

RESEARCH ARTICLE

Open Access

# Using mixed methods to select optimal mode of administration for a patient-reported outcome instrument for people with pressure ulcers

Claudia Rutherford<sup>1,4\*</sup>, Jane Nixon<sup>1</sup>, Julia M Brown<sup>1</sup>, Donna L Lamping<sup>2^</sup> and Stefan J Cano<sup>3</sup>

## Abstract

**Background:** When developing new measuring instruments or deciding upon one for research, consideration of the 'best' method of administration for the target population should be made. Current evidence is inconsistent in differentiating superiority of any one method in terms of quantity and quality of response. We trialed a novel mixed methods approach in early scale development to determine the best administration method for a new patient-reported outcome instrument for people with pressure ulcers (the PU-QOL).

**Methods:** Cognitive interviews were undertaken with 35 people with pressure ulcers to determine appropriateness of a self-completed version of the PU-QOL instrument. Quantitative analysis, including Rasch analysis, was carried out on PU-QOL data from 70 patients with pressure ulcers, randomised to self-completed or interview-administered groups, to examine data quality and differential item functioning (DIF).

**Results:** Cognitive interviews identified issues with PU-QOL self-completion. Quantitative analysis supported these findings with a large proportion of self-completed PU-QOLs returned with missing data. DIF analysis indicated administration methods did not impact the way patients from community care settings responded, supporting the equivalence of both administration versions.

**Conclusions:** Obtaining the best possible health outcomes data requires use of appropriate methods to ensure high quality data with minimal bias. Mixed methods, with the inclusion of Rasch, provided valuable evidence to support selection of the 'best' administration method for people with PUs during early PRO instrument development. We consider our approach to be generic and widely applicable to other elderly or chronically ill populations or suitable for use in limited samples where recruitment to large field tests is often difficult.

## Background

High quality health outcomes research requires patient-reported outcomes (PROs) [1-3]. PRO instruments should be reliable, valid and able to detect clinical change over time [3,4]. Consideration of appropriate administration mode should also be made. Comparisons of the two main administration methods (interviewer and self-completed) have shown mixed results: higher item-response rates were found with administered methods, while others reported inconsistent effects [5]; one study found that different methods do not have a meaningful effect on repeated

PRO measurements [6] while another reported biasing influences on the responses obtained [7]. Respondents are also less likely to give no answer or respond "don't know" when self-completing [8]. A review of PRO instruments applied in older people found best completion rates following interview administration [9]. These findings are consistent with evidence suggesting completion difficulties increase with age, declining cognition and deteriorating health [10].

Determining 'best' administration mode for PRO instruments is key in the development process and usually tested through large scale field testing [11-13]. Ascertaining the appropriateness of different methods should take into account the: population; topic and setting; anticipated response rates; acceptability; and time available [5]. Additionally, consideration of bias from sources

\* Correspondence: claudia.rutherford@sydney.edu.au

<sup>^</sup>Deceased

<sup>1</sup>Clinical Trials Research Unit (CTRU), University of Leeds, Leeds, UK

<sup>4</sup>QOL Office, University of Sydney, Sydney 2006, Australia

Full list of author information is available at the end of the article

other than non-response, for instance, equivalence of different mode versions of the same instrument, should be made. As new psychometric methods, such as Rasch Measurement Theory (RMT) [14], are able to provide useful exploratory data in small samples ( $n = 30$ ) [15], there is good potential to use these to help determine 'best' administration mode in early instrument development.

Pressure ulcers (PU) are a chronic wound that can occur when the skin and underlying tissue becomes damaged due to pressure or pressure in combination with shearing forces [16]. PUs are highly prevalent, a challenge to health-care professionals, and a major problem for high-risk populations including the mobility impaired and the elderly [16,17]. Severe PUs can become a long-term chronic condition requiring extensive management and consequently reducing health-related quality of life (HRQL) [18]. Thus, assessment of PROs is particularly important and relevant in this disease area however 'best' methods of assessment need to be determined.

Few studies have used standardised PRO instruments with elderly people with chronic wounds [19] thus, there is little evidence pertaining to acceptability and appropriateness of administration methods for this population. Previous explorations have been conducted with general samples (e.g. mixed elderly) and the current evidence is inconsistent in differentiating superiority of any one method in quantity and quality of response; failing to support choice of administration mode. Further, people who develop PUs are largely elderly, highly dependent and/or with high levels of co-morbidity, making them a unique group.

