (PROMs) data [Internet]. London: Department of Health (GB); 2012 Jul 5 [cited 2013 Dec 21]. Available from: https://www.gov.uk/government/publications/collection-and-reporting-of-patient-reported-outcome-measures-proms-data

[4] World Health Organisation. Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). Qual Life Res. 1993 Apr;2(2):153-9. PubMed PMID: 8518769.

[5] Wheeler L. (1991) A brief history of social comparison theory. In: Suls J, Wills TA, editors. Social comparison: contemporary theory and research. Hillsdale (NJ): Lawrence Erlbaum; 1991. p. 3-22.

[6] Skevington SM. Social comparisons in cross-cultural quality of life assessment. Int J Mental Health. 1994;23(2):29-47.

[7] Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. Med Care. 2005 May;43(5):493-9. PubMed PMID: 15838415; PubMed Central PMCID: PMC1188232.

[8] Wu AW, Snyder C, Clancy CM, Steinwachs DM. Adding the patient perspective to comparative effectiveness research. Health Aff (Millwood). 2010 Oct;29(10):1863-71. doi: 10.1377/hlthaff.2010.0660. PubMed PMID: 20921487.

[9] Cummins RA, Eckersley R, Pallant J, van Vugt J, Misajon RA. Developing a national index of subjective well-being: The Australian Unity Well-being Index. Soc Indic Res. 2003 Oct;64(2):159-90. doi: 10.1023/A:1024704320683

[10] Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. J Clin Epidemiol. 2002 Nov;55(11):1130-43. doi: 10.1016/s0895-4356(02)00479-1. PubMed PMID: 12507678.

[11] The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med. 1998 May;28(3):551-8. PubMed PMID: 9626712.

[12] Skevington SM, Lotfy M, O’Connell KA; WHOQOL Group. The World Health Organization’s WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res. 2004 Mar;13(2):299-310. doi: 10.1023/B:QURE.0000018486.91360.00. PubMed PMID: 15085902.

[13] Skevington SM, Sartorius N, Amir M and the WHOQOL Group. Developing methods for assessing quality of life in different cultural settings. The history of the WHOQOL

instruments. Soc Psychiatry Psychiatr Epidemiol. 2004 Jan;39(1):1-8. doi: 10.1007/s00127-004-0700-5. PubMed PMID: 15022040.

[14] Skevington SM. Measuring quality of life in Britain: introducing the WHOQOL-100. J Psychosom Res. 1999 Nov;47(5):449-59. doi: 10.1016/S0022-3999(99)00051-3. PubMed PMID: 10624843.

[15] Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. Soc Sci Med. 2003 Oct;57(7):1289-306. Review. doi:10.1016/S0277-9536(02)00503-8. PubMed PMID: 12899911.

[16] Diehr PH, Derleth AM, McKenna SP, Martin ML, Bushnell DM, Simon G, Patrick DL. Synchrony of change in depressive symptoms, health status, and quality of life in persons with clinical depression. Health Qual Life Outcomes [Internet]. 2006 Apr 25 [cited 2013 Dec 21];4(1):27. doi:10.1186/1477-7525-4-27. PubMed PMID: 16638129; PubMed Central PMCID: PMC1524937. Available from: http://www.hqlo.com/content/4/1/27

[17] Skevington SM, McCrate FM. Expecting a good quality of life in health: assessing people with diverse diseases and conditions using the WHOQOL-BREF. Health Expect. 2012 Mar;15(1):49-62. doi: 10.1111/j.1369-7625.2010.00650.x. PubMed PMID: 21281412.

[18] Skevington SM. Advancing cross-cultural research on quality of life: observations drawn from the WHOQOL development. World Health Organisation Quality of Life

Assessment. Qual Life Res. 2002 Mar;11(2):135-44. doi: 10.1023/A:1015013312456. PubMed PMID: 12018737.

[19] Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [Internet]. The Cochrane Collaboration; 2011 Mar [cited 2013 Dec 21]. Available from: www.cochrane-handbook.org.

[20] Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with continuous measurements. J Biopharm Stat. 2007 Aug;17(4): 529-69. doi: 10.1080/10543400701376480. PubMed PMID: 17613641.

[21] Leopold SS. Editor's spotlight/take 5: Comparative responsiveness and minimal clinically important differences for idiopathic ulnar impaction syndrome. Clin Orthop Relat Res. 2013 May;471(5):1403-5. doi: 10.1007/s11999-013-2886-x. Epub 2013 Mar 5. PubMed PMID: 23460486; PubMed Central PMCID: PMC3613524.

