

Rasch Analysis of the Dermatology Life Quality Index Reveals Limited Application to Chinese Patients with Skin Disease

Zehui HE^{1#}, Riccardo LO MARTIRE^{2,3#}, Chuanjian LU^{4,5}, Hongxia LIU⁶, Lin MA⁷, Ying HUANG⁸, Yongmei LI⁹, Liyun SUN¹⁰, Yanping BAI¹¹, Wali LIU¹² and Xushan ZHA¹³

¹Department of Clinical Epidemiology, ⁴Department of Dermatology, Guangzhou University of Chinese Medicine Second Affiliated Hospital, Guangzhou, China, ²Department of Aeronautical and Vehicle Engineering, KTH Royal Institute of Technology, ³Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden, ⁵Key Laboratory of Clinical Research on Traditional Chinese Medicine Syndrome, Guangdong Provincial Academy of Chinese Medical Sciences, Guangzhou, ⁶Departments of Dermatology, ⁷Xinjiang Medical University Affiliated Chinese Medicine Hospital, Urumqi, ⁸Heilongjiang Provincial Academy of Chinese Medical Sciences, Harbin, ⁹Chengdu University of Traditional Chinese Medicine Affiliated Hospital, Chengdu, ¹⁰Shanghai University of Traditional Chinese Medicine Affiliated Longhua Hospital, Shanghai, ¹¹Capital Medical University Affiliated Beijing Traditional Chinese Medicine Hospital, Beijing, ¹²China-Japan Friendship Hospital, ¹³China Academy of Chinese Medical Sciences Guanganmen Hospital, Beijing, and ¹³Guangzhou University of Chinese Medicine First Affiliated Hospital, Guangzhou, China

[#]These authors contributed equally to this work.

The objective of this study was to examine the psychometric properties of the Chinese version of the Dermatology Life Quality Index (DLQI) and to assess the invariance of its items with respect to several patient parameters via Rasch analysis. Data were aggregated from 9,845 patients with various skin diseases across 9 hospitals in different regions of China. The response structure, local independence, and reliability of the DLQI scale were analysed in a partial credit model, and differential item functioning (DIF) across region, disease, sex, and age were assessed with a Mantel-Haenszel procedure. Although acceptable scale reliability (Person Separation Index=2.3) was obtained, several problems were revealed, including disordered response thresholds, misfitting items, DIF by geographical region and disease, and mis-targeting patients with mild impairment regarding health-related quality of life (HRQL). In conclusion, the DLQI provides inadequate information on patients' impairments in HRQL, and the application of the DLQI in Chinese patients with skin disease is limited.

Key words: Dermatology Life Quality Index; skin disease; Chinese; Rasch analysis; differential item functioning.

Accepted Jul 5, 2017; Epub ahead of print Jul 5, 2017

Acta Derm Venereol 2018; 98: 59–64.

Corr: Chuanjian Lu, Department of Dermatology, Guangzhou University of Chinese Medicine Second Affiliated Hospital, 111 Da De Road, Guangzhou 510120, China. E-mail: luchuanjian888@vip.sina.com

The Dermatology Life Quality Index (DLQI) (1) has been translated into more than 90 languages and applied to over 40 different skin conditions (2). It is the most commonly used health-related quality of life (HRQL) instrument in dermatology worldwide (3, 4). The psychometric properties of the DLQI have been a controversial issue, due to contradictory results of studies using either classical or modern test theory approaches. Although acceptable psychometric properties have been reported for various DLQI translations when assessed via classical test theory approaches (5–10), investigations

based on Rasch analysis have identified several problems with the scale, including the Chinese version (11–13).

Since the translation of the DLQI into Chinese in 2004 (10), 3 peer-reviewed studies focusing on its psychometric properties have been published: 2 were classical theory-based (5, 10), and one was Rasch-based with a relatively small sample size of 150 patients with neurodermatitis (13). The psychometric properties of the DLQI have not been evaluated adequately in a large sample of patients with skin disease, nor have its item response functions for 2 or more subgroups of skin diseases. Therefore, this study examined the response category structure, fitness of items and persons, and local independence of items of the Chinese version of the DLQI via Rasch analysis, and assessed the invariance of items with respect to several patient subgroups in 9,845 Chinese dermatology patients.

