

Test–retest reliability of the Food Allergy Quality of Life Questionnaires (FAQLQ) for children, adolescents and adults

Jantina L. van der Velde · Bertine M. J. Flokstra-de Blok · Berber J. Vlieg-Boerstra ·
Joanne N. G. Oude Elberink · Jan P. Schouten · Audrey DunnGalvin ·
Jonathan O'B Hourihane · Eric J. Duiverman · Anthony E. J. Dubois

Accepted: 12 December 2008 / Published online: 14 January 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract

Objective The self-administered Food Allergy Quality of Life Questionnaire-Child Form (FAQLQ-CF), -Teenager Form (FAQLQ-TF) and -Adult Form (FAQLQ-AF) were recently developed within EuroPrevall, a multi-centred study of food allergy in Europe. The primary aim of this study was to evaluate the test-retest reliability of the FAQLQ-CF, -TF and -AF.

Methods One hundred and one Dutch patients (31 children, 34 adolescents and 36 adults) completed the FAQLQ twice with a 10–14 day interval. The intraclass correlation coefficient (ICC), Lin's concordance correlation coefficient (CCC) and Bland-Altman plots were used to assess test-retest reliability.

Results Test-retest reliability was excellent with ICCs and CCCs above 0.907, 0.975 and 0.951 for the FAQLQ-

CF, -TF and -AF, respectively. Bland-Altman plots showed that the mean differences of the test and re-test were all close to zero for the FAQLQs.

Conclusions The FAQLQs are reliable over a short time interval. The FAQLQs are not only excellent tools for group comparison studies, but also for monitoring individual patients.

Keywords EuroPrevall · Food allergy · Health-Related Quality of Life · Reproducibility · Test-retest reliability

Abbreviations

CCC	Concordance Correlation Coefficient
HRQL	Health-Related Quality of Life
ICC	Intraclass Correlation Coefficient
FAQLQ-CF	Food Allergy Quality of Life Questionnaire-Child Form
FAQLQ-TF	Food Allergy Quality of Life Questionnaire-Teenager Form
FAQLQ-AF	Food Allergy Quality of Life Questionnaire-Adult Form

J. L. van der Velde · B. M. J. Flokstra-de Blok (✉) ·
B. J. Vlieg-Boerstra · E. J. Duiverman · A. E. J. Dubois
Department of Pediatrics, Division of Pediatric Pulmonology
and Pediatric Allergy, University Medical Center Groningen,
University of Groningen, Groningen, P.O. Box 30.001, 9700 RB,
The Netherlands
e-mail: b.m.j.flokstra.de.blok@bkk.umcg.nl

J. N. G. Oude Elberink
Department of Internal Medicine, Division of Allergy,
University Medical Center Groningen, University of Groningen,
Groningen, P.O. Box 30.001, 9700 RB, The Netherlands

J. P. Schouten
Department of Epidemiology, University Medical Center
Groningen, University of Groningen, Groningen, P.O. Box
30.001, 9700 RB, The Netherlands

A. DunnGalvin · J. O. Hourihane
Department of Paediatrics and Child Health, University College,
Cork, Ireland

Introduction

Food allergy affects almost 4% of the general population in westernized countries [1], and it is the primary cause of anaphylaxis presenting to emergency departments [2]. The only proven therapy is careful avoidance of the causal food(s) and provision of medication for emergency treatment [3]. Consequently, patients often fear an allergic reaction and are continuously faced with dietary and social restrictions in their daily lives, which can have a negative impact on quality of life [4–11].

To measure Health-Related Quality of Life (HRQL), disease-specific questionnaires are significantly more sensitive than generic ones, and they are important for estimating the general burden of food allergy as well as measuring the response to interventions or future treatments. However, generic HRQL instruments allow comparison of the burden of disease between patient populations with different diseases [12]. Recently, as part of the EuroPrevall project, the first self-administered HRQL questionnaires specific for food allergy have been developed and validated: the Food Allergy Quality of Life Questionnaire-Child Form, -Teenager Form and -Adult Form (FAQLQ-CF, -TF, -AF). The FAQLQs showed good validity, internal consistency and discriminative abilities [13–16], but test-retest reliability was not extensively investigated.

Reliability measures are important to ensure that what the questionnaire is measuring is dependable and repeatable [12] and that it allows sample sizes to be determined for clinical trials [17]. The aim of this study was therefore to assess the test-retest reliability of the self-administered FAQLQ-CF, -TF and -AF.