[22] Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T; Clinical Significance Consensus Meeting Group. Estimating clinically significant differences in quality of life outcomes. Qual Life Res. 2005 Mar;14(2):285-95. PubMed PMID: 15892420.

[23] Cooper H, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis. 2nd ed. New York:Russell Sage Foundation; 2009 Feb. 632 p.

[24] Orwin RG. Evaluating coding decisions. In: Cooper H, Hedges LV, editors. The handbook of research synthesis. 1st ed. New York: Russell Sage Foundation; 1994. p. 139-63.

[25] Alshubaili AF, Awadalla AW, Ohaeri JU, Mabrouk AA. Relationship of depression, disability, and family caregiver attitudes to the quality of life of Kuwaiti persons with multiple sclerosis: a controlled study. BMC Neurol. 2007 Sep 18;7:31. doi: 10.1186/1471-2377-7-31. PubMed PMID: 17877820; PubMed Central PMCID: PMC2041952.

[26] Awadalla AW, Ohaeri JU, Salih AA, Tawfiq AM. Subjective quality of life of community living Sudanese psychiatric patients: comparison with family caregivers' impressions and control group. Qual Life Res. 2005 Oct;14(8):1855-67. doi: 10.1007/s11136-005-4328-z. PubMed PMID: 16155773.

[27] Awadalla AW, Ohaeri JU, Tawfiq AM, Al-Awadi SA. Subjective quality of life of outpatients with diabetes: comparison with family caregivers' impressions and control group. J Natl Med Assoc. 2006 May;98(5):737-45. PubMed PMID: 16749649; PubMed Central PMCID: PMC2569271.

[28] Bahrami M, Parker S, Blackman I. Patients' quality of life: a comparison of patient and nurse perceptions. Contemp Nurse. 2008 May;29(1):67-79. PubMed PMID: 18844544.

[29] Bhandari RP and Joshi SR. Quality of life in dementia patients in a developing country: Comparison of three approaches. Alzheimer's and Dementia. 2011 Jul;7(4 Suppl):S638. doi: 10.1016/j.jalz.2011.05.1825.

[30] Chachamovich JR, Chachamovich E, Ezer H, Fleck MP, Knauth DR, Passos EP. Agreement on perceptions of quality of life in couples dealing with infertility. J Obstet Gynecol Neonatal Nurs. 2010 Sep-Oct;39(5):557-65. doi: 10.1111/j.1552-6909.2010.01168.x. PubMed PMID: 20920002.

[31] Herrman H, Hawthorne G, Thomas R. Quality of life assessment in people living with psychosis. Soc Psychiatry Psychiatr Epidemiol. 2002 Nov;37(11):510-8. doi: 10.1007/s00127-002-0587-y. PubMed PMID: 12395140.

[32] Kim EJ, Song DH, Kim SJ, Park JY, Lee E, Seok JH, Jon DI, Cho HS. Proxy and patients ratings on quality of life in patients with schizophrenia and bipolar disorder in Korea. Qual Life Res. 2010 May;19(4):521-9. doi: 10.1007/s11136-010-9617-5. PubMed PMID: 20204707.

[33] Novelli MM, Nitrini R, Caramelli P. Validation of the Brazilian version of the quality of life scale for patients with Alzheimer’s disease and their caregivers (QOL-AD). Aging Ment Health. 2010 Jul;14(5):624-31. doi: 10.1080/13607861003588840. PubMed PMID: 20480421.

[34] Rabin EG, Heldt E, Hirakata VN, Bittelbrunn AC, Chachamovich E, Fleck MP. Depression and perceptions of quality of life of breast cancer survivors and their male partners. Oncol Nurs Forum. 2009 May;36(3):E153-8. doi: 10.1188/09.ONF.E153-E158. PubMed PMID: 19403443.

[35] Schmidt S, Power M, Green A, Lucas-Carrasco R, Eser E, Dragomirecka E, Fleck M. Self and proxy rating of quality of life in adults with intellectual disabilities: results from the DISQOL study. Res Dev Disabil. 2010 Sep-Oct;31(5):1015-26. doi: 10.1016/j.ridd.2010.04.013. PubMed PMID: 20478692.

[36] Alshubaili AF, Ohaeri JU, Awadalla AW, Mabrouk AA. Family caregiver quality of life in multiple sclerosis among Kuwaitis: a controlled study. BMC Health Serv Res. 2008 Oct 7;8:206. doi: 10.1186/1472-6963-8-206. PubMed PMID: 18840287; PubMed Central PMCID: PMC2576463.