METHODS

Design and participants

In this cross-sectional study, 9,845 dermatology patients were consecutively recruited in 9 hospitals from different geographical regions of mainland China between 2013 and 2015. Inclusion criteria were: minimum age 16 years, diagnosed skin disease, and ability to understand and read Chinese. Exclusion criteria were: mental or physical incapacity resulting in inability to complete the survey. This study was approved by the ethics committee of the Guangzhou University of Chinese Medicine Second Affiliated Hospital and conformed to the principles of the Declaration of Helsinki.

Initially, patients received information about the study and signed informed consent forms. Then they provided their demographic information and self-completed the DLQI. Dermatologists confirmed the skin disease diagnoses and assessed their severity on a 5-point Likert-type response from “very mild” to “very severe”.

Dermatology Life Quality Index

The DLQI is a self-administered questionnaire used to assess the impact of skin disease on HRQL. It contains 10 items covering 6 aspects of quality of life: symptoms and feelings, daily activities, leisure, work and school, personal relationships and problems

with treatment. Nine items are rated on a 4-point Likert-type scale, with scores 3 (“very much”), 2 (“a lot”), 1 (“a little”) and 0 (“not at all”). Item 7 is divided into 2 steps: the first inquiring whether work or school have been prevented: A “yes” is scored as 3; if “no” is selected, the patient specifies to what degree the skin condition has been a problem at work or school, scored as 2 (“a lot”), 1 (“a little”) or 0 (“not at all”). For 8 of the 10 items, a “not relevant” option is also available, which is scored as 0. Individual item scores are summed to a total score of 0–30, with higher scores corresponding to a larger impact on HRQL.

Rasch analysis

The psychometric properties of the DLQI were assessed in a polytomous Rasch model (Winsteps® Rasch measurement program v3.92.1, John M. Linacre, Oregon, USA) conforming to prior recommendations (14). All DLQI items were analysed together first, and scale optimization was then attempted.

To determine whether a partial credit model (PCM) (15) or a rating scale model (RSM) (16) was the most suitable, the likelihood ratio test, Akaike’s information criterion (AIC) and Schwartz’s Bayesian information criterion (BIC) were used. A significant likelihood ratio test and smaller AIC or BIC values suggested that the PCM provided a better model fit.

The structure of the response categories of the DLQI was assessed via the response distribution, categorical measure advancement, and goodness-of-fit (17). According to recommendations, a minimum of 10 observations per response category is necessary to avoid imprecise and unstable model estimates. Mean measures also must advance logically with their respective categories, and response categories must have an acceptable model fit (17). Following response category assessment, the fit of individual items and persons was examined. To evaluate model fit, unstandardized mean square values (MNSQ) with a χ^2 distribution, or standardized MNSQ with a *t*-distribution are commonly used (18). In this study, infit and outfit MNSQ in the range 0.6–1.4 were considered an acceptable model fit, and lower and higher values suggested overfit (redundancy) and underfit (unpredictability), respectively (19). Standardized MNSQ were not used in the model fit evaluation because of their sensitivity to sample size (18).

To identify invariance failures of DLQI items across subgroups, differential item functioning (DIF) was investigated between sexes (females vs. males), age groups (16–30 vs. 31–50 vs. 51–91 years), hospital’s geographical location (north vs. south vs. east vs. west China), disease (acne vs. eczema vs. psoriasis), and diagnosed disease severity (“very mild”–“mild” vs. “moderate” vs. “severe”–“very severe”). Mean item measures were initially compared between groups, with differences of 0.5 logits or more considered meaningful (20), and further analysed with a Mantel-Haenszel procedure to ascertain their statistical significance (21). To avoid biases related to group size differences in the analyses, simple randomization was used to select subsamples corresponding to the smallest group size. Lastly, the uniformity of statistically significant DIF was assessed. Item characteristic curves (ICC) were first visually inspected, with differences consistent and non-consistent over the measure range defined as uniform and non-uniform, respectively (14). Ordinal logistic regression (MASS, v7.3-45, Ripley et al. 2016, <https://CRAN.R-project.org/package=MASS>) was then used to statistically evaluate DIF uniformity according to a previously described procedure (22). Alpha was set at 0.05 and Bonferroni-adjusted for all analyses to diminish the risk of alpha inflation due to multiple comparisons.