We previously developed a PRO instrument for people with PUs (the PU-QOL instrument) intended for patient self-completion [20]. However, pretesting identified problems with item-response rates, questioning the suitability of self-completion for this patient group, particularly those aged over 70 years. This study uses a novel mixed methods approach to provide direction for the 'best' administration mode for the PU-QOL instrument. Specifically, we investigated differences between two administration groups to determine whether one instrument could be developed for use with both self-completed and interview-administered methods (similar responses between groups would support one version suitable for both methods) or whether two mode-specific versions were required (divergent responses would require two administration mode-specific versions).

## Methods

### Study design and sample

We investigated 'best' administration mode through: 1) semi-structured cognitive interviews with 35 participants with PUs to determine the appropriateness of and reasons for any difficulty with self-completion (study methods

described elsewhere [20]); and 2) quantitative methods with the inclusion of RMT on PU-QOL data from patients randomised to self-completed or interview-administered groups to examine data quality and differential item functioning (DIF). We anticipated a sample of around 100 would meet the data requirement for DIF analyses [15].

Consecutive patients from 31 hospital and community National Health Services (NHS) around the UK, with existing PUs of any severity [16], location or duration; aged over 18 years; and able to understand English were recruited between September 2009 and August 2010. Patients with only moisture lesions or who were unconscious, confused, cognitively impaired or deemed ethically inappropriate to approach (e.g. death was imminent) were excluded. To ensure equivalent clinical presentation in both administration groups, only patients able to read and write in English were included. Ethical approval was provided by a UK NHS Research Ethics Committee and all participants gave written informed consent to participation.

### Data collection procedures

To ensure the DIF analysis was a valid interpretation of group differences - in this instance, differences dependent on administration mode and not an artefact of differences within groups - through application of the eligibility criteria, participants were matched on clinical presentation and relevant underlying ability (e.g. with an existing PU; able to read and write independently) before determining equivalence of responses to scale items. Participants were then randomised to one of two groups: self-completed or interview-administered groups through a 2:1 ratio. The 2:1 ratio was used to account for the likelihood of increased missing data from self-completed PU-QOLs [20]. Randomisation was stratified by: age ( $\leq 70$ ,  $>70$  years), wound severity (superficial, severe) and healthcare setting (hospital, community).

Patients randomised to the self-complete group were provided with the PU-QOL and instructed to complete the instrument on their own. Those randomised to the interview-administered group had the PU-QOL administered to them by a tissue viability team member, following an interview user manual. Training in administering the PU-QOL was provided by one researcher (CR) to ensure standardisation across administrations.

### PU-QOL instrument

The PU-QOL version used in this study consisted of 13 scales (87-items): pain; exudate; odour; sleep; vitality; mobility; daily activities; mood; anxiety; self-consciousness and appearance; autonomy; isolation; and participation. Scales represent unique outcomes represented in a conceptual framework of HRQL specific to PUs [21]. Questions focused upon the impact of PUs on these constructs, rated by the amount of bother attributed (e.g. "During the

past week, how much have you been bothered by...?”) on a 4-point response scale (e.g. 0 = no bother – 3 = a lot of bother). A recall period of the past-week was chosen on clinical grounds, as changes in PU severity and symptomatology often occur over days and thus a longer recall period would risk not capturing relevant impact on HRQL.

### Analyses

The qualitative analysis involved identifying dominant trends (e.g. issues occurring repeatedly) and key findings (e.g. issues reported once but considered severe). Findings were categorised by mode preference, ease of self-completion, and reasons for any difficulty. We calculated the proportion of: completed and returned PU-QOLs (response rate) and missing data (data quality) per PU-QOL and per item by mode group. A Rasch analysis was performed on each of the 13 PU-QOL scales to examine DIF [14,22,23]. The measurement properties of the PU-QOL instrument were subsequently tested in a large field test [24].

RMT provides a formal method for evaluating scale functioning against a sophisticated mathematical measurement model [25]. The Rasch model defines how a set of items should perform to generate reliable and valid measurements [26] and evaluates the legitimacy of summing items to generate those measurements [14,22]. The extent to which observed data (patients' actual responses to scale items) are concurrent with ('fit') predictions of those responses from the Rasch model are examined; whereby the difference between expected and observed scores indicates the degree to which rigorous measurement is achieved [27]. The expected response structure is a probabilistic Guttman pattern, which assumes that for the same person ability, the probability of endorsing an easy item is higher than the probability of endorsing a more difficult item, and vice versa [28]. When a PRO instrument is used to discriminate between persons with different abilities, someone with higher ability is expected to affirm all items endorsed by a person with lower ability in addition to items representative of higher ability.