Methods

Patients

We contacted Dutch children (8–12 years), adolescents (13–17 years) and adults (≥ 18 years) with food allergy, who were recruited from our clinic or by advertisement. We included patients with the most prevalent food allergies.

Questionnaires

The FAQLQ-CF contains 24 items and 4 domains, the FAQLQ-TF contains 23 items and 3 domains, and the FAQLQ-AF contains 29 items and 4 domains [13–15]. The total FAQLQ score is the sum of all the items divided by the number of items and ranges from 1 (minimal impairment in HRQL) to 7 (maximal impairment in HRQL) [18, 19].

Procedures

We sent the FAQLQs by mail to be completed at home. Regarding the FAQLQ-CF, parents were instructed that they were allowed to explain a question when needed, but they were not allowed to tell the child which answer to give. All patients who completed the first questionnaires (test) received the second questionnaires (re-test) 10–14 days after completion of the first. Patients who did not

respond in time were excluded from the study [20, 21] as well as patients who reported a clinically important change in disease between the measurements or within 2 months before the study. We defined a clinically important change in disease that could influence HRQL as a food allergic reaction of grade 3 or 4 according to the Mueller classification [22]. The study was approved by the local medical ethics review commission (METc 2005/051).

Statistical analysis

Data were analysed using SPSS software for Windows (version 14.0). To investigate test-retest reliability of the FAQLQs, we used the intraclass correlation coefficient (ICC), using a one-way ANOVA [20, 21, 23]. Values should be above 0.70 for group comparison studies and above 0.90–0.95 for individual measurements over time [24].

As a second measure of test-retest reliability, we calculated the Lin's concordance correlation coefficient (CCC). The different components of the CCC [Pearson correlation coefficient (measure of precision), location shift and scale shift (measures of accuracy)] were calculated. We plotted the first measurement against the second measurement, and we used major axis analyses to calculate the best fitting line [25].

Visual assessment of test-retest agreement was obtained by use of Bland-Altman plots [26]. Differences between the first and the second measurement were plotted against the mean of the first and the second measurement. Limits of agreement (mean difference $\pm 1.96 \times$ SD of the difference) were calculated, which reflect the interval within which about 95% of the differences between the two measurements should lie [27, 28]. A regression coefficient (r) was calculated to estimate a relationship between the difference and the mean [26].

Table 1 Patient recruitment

Patients	Children	Adolescents	Adults	Total
Contacted (n)	48	51	49	148
Returned 1st questionnaire (n)	41	47	43	131
Returned 2nd questionnaire (n)	38	38	38	114
Excluded (n)	7	4	2	13 ^a
Analysed (n)	31	34	36	101

^a Seven patients (three children, three adolescents and one adult) were excluded, because they completed the second questionnaire more than 14 days after completion of the first. One child and one adult were excluded because of a grade 3 or 4 allergic reaction between the first and second measurement. One child was excluded because she was aged under 8 years. Two children and one adolescent were excluded because they experienced their most severe reaction ever within 2 months before the first measurement