The local independence of items was examined by the dimensionality and the response dependency of DLQI (14). Dimensionality was assessed via principal component analysis of the residuals (PCAR), with eigenvalues of residual components of less than

2.0 considered as supporting unidimensionality (23). Response dependency was evaluated via the correlation between the items’ standardized residuals (14), with correlation coefficients of more than 0.3 considered unacceptably high.

The HRQL impairment requirement of the DLQI response categories relative to patients’ HRQL impairment was assessed via a Wright map (24). Finally, the internal reliability in distinguishing between persons according to disease severity was determined via the Person Separation Index (PSI), with 1.50 considered acceptable and 2.00 good. This means it can discern between 2 and 3 satisfaction levels, respectively (24). All the properties evaluated in this study were listed in Table S1[†] to facilitate interpretation of Rasch analysis.

RESULTS

Sample characteristics

Of the 9,845 dermatological patients participating in this study, 63% were female, the mean age was 33 years, and the 4 most common diseases were acne, eczema, dermatitis, and psoriasis. **Table I** presents the sample’s demographic characteristics in more detail.

Rasch analysis

Rasch model selection. The likelihood ratio test ($\chi^2_{3244} = 2937$; $p < 0.0001$), AIC ($\Delta 9425$), and BIC ($\Delta 32524$) all supported a PCM over a RSM, and therefore a PCM was chosen.

[†]<https://doi.org/10.2340/00015555-2742>

Table I. Sample characteristics

Characteristics	
Sex (<i>n</i> , %)	
Female	6,214 (63.1)
Male	3,564 (36.2)
Missing	67 (0.7)
Age, years, mean (SD; min–max)	33.0 (13.5; 16–91)
Hospital’s geographical location (<i>n</i> , %)	
North China	3,154 (32.0)
South China	2,561 (26.0)
West China	3,099 (31.5)
East China	1,031 (10.5)
Marital status (<i>n</i> , %)	
Married	4,989 (50.7)
Single	4,540 (46.1)
Missing	316 (3.2)
Disease (<i>n</i> , %)	
Acne	2,947 (29.9)
Eczema	1,275 (13.0)
Dermatitis ^a	1,218 (12.4)
Psoriasis	1,335 (13.6)
Urticaria	721 (7.3)
Other	2,349 (23.9)
Diagnosed disease severity (<i>n</i> , %)	
Very mild	321 (3.3)
Mild	2,995 (30.4)
Moderate	4,761 (48.4)
Severe	1,220 (12.4)
Very severe	118 (1.2)
Missing	430 (4.4)
DLQI score, median (IQR; min–max)	7 (3, 12; 0–30)

^aDermatitis includes atopic, contact-, neuro-, non-specific-, and seborrhoeic dermatitis. Discrepancies in percentiles are due to missing values. SD: standard deviation; IQR: interquartile range.

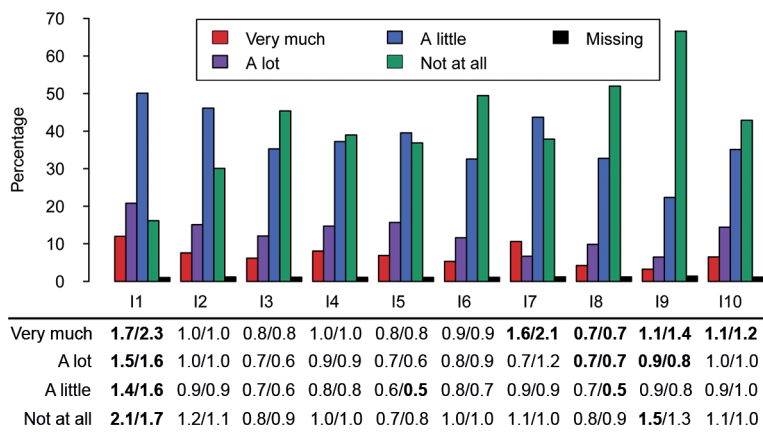


Fig. 1. Dermatology Life Quality Index (DLQI) raw score item frequencies. The category "Not at all" includes item responses from the "not relevant" category. The attached table displays infit and outfit mean square values by category across items, with **bold** numbers denoting misfitted values. *n*=9,845.