### Rasch analysis: differential item functioning (DIF)

DIF analysis [29] is a technique for investigating conditional relationships between item response and group membership [30]. It is based on the assumption that respondents with similar ability (determined by total scores) should respond in similar ways to individual items regardless of gender, age or ethnicity. Groups are selected based on theoretical considerations about whether or not the construct measured is hypothesised to have the same conceptual meaning across groups. We proposed that the PU-QOL instrument's scales should measure the same

constructs - here measured HRQL specific to PUs - across administration mode groups.

DIF involves a between group analysis, indicating any patterns of responses. Using RUMM2030 [31], we examined: uniform DIF - indicated by the same amount of DIF between groups measured, regardless of person ability/disability level - and non-uniform DIF - indicated by varying magnitudes of DIF according to ability/disability level. DIF was considered at both the 1% and 5% level. Bonferroni corrections were applied to both levels to take account of multiple testing [32]. This is a method for adjusting the significance levels of individual tests when multiple tests are performed on the same data (the test-wise significance levels are divided by the number of tests) [33,34]. An exact probability value using Bonferroni adjustment is calculated in RUMM2030.

## Results

### Qualitative analysis

Qualitative findings indicated problems with PU-QOL self-completion. Despite assessed as able to self-complete, almost half the sample (43%) required assistance with completion; eight were aged  $\geq 70$  and seven  $< 70$  years (see Gorecki et al 2013 for additional results from the qualitative study [20]). Reasons for needing assistance included: i) too weak/ill; ii) unable to hold a pen; iii) visually impaired (e.g. glasses not accessible); and iv) co-morbidity (e.g. acute or chronic illness). Respondents did not read instructions, expressed difficulty selecting an appropriate response option, or left items blank rather than indicating “no bother” if: i) they had not experienced what the item referred to; ii) they experienced it but not because of PUs; or iii) it applied only in the past. These issues did not emerge when PU-QOLs were administered.

### Quantitative analysis

We screened 427 patients from 21 hospitals, 10 community services and one hospice. Eligibility was assessed for 227 (53.2%), of which 142 were eligible (62.6%); 75 (52.8%) consented to participation. Cognitive impairment and inability to self-complete were the main reason for ineligibility (47.7% and 26% respectively). Patient characteristics are presented in Table 1.

### Response rates and data quality

Of the 75 patients recruited, 70 completed and returned PU-QOLs indicating a 93% response rate; no difference in response rate was observed by mode group. Table 2 indicates the percentage of missing data by groups: mode (self-complete and administered), age ( $< 70$  years and  $\geq 70$  years) and healthcare setting (hospital and community). For the administered group, the possible range of missed items was 0-1827 (i.e. 87 items per PU-QOL  $\times$  21 administrations = 1827 total items); a total of three

**Table 1 Patient characteristics**

	Self-completed (n = 49)	Administered (n = 21)	Total (n = 70)
Patient age (years)			
Mean (SD)	65 (15)	62 (16)	64 (15)
Median (range)	68 (21-85)	65 (27-93)	66 (21-93)
Under 70 years of age	25	14	39
70 years or older	24	7	31
Missing	0	0	0
Gender			
Male	33 (67.3%)	14 (66.7%)	47 (67.1%)
Female	16 (32.7%)	7 (33.3%)	23 (32.9%)
Type of healthcare setting			
Acute hospital	26 (53.0%)	12 (57.1%)	38 (54.3%)
Community	23 (47.0%)	9 (42.9%)	32 (45.7%)
Pressure ulcer severity			
Superficial grades 1/2	28 (57.1%)	12 (57.1%)	40 (57.1%)
Severe grades 3/4	21 (42.9%)	9 (42.9%)	30 (42.9%)

PU-QOLs were returned with 29 items missed (1.6%). For the self-completed group, the possible range of missed items was 0-4263; 19 PU-QOLs were returned with 619 missed items (14.5%).

Of the participants under 70 years of age who self-completed, 48% returned PU-QOLs with items missed compared to 29% of those 70 years or older that self-completed (Table 2). Of the administered group, two PU-QOLs had three items missed from those under 70 years and one PU-QOL with 26 items missed from those 70 years or older; this patient requested early completion due to feeling unwell.