Table 2 Demographics and clinical characteristics

	Children (n = 31)	Adolescents (n = 34)	Adults (n = 36)
Mean age, years (SD)	10.6 (1.5)	15.0 (1.5)	37.3 (14.5)
Gender, n (%)			
Male	17 (55%)	18 (53%)	7 (19%)
Female	14 (45%)	16 (47%)	29 (81%)
Type of food allergy, n (%)			
Peanuts	25 (71%)	30 (88%)	25 (69%)
Nuts	17 (49%)	28 (82%)	25 (69%)
Milk	15 (43%)	15 (44%)	15 (42%)
Eggs	14 (40%)	16 (47%)	7 (19%)
Wheat	5 (14%)	4 (12%)	7 (19%)
Soy	9 (26%)	13 (38%)	8 (22%)
Sesame	7 (20%)	9 (26%)	6 (17%)
Fish	2 (6%)	5 (15%)	9 (25%)
Shellfish	6 (17%)	8 (24%)	12 (33%)
Celery	0 (0%)	4 (12%)	8 (22%)
Fruit	14 (40%)	13 (38%)	26 (72%)
Vegetables	6 (17%)	6 (18%)	10 (28%)
Others	25 (71%)	24 (71%)	13 (36%)
Number of food allergies, n (%)			
1 food	6 (19%)	3 (9%)	1 (3%)
2 foods	4 (13%)	4 (12%)	3 (8%)
3 foods	4 (13%)	8 (24%)	10 (28%)
>3 foods	17 (55%)	19 (56%)	22 (61%)
Severity of symptoms			
Mueller classification, n (%)			
Grade 1	6 (19%)	2 (6%)	3 (8%)
Grade 2	2 (6%)	3 (9%)	3 (8%)
Grade 3	17 (55%)	18 (53%)	13 (36%)
Grade 4	6 (19%)	9 (26%)	17 (47%)
Other ^a	0 (0%)	2 (6%)	0 (0%)
Most severe reaction, years ago (SD)	4.6 (3.6)	7.1 (5.4)	5.2 (7.5)
Diagnosed by, n (%)			
Specialist ^b	26 (83%)	25 (74%)	25 (69%)
Dietician	0 (0%)	1 (3%)	0 (0%)
General practitioner	4 (13%)	6 (18%)	3 (8%)
Alternative physician	1 (3%)	0 (0%)	3 (8%)
Patient	0 (0%)	0 (0%)	4 (11%)
Parents	0 (0%)	2 (6%)	1 (3%)

^a Other food allergy types not specified in the Mueller Classification, for example, the Oral Allergy Syndrome

^b Allergist, dermatologist or paediatrician

Results

Patients

We contacted 148 patients, of which 131 patients completed and returned the first questionnaire and 114 responded to the second questionnaire. This resulted in an overall response rate of 77%. A few patients were excluded, resulting in 101 patients that were eligible for analysing test-retest reliability (Table 1). The descriptive

characteristics are shown in Table 2. Mean duration between the first and second measurement was 11 days for all three age groups.

Analysis of FAQLQs

ICCs were ≥ 0.900 for the FAQLQs, and CCCs were comparably high. Location shift and scale shift should both be considered minimal according to Lin's examples [29].

Pearson correlation should be considered moderate in the FAQLQ-CF and good in the FAQLQ-TF and -AF (Table 3). Comparable results were found for the individual domains of the FAQLQs (data not shown).

Figure 1 illustrates the correlation between the first and second measurement. Major axis analysis revealed no significant differences of the slope and intercept of the best fitting line from the concordance line for the FAQLQ-CF and -TF. For the FAQLQ-AF there were significant but modest differences of the slope (1.10, $P = 0.046$) and the intercept (-0.612 , $P = 0.019$) of the best fitting line from the concordance line. The slope and intercept of the best fitting line of the FAQLQ-CF, -TF and -AF did not differ significantly from each other.

The Bland-Altman plots are shown in Fig. 2. About 95% of the differences lie within the 1.96 SD limits of agreement. There was no significant correlation between the mean of both scores and the differences of both scores for the FAQLQ-CF and -TF. There was a significant but modest correlation between the mean of both scores and the differences of both scores for the FAQLQ-AF ($r = -0.334$; $P = 0.046$). No significant systematic bias was observed, which means that mean differences of both scores were all close to zero. The limits of agreement are most narrow for FAQLQ-TF and wider for FAQLQ-CF and -AF.

Discussion

This article describes the evaluation of the test-retest reliability of the recently developed self-administered FAQLQ-CF, -TF and -AF. Overall, reliability was considered to be excellent for the FAQLQs as measured with the ICC and CCC. Additionally, Bland-Altman plots showed that mean differences were all close to zero, supporting the high reliability of the FAQLQs.

In this study we used ICCs calculated by a one-way ANOVA, CCCs and Bland-Altman plots to assess test-retest reliability. However, different methods can be used to assess test-retest reliability, and there is much discussion in literature on the best way to do this [20]. A disadvantage of the ICC is that if patient groups are very homogeneous, the ICC tends to be low, because the ICC compares variance among patients to total variance. If patient groups are very heterogeneous, the ICC tends to be high. Thus, the ICC would only generalise to similar populations. Additionally, the one-way ICC does not take into account the order in which observations were taken [29]. Therefore, the CCC is a useful additional measure. The CCC takes into account not only mean differences between the first and second measurement, such as ICCs calculated by a one-way ANOVA, but also takes into account variance