Structure of response categories

Fig. 1 shows the response category distribution of 9,845 subjects across the DLQI items. Maximum and minimum scores were observed in 37 (0.4%) and 511 subjects (5.2%), respectively, 85 subjects (<0.1%) had completely missing data, and no response category had less than 315 observations. Excluding the 2 highest categories of item 7, which were disordered (0.58 followed by 0.76 logits), all item categories advanced logically, with mean (SD) categorical step differences in mean measures of 1.42 (0.13), 1.34 (0.20), and 2.18 (0.17) logits between the 4th (highest) and 3rd, 3rd and 2nd, and 2nd and 1st (lowest), respectively. Moreover, items 1, 7, and 9 had underfitted categories, and items 5 and 8 had overfitted categories (Fig. 1).

Person and item fit

In total, 771 persons (7.8%) underfitted and 715 persons (7.3%) overfitted the model. Group comparison of overfitted, fitted, and underfitted patients showed a similar distribution in age, sex, diagnosed disease severity, and overall diseases (with the exception of acne). Persons with acne and who were in western China tended to overfit the model more frequently than other persons. The DLQI item characteristics are detailed in **Table II**.

II. As displayed, item 1 underfitted and no other items overfitted the model.

Differential item functioning

In total, DIF was observed in 4 of 10 items, and was associated with the hospital's geographical location for item 7 and with the disease for items 1, 2 and 5 (Table II). The visual inspection suggested that all DIF except for that of item 5 related to disease (**Fig. 2**) were non-uniform. On the other hand, the ordinal logistic regression classified all DIF as uniform. No DIF was observed for sex, age, or diagnosed disease severity.

Local independence

The PCAR identified no substantive residual latent dimensions (eigenvalue ≤ 1.5), thereby supporting unidimensionality. In contrast, item 1 had a considerable residual correlation with item 5 (*r*=-0.31; *p*<0.001), suggesting response co-dependency.

Logits coverage and scale targeting

Mean DLQI item measures spanned -1.26 to 1.11 logits (Table II), with the item total range covering 4.92 to 7.58 logits. **Fig. 3** shows a Wright map of the patients'

Table II. Dermatology Life Quality Index (DLQI) item fit statistics and differential item functioning

Item description	Raw score	Latent measure	MNSQ _{infit} (ZSTD)	MNSQ _{outfit} (ZSTD)	Differential item functioning
(1) Itchy, sore, painful, or stinging	1.29	-1.26	1.63 (9.9)	1.80 (9.9)	Ac>Ec 1.34 ^{b,*} ; Ps>Ec 1.19 ^{b,*}
(2) Embarrassment/self-consciousness	1.00	-0.38	1.01 (0.5)	1.00 (0.1)	Ec>Ac 0.95 ^{b,*} ; Ps>Ac 0.63 ^{b,*}
(3) Interferes with shopping/looking after home/garden	0.79	0.17	0.76 (-9.9)	0.72 (-9.9)	
(4) Influences choice of clothes	0.92	-0.18	0.93 (-4.8)	0.91 (-5.0)	
(5) Affects social/leisure activities	0.92	-0.15	0.70 (-9.9)	0.66 (-9.9)	Ec>Ac 0.54 ^{a,*}
(6) Affects ability to do sports	0.72	0.36	0.91 (-5.9)	0.86 (-6.5)	
(7) Prevents working/studying	0.90	-0.27	1.17 (9.8)	1.21 (9.9)	E>W 0.58 ^{b,*}
(8) Creates problems with partner/close friends/relatives	0.66	0.58	0.74 (-9.9)	0.67 (-9.9)	
(9) Causes sexual difficulties	0.45	1.11	1.11 (5.7)	0.93 (-1.9)	
(10) Problem with treatment	0.84	0.05	1.01 (0.8)	1.04 (2.2)	

Raw score and latent measure show mean values, and bold numbers denote misfitted values. Differential item functioning (DIF) associated with disease (*n*=3,744) or hospital's geographical location (*n*=4,068), expressed in logits and significant at *p*<0.0021 (i.e. alpha level at 0.05 Bonferroni-adjusted for 24 analyses). ^aUniform DIF classified based on visual inspection. ^bNon-uniform DIF classified based on visual inspection. *Uniform DIF classified based on ordinal logistic regression. MNSQ: mean square; ZSTD: z-score; Ac: acne; Ec: eczema; Ps: psoriasis; E: eastern China; W: western China.