A larger proportion of self-completed PU-QOLs were returned with missing data from hospitalised patients

compared to those living in the community who self-completed (Table 2). Of administered PU-QOLs, two returned with 28 items missed from patients hospitalised compared to only one PU-QOL returned with one item missed from those living in the community (Table 2). A difference was observed by healthcare setting; hospitalised patients that self-completed returned PU-QOLs with the largest amount of missing data.

#### Qualitative observations

PU-QOLs were examined to investigate any patterns in missing responses. The following observations were noted. Of the 19 self-completed PU-QOLs with missing data, four respondents wrote 'n/a' next to items missed,

**Table 2 Data quality – missing data**

	Self-completed (n = 49)	Administered (n = 21)	Total* (n = 70)
PU-QOLs with missing data	<b>19</b> (38.8%)	<b>3</b> (14.3%)	<b>22</b> (31.4%)
Total number of PU-QOL items missed (range 1-87 items per PU-QOL)	619 (14.5%)	29 (1.6%)	648 (10.6%)
Age			
Number under 70 years	(n = 12/25) <sup>+</sup>	(n = 2/14) <sup>+</sup>	(n = 14/39) <sup>+</sup>
Number items missed	336 (15.5%)	3 (0.3%)	345 (10.2%)
Number 70 years or older	(n = 7/24) <sup>+</sup>	(n = 1/7) <sup>+</sup>	(n = 8/31) <sup>+</sup>
Number items missed	283 (13.6%)	26 (4.3%)	309 (11.5%)
Type of healthcare setting			
Number acute	(n = 16/26) <sup>+</sup>	(n = 2/12) <sup>+</sup>	(n = 18/38) <sup>+</sup>
Number items missed	604 (26.7%)	28 (2.7%)	632 (19.1%)
Number community	(n = 3/23) <sup>+</sup>	(n = 1/9) <sup>+</sup>	(n = 4/32) <sup>+</sup>
Number items missed	15 (0.8%)	1 (0.1%)	16 (0.6%)

<sup>+</sup>Refers to the number of patients with missing data versus the total number of patients in the respective subgroup.

\*A total of 70 PU-QOLs were returned and analysed (5 completed PU-QOLs were lost in the post).

**Table 3 Summary of DIF by administration mode for each PU-QOL scale**

Scale (no. of items) (sample size)	Uniform		Non-uniform	
	p < 0.01*	p < 0.05*	p < 0.01*	*p < 0.05*
<b>Pain (11)</b>				
(n = 70)	0	0	0	1
(Adjusted n = 200)	1	4	1	1
(Adjusted n = 300)	4	6	1	4
<b>Exudate (8)</b>				
(n = 70)	0	0	0	0
(Adjusted n = 200)	2	3	1	2
(Adjusted n = 300)	2	4	1	2
<b>Odour (6)</b>				
(n = 70)	0	1	0	0
(Adjusted n = 200)	3	3	3	3
(Adjusted n = 300)	3	4	3	4
<b>Sleep (6)</b>				
(n = 70)	0	0	0	1
(Adjusted n = 200)	2	3	3	3
(Adjusted n = 300)	3	4	3	3
<b>Malaise (3)</b>				
(n = 70)	0	1	0	0
(Adjusted n = 200)	2	3	1	1
(Adjusted n = 300)	2	3	1	2
<b>Mobility (11)</b>				
(n = 70)	0	1	0	3
(Adjusted n = 200)	2	3	7	7
(Adjusted n = 300)	2	4	7	9
<b>Daily activities (9)</b>				
(n = 70)	0	1	0	1
(Adjusted n = 200)	2	3	2	5
(Adjusted n = 300)	3	4	4	6
<b>Mood (7)</b>				
(n = 70)	0	0	0	1
(Adjusted n = 200)	0	2	2	4
(Adjusted n = 300)	0	4	3	5
<b>Anxiety (3)</b>				
(n = 70)	0	0	0	0
(Adjusted n = 200)	0	0	0	0
(Adjusted n = 300)	0	0	0	0
<b>Self-consciousness (7)</b>				
(n = 70)	0	0	0	0
(Adjusted n = 200)	1	4	1	2
(Adjusted n = 300)	3	4	1	3

**Table 3 Summary of DIF by administration mode for each PU-QOL scale (Continued)**

<b>Autonomy (3)</b>				
(n = 70)	0	0	0	0
(Adjusted n = 200)	0	0	2	2
(Adjusted n = 300)	0	0	2	2
<b>Isolation (4)</b>				
(n = 70)	0	0	0	1
(Adjusted n = 200)	1	2	2	3
(Adjusted n = 300)	2	3	3	3
<b>Participation (9)</b>				
(n = 70)	0	0	0	2
(Adjusted n = 200)	4	8	7	7
(Adjusted n = 300)	6	8	7	7

\*Indicates the number of items with DIF at the specified p value.

suggesting that the response option 'My PU did not give me this problem' was not used as intended. Six respondents completed only one item per scale; five missed items at random; two missed a page; one missed items from only the daily activities scale; and one mostly missed items at the beginning of the instrument. For the three administered PU-QOLs with missing data, one had one item missed; one had two items missed; and one hospital patient missed 26 items due to feeling unwell. No obvious patterns in responses emerged.