Table 3 Reliability and agreement measures of the FAQLQs

	FAQLQ-CF	FAQLQ-TF	FAQLQ-AF
M 1 (SD)	4.13 (1.15)	4.37 (1.20)	4.49 (1.44)
M 2 (SD)	4.08 (1.34)	4.42 (1.29)	4.34 (1.59)
MB (SD)	4.11 (1.22)	4.40 (1.24)	4.41 (1.50)
MD (SD)	0.045 (0.537)	-0.051 (0.274)	0.147 (0.451)
Limits of agreement (1.96 SD)	-1.008 to 1.097	-0.588 to 0.486	-0.737 to 1.031
ICC one-way (95% CI)	0.910 (0.823–0.955)	0.976 (0.952–0.988)	0.952 (0.909–0.975)
Error variance	0.147	0.038	0.102
CCC (95% CI)	0.907 (0.847–0.967)	0.975 (0.959–0.991)	0.951 (0.921–0.981)
Scale shift	1.162	1.077	1.104
Location shift	0.036	-0.041	0.097
Pearson	0.918	0.978	0.960
Kendall's tau-b	0.759	0.888	0.780

M 1 = Total FAQLQ score measurement 1

M 2 = Total FAQLQ score measurement 2

MB = Mean FAQLQ score of both measurements

MD = Mean difference between measurement 1 and 2 ($M1 - M2$)

SD = Standard deviation

CI = Confidence interval

Limits of agreement: $MD \pm 1.96$ SD of the MD

ICC = Intraclass correlation coefficient

CCC = Concordance correlation coefficient

Scale shift ($SD2/SD1$)

Location shift: $(M1 - M2)/\sqrt{SD1 * SD2}$

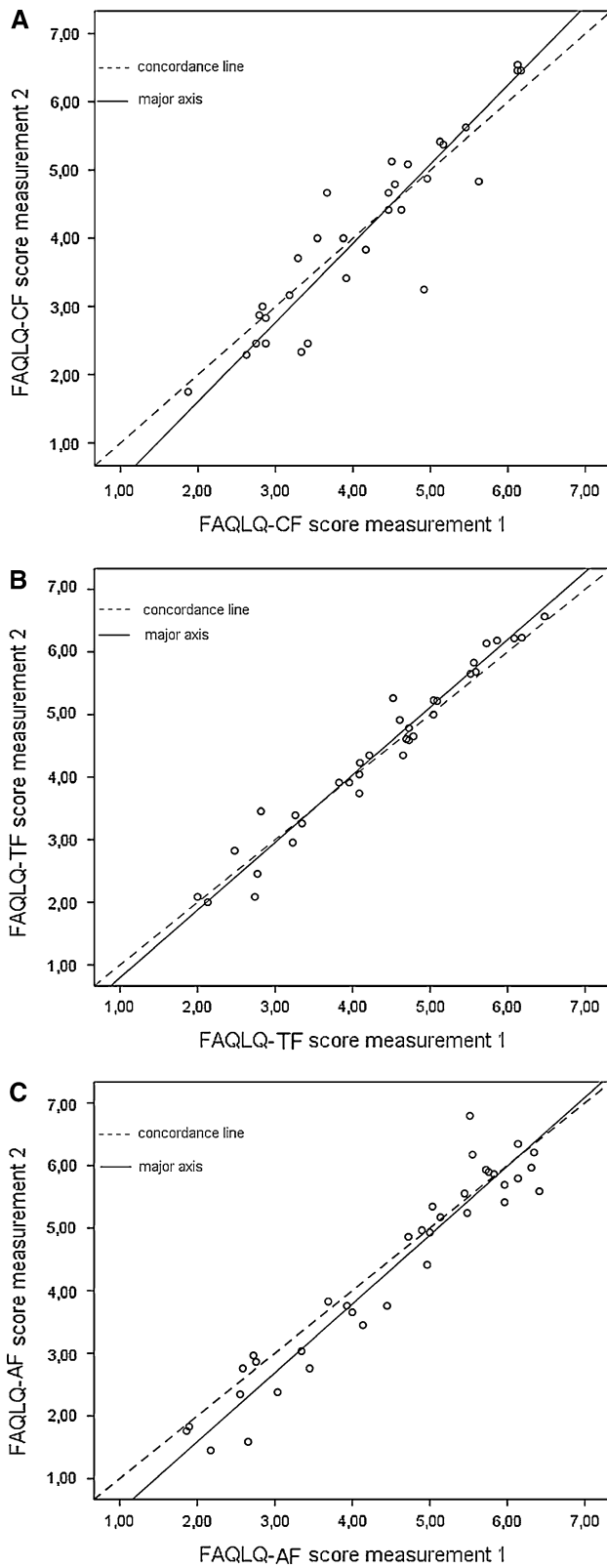


Fig. 1 FAQLQ score of the first measurement against the FAQLQ score of the second measurement with 45° line through the origin in (A) children, (B) adolescents and (C) adults