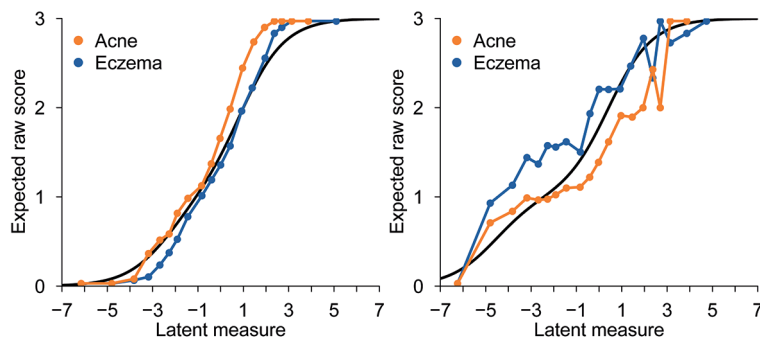
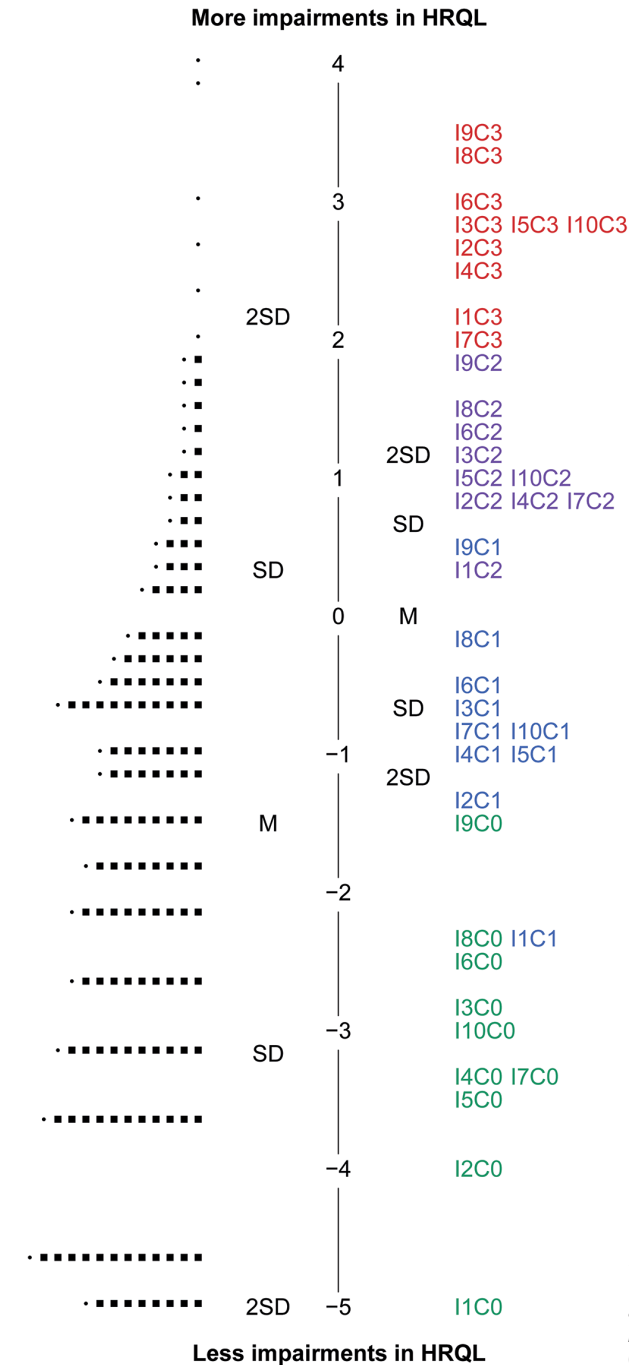


Fig. 2. Example of differential item functioning (DIF) for acne compared with eczema visualized on item characteristic curves. Based on visual inspection, left and right side shows DIF classified as uniform (item 5) and non-uniform (item 1), respectively. $n=2,496$.