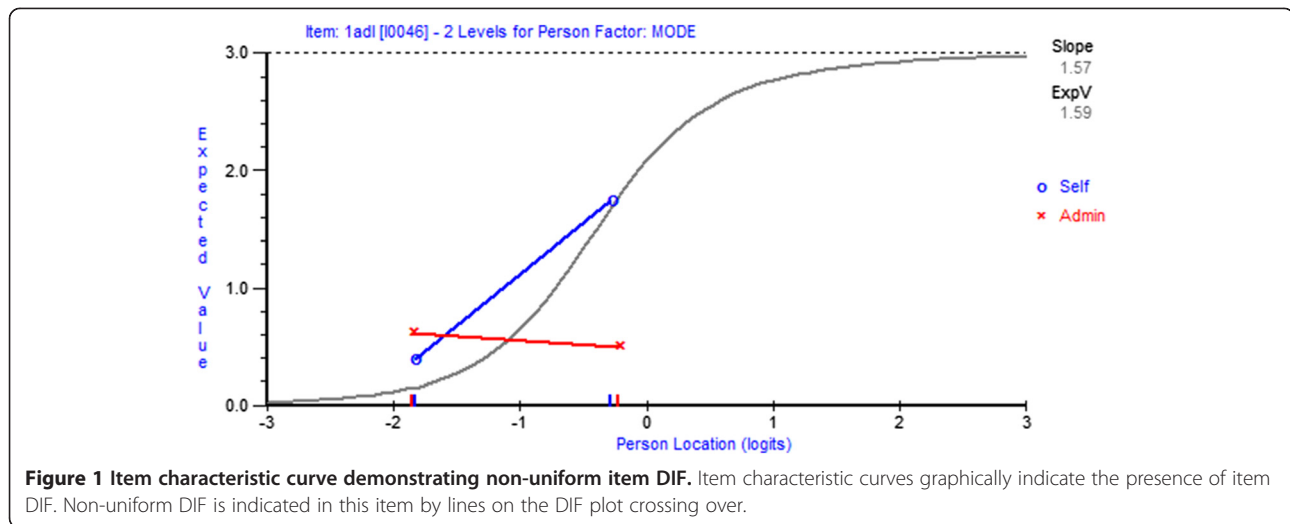
### Differential Item Functioning

Statistically there were no items with significant DIF by mode at the 1% confidence level (Table 3); thus supporting the equivalence of self-completed and interview-administered versions. A few items emerged with DIF at the 5% confidence level; however, the DIF observed was marginal (DIF was demonstrated in 9/13 scales but only ≤3 items for seven scales; Table 3). Figures 1 and 2 provide a graphical illustration of an item with and without DIF, respectively.

Additional exploration of DIF was undertaken with two hypothetical samples (n = 200 and n = 300); RUMM software has a function enabling multiplication of the original analysis sample (n = 70). In both adjusted samples, a significant proportion of items emerged with both uniform and non-uniform DIF (Table 3); highlighting areas warranting further investigation if pursuing a self-completed version in the future. Increasing the sample from 200 to 300 did not improve the detection of items with DIF (Table 3).

### Discussion

The PU-QOL instrument provided a vehicle for demonstrating a novel mixed methods approach to guide selection of the 'best' administration mode. Our findings