[37] Awadalla AW, Ohaeri JU, Salih AA, Tawfiq AM. Subjective quality of life of family caregivers of community living Sudanese psychiatric patients. Soc Psychiatry Psychiatr Epidemiol. 2005 Sep;40(9):755-63. doi: 10.1007/s00127-005-0947-x. PubMed PMID: 16151598.

[38] Awadalla AW, Ohaeri JU, Al-Awadi SA, Tawfiq AM. Diabetes mellitus patients' family caregivers' subjective quality of life. J Natl Med Assoc. 2006 May;98(5):727-36. PubMed PMID: 16749648; PubMed Central PMCID: PMC2569270.

[39] Kim EJ, Kim SJ, Cho HS. Quality of life in patients with schizophrenia and bipolar disorder: Comparison of self-report and proxy assessment. European Psychiatry. 2009;24 Suppl 1:S569. doi: 10.1016/S0924-9338(09)70802-8.

[40] Perry J, Felce D. Objective assessments of quality of life: How much do they agree with each other? J Community Appl Soc Psychol. 1995 Feb;5(1):1-19. doi: 10.1002/casp.2450050102.

[41] Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med. 1999 Jun;48(11):1507-15. doi: 10.1016/S0277-9536(99)00045-3. PubMed PMID: 10400253.

[42] Pickard AS, Lin HW, Knight SJ, Sharifi R, Wu Z, Hung SY, Witt WP, Chang CH, Bennett CL. Proxy assessment of health-related quality of life in African American and white respondents with prostate cancer: perspective matters. Med Care. 2009 Feb;47(2):176-83. doi: 10.1097/MLR.0b013e31818475f4. Erratum in: Med Care. 2009 Apr;47(4):491. Knight, Sara L [corrected to Knight, Sara J]. PubMed PMID: 19169118; PubMed Central PMCID: PMC3215256.

[43] Power MJ, Green AM; WHOQOL-Dis Group. Development of the WHOQOL disabilities module. Qual Life Res. 2010 May;19(4):571-84. doi: 10.1007/s11136-010-9616-6. PubMed PMID: 20217246.

Acknowledgements

Sally Dalton, University of Leeds for support in development and execution of the search strategy.Elizabeth Teale, Bradford Institute for Health Research for statistical advice and support.

John Young and Anne Forster, Leeds Institute of Health Sciences and Bradford Institute for Health Research for providing TC and JS with the time for this research to be conducted.

Lesley Brown and Anne Forster for providing suggestions based on previous drafts.

Table and figure legends

Figure 1. PRISMA flow diagram

Table 1. Characteristics of studies

† For infertile couples in Chachamovich 2010 [30], both women and men were proxies for each other. After tossing a coin to select one pairing, we designated men as persons and women as proxies.

‡ In Schmidt 2010 [35] some persons had two proxies, and these analyses were presented separately. Our analysis took the responses of proxy one only, as this data set was more complete (n=614 (proxy one) vs. n=257 (proxy two)).

Table 2. Authors' judgement of risk of bias across four domains

↓ indicates low risk of bias, ↑ indicates high risk of bias, ? indicates an unclear risk of bias

Figure 2. Forest plot of mean difference and correlation between person and proxy for the four domains of the WHOQOL-BREF

Meta-analysis using inverse-variance method and the random-effects model. Plot indicates mean and 95% CI. Full details available in appendix A

Country codes AU: Australia; BR: Brazil; KR: Korea, Republic of; KW: Kuwait; SD: Sudan; XZ: Multinational (Brazil, Czech Republic, Germany, Spain, Turkey and UK)

Table 3. Mean differences and correlations of proxy responses for different person groups across the four domains of the WHOQOL-BREF

† Mean difference is between person and proxy ratings on the WHOQOL-BREF 0-100 scale. Difference was calculated as person minus proxy. Therefore, higher scores indicate underestimation by the proxy

‡ Chi^2 test for subgroup differences.

§ r^2 indicates the proportion of variance in the scores of persons explained by the scores estimated by the proxies

* indicates P<0.05, ** indicates P<0.01, *** indicates P<0.001

Table 4. Mean differences and correlations of proxy responses for different proxy groups across the four domains of the WHOQOL-BREF

† Mean difference is between person and proxy ratings on the WHOQOL-BREF 0-100 scale. Difference was

calculated as person minus proxy. Therefore, higher scores indicate underestimation by the proxy

‡ Chi^2 test for subgroup differences.

§ r^2 indicates the proportion of variance in the scores of persons explained by the scores estimated by the proxies

* indicates $P<0.05$, ** indicates $P<0.01$, *** indicates $P<0.001$