HRQL impairment relative to the HRQL impairment requirement of the DLQI item response categories. As illustrated, the mean person measure of -1.66 logits shows that DLQI mistargeted the sample, having too high of an impairment requirement for most patients. Indeed, for patients with an impairment below mean, only item 1 had sufficient sensitivity to identify improvement in HRQL.

Person Separation Index

The PSI of 2.30 ($r=0.84$) suggested that DLQI had adequate internal reliability and was able to distinguish between the 3 sample subgroups in HRQL impairments.

Scale optimization

In order to improve the psychometric properties of the DLQI, data were reanalysed following stepwise modifications (**Table III**). First, patients who misfitted the model were excluded, which resolved the disordering in item 7 response categories. However, all item 1 categories, the 2 highest item 7 categories, and the 2 lowest item 9 categories still underfitted the model. Next, item 1 categories were dichotomized, and the aforementioned misfitted categories in items 7 and 9 were each combined. While resolving the categorical misfits of items 7 and 9, item 1 categories still underfitted the model. Subsequently, item 1 was excluded, resulting in item 9 overfitting the model; however, following the exclusion of item 9, all items and item categories fitted the model. A considerable DIF associated with disease and hospital's geographical location remained for items 2 and 7, respectively, and these were therefore split in the final analysis. In the resulting model, no items or item categories misfitted the model, and no residual latent dimensions or correlations were observed. However, a total of 1,035 (10.7%) patients underfitted and 819 (8.5%)

Fig. 3. Wright map. The sample's impairment in health-related quality of life (HRQL) due to skin disease, relative to the HRQL impairment requirement of the Dermatology Life Quality Index (DLQI) item response categories measured on a logit scale. Symbols to the left denote patients (square: $n=62$ patients; dot: $n=1-61$) and characters to the right mark DLQI item (I) response categories (C: 0="not at all"/"not relevant", 1="A little", 2="A lot", 3="very much"). $n=9,690$.

Table III. Dermatology Life Quality Index scale modifications

Ana-lysis No.	Description	<i>n</i>	PSI (<i>r</i>)	Item misfit <i>n</i> (%)	Person misfit <i>n</i> (%)
1	All of sample	9,690	2.30 (0.84)	1 (10.0)	1,486 (15.3)
2	Misfitted persons excluded	8,204	2.40 (0.85)	1 (10.0)	198 (2.4)
3	Collapse response categories for I1: 0&1, 2&3; I7: 2&3; I9: 0&1	8,204	2.23 (0.83)	1 (10.0)	432 (5.3)
4	Delete I1	8,204	2.26 (0.84)	1 (11.1)	374 (4.6)
5	Delete I9	8,204	2.25 (0.83)	0 (0)	359 (4.4)
6	Split I2 by disease and I7 by hospital's location	8,204	2.26 (0.84)	0 (0)	368 (4.5)

PSI: person separation index; *r*: person reliability; *n*: sample size; I: item.

patients overfitted the model, and the scale's mistargeting remained at similar magnitudes.

DISCUSSION

DLQI measurement properties

In this study, the psychometric properties of the Chinese version of DLQI were examined with Rasch analysis in a large sample of dermatological patients, and the item DIF relative to several patient characteristics was assessed. Our findings suggest that the scale has several structural problems and that the score of DLQI therefore does not provide accurate information on patients' HRQL impairments. They also show that the comparability of item scores between patient characteristic groups is limited.

In agreement with previous Rasch studies (11–13, 25), our results suggested that although DLQI has good internal reliability in distinguishing between patients with respect to the degree of their HRQL impairments, there are several issues with the scale. Item 7 had logically disordered response categories; items 5, 7, 8, 9 contained several categories which misfitted the model; and item 1 misfitted the model both on a category and item level. This supports that the response patterns on those items were inconsistent with the model's predicted response pattern. Thus, in contrast to PCAR, which supported unidimensionality for the scale globally, the item misfits suggested that there are local violations against unidimensionality and that DLQI therefore does not conform to the unidimensionality assumption (18). These findings agree with those of previous Rasch studies, which also identified problems with the DLQI for several dermatological diseases (11–13, 25). Thus, the available research supports that the DLQI does not measure a single latent construct and that the summary score therefore does not provide accurate information on patients' impairments in HRQL.