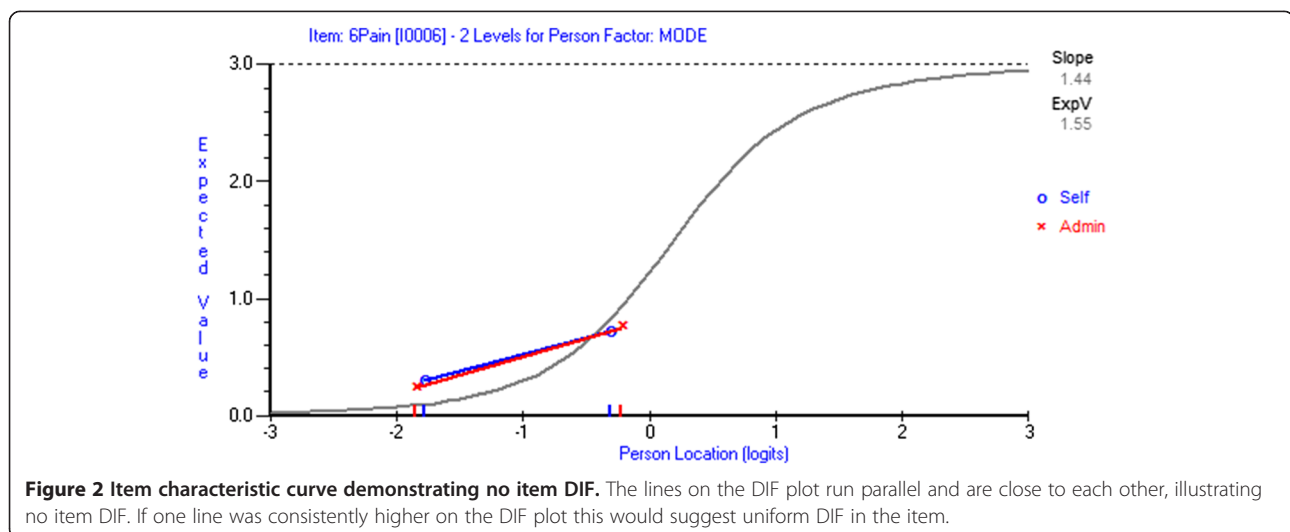
confirm the usefulness of our strategic approach for investigating response rate, data quality and measurement equivalence between two administration methods during early PRO instrument development or in limited samples where recruitment to large field tests is often difficult.

Qualitative data informed modifications to the PU-QOL instrument. Despite modifications intended to promote self-completion, almost half the sample required assistance with completion, of which half were aged 70 years or older; findings consistent with others [35,36]. Elderly patients were more likely to miss multiple items and expressed a preference for assistance with completion. The interpersonal interaction (interviewer can provide clarification); enabling those with reading or writing difficulties to be included in research; and enhancing data quality through facilitation with visual aids or checking for data completeness makes administration of PRO instruments a suitable

method for people with PUs and potentially other elderly or chronically ill populations.

A difference in data quality was observed; a large proportion of PU-QOLs that were self-completed by acute hospital patients had missing data; indicating the method was inappropriate for these patients. No difference in data quality was observed by mode for the community setting group, thus a self-completed version may be feasible for community patients; but the sample size was relatively small. Initially we had planned to include around 100 participants into this exploratory methodological study, however due to time constraints and objectives for the larger study [24], we only recruited 75 patients.

The DIF observed was marginal thus providing preliminary evidence of stable item performance across administration methods; suggesting PU-QOL scales could be measured on a common metric. However, when



investigating DIF in small samples, failure to detect no DIF at the 1% confidence level does not imply that no problems exist, rather that we might not have enough power to detect measurement issues. Using the 95% confidence level indicated that the few items with DIF did not warrant two administration mode-specific versions. However, items to be cognisant of if pursuing a self-completed version in the future were identified.

Determining DIF is valuable as detection of any severely problematic items (those presenting with significant DIF) would be expected even in small samples. However, as DIF is a product of the sample and not the scale (e.g. probabilities are sample size dependent), additional exploration of DIF was undertaken. To provide confidence in our findings of marginal DIF by administration mode, we inflated the sample size to provide a better feel for the behaviour of the data and increase the likelihood of revealing any DIF [37]. Despite encouraging preliminary results, re-examination in inflated samples detected measurement non-equivalence between administration methods on some scale items. Increasing the sample from 200 to 300 did not improve detection of items with DIF, suggesting that a sample of around 200 might be required for revealing significant DIF; however optimum sample size needs to be empirically determined.

The appropriateness of different administration methods will vary depending on the population being measured, the topic and content of the scale, and the setting of the data collection. This will differ from population to population, and scale to scale, and should be empirically tested. Based on our findings, we selected interview-administered mode to ensure suitability of the PU-QOL instrument across the wide spectrum of patients with PUs and to increase clinical meaningfulness; a self-completed PU-QOL would limit the type of people that could be assessed. In longitudinal research, this can be problematic as the progress of PUs and the impact on patients may not be accurately measured due to high levels of missing responses on repeated measurement. Finally, we provide preliminary evidence for the feasibility of a community self-completed version but as this study was not powered accordingly (e.g. once the  $n = 33$  community patients are split over the class interval groups used in the DIF analysis, a very small sample will be included in each class interval group), more work is needed to confirm appropriateness.

## Conclusion

Obtaining the best possible health outcomes data requires use of appropriate methods to ensure high quality data with minimal bias. Mixed methods, with the inclusion of RMT, provided both qualitative and empirical evidence for selection of the 'best' administration method for people with PUs. RMT/DIF analyses thus provide a complementary method alongside standard testing for examining key

clinically reasonable variables, with the intention of flagging issues with DIF for further examination. Parallel use of qualitative methods may assist in: explaining reasons for DIF; resolving them (i.e. adapt/improve items); and testing any changes made to instruments early in the development process.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

CG contributed to study concept and design, acquisition of qualitative data, analysis and interpretation of data, and preparation of manuscript. JN, JB and DL contributed to study design, interpretation of data and preparation of manuscript. SC contributed to study concept and design, analysis and interpretation of data, and preparation of manuscript.

## Acknowledgments

This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its programme Grants for Applied research funding scheme (RP-PG-0407-10056). The views expressed in this paper are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We would like to thank all research teams at participating centres around the United Kingdom and the participants for taking time to be involved in this research. The sponsor had no role in the design, methods, subject recruitment, data collections, analysis and preparation of paper.

## Funding

This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its programme Grants for Applied research funding scheme (RP-PG-0407-10056). The views expressed in this paper are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

## Author details

<sup>1</sup>Clinical Trials Research Unit (CTRU), University of Leeds, Leeds, UK. <sup>2</sup>Department of Health Services Research and Teaching, London School of Hygiene & Tropical Medicine, London, UK. <sup>3</sup>Clinical Neurology Research Group, Plymouth University Peninsula School of Medicine and Dentistry, Plymouth, UK. <sup>4</sup>QOL Office, University of Sydney, Sydney 2006, Australia.

Received: 10 December 2012 Accepted: 6 February 2014

Published: 12 February 2014

## References

1. Greenfield S, Nelson E: Recent developments and future issues in the use of health status assessment measures in clinical settings. *Med Care* 1992, **20**:S23–41.
2. Hobart J, Cano S, Zajicek J, Thompson A: Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007, **6**:1094–95.
3. US Department of Health & Human Services FDA: Patient reported outcome measures: use in medical product development to support labelling claims. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm193282.pdf>. 2009. MD, US Department of Health & Human Support Food & Drug Administration.
4. Scientific Advisory Committee of the Medical Outcomes Trust: Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002, **11**:193–205.
5. McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, et al: Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess* 2001, **5**:1–256.
6. Puhon M, Ahuja A, Van Natta M, Ackatz L, Meinert C: Interviewer versus self-administered health-related quality of life questionnaires-Does it matter? *Health Qual Life Outcomes* 2011. doi:10.1186/1477-7525-9-30.
7. Bowling A: Mode of questionnaire administration can have serious effects on data quality. *J Public Health* 2005, **27**:281–291.