To rule out the possibility that the observed item misfit was a consequence of item invariance violations across patients with different characteristics, separate Rasch analyses were conducted for each patient group for which DIF was identified. The results were similar to

those obtained for the complete patient sample. Disease-related DIF was observed for items 1, 2, and 5, and location-related DIF for item 7. Most available studies of DLQI have also reported DIF, albeit related to sex, age, disease and cultural differences (11, 12, 25). The combined results therefore suggest that comparability between patients with different characteristics is limited.

Consistent with previous studies (12, 13, 25), we found that DLQI considerably mistargeted the patients. Few items covered the lower part of the logit scale and large step differences between the lower response categories impeded the evaluation of improvements in patients with minor HRQL impairments (i.e. the second lowest response categories were grouped around the person measure mean). Thus, DLQI is better suited for patients with severe impairments in HRQL. To improve the measurement properties of the DLQI, we made several changes to the scale. Although this resolved the misfitting items, the scale's mistargeting remained, and nearly 20% of the patient sample did not fit the model.

Strengths and limitations

To our knowledge, this is one of the most comprehensive DLQI studies conducted to date. The heterogeneous sample of nearly 10,000 patients with several different dermatological diagnoses recruited from across China, provided sufficient power for robust item estimates with small standard errors, while also enabling comparisons between multiple sample subgroups. In addition, small amounts of missing data limited related biases to an insignificant level. However, some limitations must be considered when interpreting the results. The analyses were based on data with the response categories "Not at all" and "Not relevant" combined, which probably resulted in some bias, since "Not relevant" does not signify that patients did not have severe disease. Moreover, it is possible that patient heterogeneity inflated model noise, since different patient characteristic combinations could have engendered precarious response patterns. Finally, approximately 15% of the patient sample responded differently than expected by the model. This means that our results are limited to a subgroup of the dermatological patient population.

Conclusion

The DLQI is the most commonly used HRQL instrument for dermatology patients. Whereas its psychometric properties have been reported adequate when tested with classical test theory, analysis with modern test theory has revealed that the DLQI fails to fulfil the proper requirements. Our findings, based on Rasch analysis, verified several problems with its measurement properties and uncovered DIF for several items across patient characteristic groups. Thus, the application of the DLQI in Chinese patients with skin disease is limited, and these

factors need to be considered when interpreting results based on data from the DLQI in its current form.

ACKNOWLEDGEMENTS

This work was funded by the National Key Technology R&D Program for the 12th Five-year Plan of Ministry of Science and Technology, China [grant number 2013BAI02B03]; the Financial Industry Technology Research and Development Program of Guangdong Province, China [grant number 2011 (285)05]; and the Guangdong Science and Technology Project, China [grant number 2014A020221040, 2014B010118005].

The authors would like to thank Lea Constan for proofreading the manuscript. The authors also thank the collaborating investigators in each of the following hospitals: Dr Ruiqiang Fan, Guangzhou University of Chinese Medicine Second Affiliated Hospital, Guangzhou; Dr Fei Guo, Xinjiang Medical University Affiliated Chinese Medicine Hospital, Urumqi; Dr Lianwei Kong and Dr Fanhui Yuan, Heilongjiang Provincial Academy of Chinese Medical Sciences, Harbin; Dr Chunxiao Li and Tianhao Li, Chengdu University of Traditional Chinese Medicine Affiliated Hospital, Chengdu; Dr Shangpu Gao, Shanghai University of Traditional Chinese Medicine Affiliated Longhua Hospital, Shanghai; Dr Guangzhong Zhang and Jie Su, Capital Medical University Affiliated Beijing Traditional Chinese Medicine Hospital, Beijing; Dr Wenqi Zhang and Haoyu Yang, China-Japan Friendship Hospital; Dr Feixing Zheng, China Academy of Chinese Medical Sciences Guanganmen Hospital, Beijing; and Dr Jingyang Yu, Guangzhou University of Chinese Medicine First Affiliated Hospital, Guangzhou, China.