8. Newton R, Prensky D, Schuessler K: **Form effect in the measurement of feeling states.** *Soc Sci Res* 1982, **11**:301–17.
9. Haywood KL, Garratt AM, Fitzpatrick R: **Quality of life in older people: a structured review of generic self-assessed health instruments.** *Qual Life Res* 2005, **14**:1651–1668.
10. McHorney CA: **Measuring and monitoring general health status in elderly persons: Practical and methodological issues in using the SF-36 health survey.** *Gerontologist* 1996, **36**:571–583.
11. McHorney C, Kosinski M, Ware J: **Comparisons of the cost and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey.** *Med Care* 1994, **32**:551–67.
12. Sikorski A, Given C, Given B, Jeon A, You M: **Differential symptom reporting by mode of administration of the assessment: automated voice response system versus a live telephone interview.** *Med Care* 2009, **47**:866–874.
13. Weinberger M, Oddone E, Samsa G, Landsman P: **Are health-related quality of life measures affected by the mode of administration?** *J Clin Epidemiol* 1996, **49**:135–140.
14. Rasch G: *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago; 1960.
15. Linacre J: **Sample size and item calibration stability.** *Rasch Meas Trans* 1994, **7**:328.
16. Coleman S, Gorecki C, Nelson E, Closs J, Defloor T, Halfens R, et al: **Patient risk factors for pressure ulcer development: systematic review.** *Int J Nurs Stud* 2013. <http://dx.doi.org/10.1016/j.ijnurstu.2012.11.019>.
17. NICE: *CG7 Pressure ulcer prevention.* web . 2003. from the NICE, UK website.
18. Gorecki C, Brown J, Nelson E, Briggs M, Schoonhoven L, Dealey C, et al: **Impact of pressure ulcers on quality of life in older patients: a systematic review.** *J Am Geriatr Soc* 2009, **57**:1175–1183.
19. Gorecki C, Nixon J, Lamping DL, Alavi Y, Brown JM: **Patient-reported outcome measures for chronic wounds with particular reference to pressure ulcer research: A systematic review.** *Int J Nurs Stud* 2013, **51**:157–65.
20. Gorecki C, Lamping D, Nixon J, Brown J, Cano S: **Applying mixed methods to pretest the Pressure Ulcer Quality of Life (PU-QOL) instrument.** *Qual Life Res* 2012, **21**:441–451.
21. Gorecki C, Lamping DL, Brown JM, Madill A, Firth J, Nixon J: **Development of a conceptual framework of health-related quality of life in pressure ulcers: a patient-focused approach.** *Int J Nurs Stud* 2010, **47**:1525–1534.
22. Andrich D: *Rasch models for measurement.* Beverly Hills: Sage Publications; 1988.
23. Hagquist C, Andrich D: **Is the Sense of Coherence-instrument applicable on adolescents? A latent trait analysis using Rasch modelling.** *Pers Individ Differ* 2004, **36**:955–968.
24. Gorecki C, Brown J, Cano S, Lamping D, Briggs M, Coleman S, et al: **Development and validation of a new patient-reported outcome measure for patients with pressure ulcers: the PU-QOL instrument.** *Health Qual Life Outcomes* 2013, **11**:95. <http://www.hqlo.com/content/11/1/95>.
25. Vileikyte L, Peyrot M, Bundy C, Rubin RR, Leventhal H, Mora P, et al: **The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument.** *Diabetes Care* 2003, **26**(9):2549–2555. (26 ref) 2003, 2549-2555.
26. Andrich D: **Distinctions between assumptions and requirements in measurements in the social sciences.** In *Mathematical and Theoretical Systems: Vol 4.* Edited by Keats JA, Taft R, Heath RA, Lovibond SH. North Holland: Elsevier Science Publishers; 1989:7–16.
27. Cano S, Barrett L, Zajicek J, Hobart J: **Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis.** *Mult Scler J* 2011, **17**:214–222.
28. Andrich D: **An elaboration of Guttman scaling with Rasch models for measurement.** In *Sociological Methodology.* Edited by Brandon-Tuma N. San Francisco: Jossey-Bass; 1985:33–80.
29. Zumbo BD: *A handbook on the theory and methods of Differential Item Functioning (DIF): logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* 1999. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
30. Teresi J, Ramirez M, Lai J, Silver S: **Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: description of DIF methods, and review of measures of depression, quality of life and general health.** *Psychol Sci Q* 2008, **50**:538.
31. Andrich D, Sheridan B, Luo G: *RUMM 2030. 4.0 for windows (upgrade 4600.0109).* Perth, WA: RUMM Laboratory; 2010. PTY LTD.
32. Bland J, Altman D: **Multiple significance tests: the Bonferroni method.** *BMJ* 1995, **310**:170.
33. Bonferroni C: *Teoria statistica delle classi e calcolo delle probabilita.* Pubblicazionidel R Istituto Superiore di Scienze Economichee Commerciali di Firenze. 2011, **8**:3–62.
34. Miller R: *Simultaneous statistical inference.* 2nd edition. Verlag: Springer; 1981.
35. Fletcher A, Bulpitt C: **Quality of life and hypertensive drugs in the elderly.** *Ageing Clin Exp Res* 1992, **4**:115–123.
36. Hayes V, Morris J, Wolfe C, Morgan M: **The SF-36 health survey questionnaire: is it suitable for use with older adults?** *Age Ageing* 1995, **24**:120–125.
37. Hobart J, Cano S: **Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.** *Health Techno Assess* 2009, **13**:1–177.

doi:10.1186/1471-2288-14-22

**Cite this article as:** Rutherford et al.: Using mixed methods to select optimal mode of administration for a patient-reported outcome instrument for people with pressure ulcers. *BMC Medical Research Methodology* 2014 **14**:22.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