The authors have no conflicts of interests to declare.

REFERENCES

1. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; 19: 210–216.
2. Department of Dermatology of Cardiff University. Quality of Life Questionnaires. [accessed 2016 Dec 5]. Available from: <http://www.cardiff.ac.uk/dermatology/quality-of-life/>.
3. Basra MKA, Fenech R, Gatt RM, Salek MS, Finlay AY. The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results. *Br J Dermatol* 2008; 159: 997–1035.
4. 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.
5. He Z, Lu C, Basra MKA, Ou A, Yan Y, Li L. Psychometric properties of the Chinese version of Dermatology Life Quality Index (DLQI) in 851 Chinese patients with psoriasis. *J Eur Acad Dermatol Venereol* 2013; 27: 109–115.
6. Mork C, Wahl A, Moum T. The Norwegian version of the dermatology life quality index: a study of validity and reliability in psoriatics. *Acta Derm Venereol* 2002; 82: 347–351.
7. Takahashi N, Suzukamo Y, Nakamura M, Miyachi Y, Green J, Ohya Y, et al. Japanese version of the Dermatology Life Quality Index: validity and reliability in patients with acne. *Health Qual Life Outcomes* 2006; 4: 46.
8. Ozturkcan S, Ermertcan AT, Eser E, Sahin MT. Cross validation of the Turkish version of dermatology life quality index. *Int J Dermatol* 2006; 45: 1300–1307.
9. Mazzotti E, Barbaranelli C, Picardi A, Abeni D, Pasquini P. Psychometric properties of the Dermatology Life Quality Index (DLQI) in 900 Italian patients with psoriasis. *Acta Derm Venereol* 2005; 85: 409–413.
10. Wang XL, Zhao TE, Zhang XQ. Assessment on the reliability and validity of the Dermatology Life Quality Index in Chinese version. *Chin J Epidemiol* 2004; 25: 791–793.
11. Nijsten T, Meads DM, de Korte J, Sampogna F, Gelfand JM, Ongenae K, et al. Cross-cultural inequivalence of dermatology-specific health-related quality of life instruments in psoriasis patients. *J Invest Dermatol* 2007; 127: 2315–2322.
12. Twiss J, Meads DM, Preston EP, Crawford SR, McKenna SP. Can we rely on the Dermatology Life Quality Index as a measure of the impact of psoriasis or atopic dermatitis? *J Invest Dermatol* 2012; 132: 76–84.
13. Liu Y, Li T, An J, Zeng W, Xiao S. Rasch analysis holds no brief for the use of the Dermatology Life Quality Index (DLQI) in Chinese neurodermatitis patients. *Health Qual Life Outcomes* 2016; 14: 17.
14. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007; 57: 1358–1362.
15. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982; 47: 149–174.
16. Andrich D. A Rating formulation for ordered response categories. *Psychometrika* 1978; 43: 561–573.
17. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002; 3: 85–106.
18. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med Res Methodol* 2008; 8: 33.
19. Wright BD, Linacre JM, Gustafson, JE, Martin-Lof, P. Reasonable mean-square fit values. *Rasch measurement transactions* 1994; 8: 370.
20. Lai JS, Teresi J, Gershon R. Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Eval Health Prof* 2005; 28: 283–294.
21. Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J Amer Stat Assoc* 1963; 58: 690–700.
22. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques. *DIFdetect and difwithpar*. *Med Care* 2006; 44: S115–123.
23. Linacre JM. Local independence and residual covariance: a study of Olympic figure skating ratings. *J Appl Meas* 2009; 10: 157–169.
24. Duncan PW, Bode RK, Lai SM, Perera S. Rasch analysis of a new stroke-specific outcome Scale: The Stroke Impact Scale. *Arch Phys Med Rehabil* 2003; 84: 950–963.
25. Ofenloch RF, Diepgen TL, Weisshaar E, Elsner P, Apfelbacher CJ. Assessing health-related quality of life in hand eczema patients: how to overcome psychometric faults when using the Dermatology Life Quality Index. *Acta Derm Venereol* 2014; 94: 658–662.