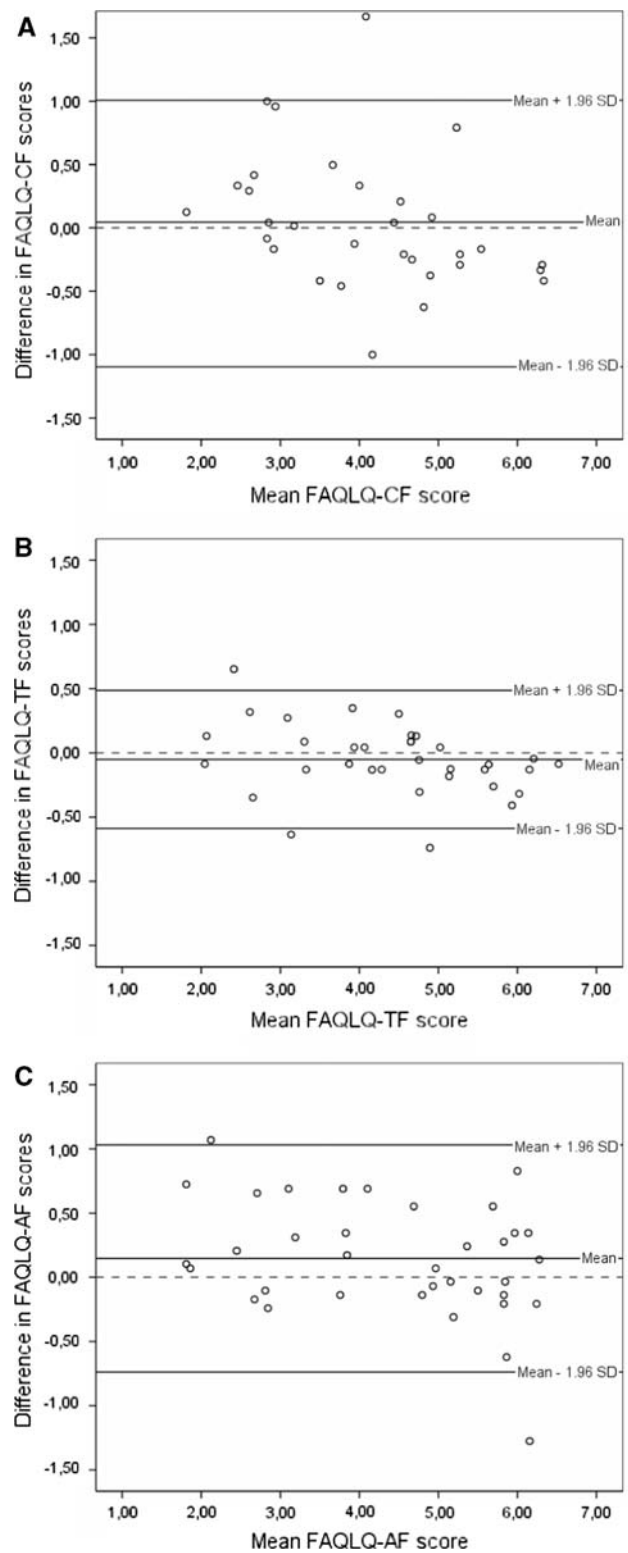


Fig. 2 Bland-Altman plots for the FAQLQs in (A) children, (B) adolescents and (C) adults. The mean of both measurements are plotted against the difference of both measurements (calculated as first measurement minus second measurement)

differences between the first and second measurement by reducing the magnitude of the resulting test-retest reliability estimate. In addition, the CCC is a better tool to distinguish between bias and imprecision [20, 29]. There can be large differences in ICC and CCC scores, especially in studies with heterogeneous groups. The similar scores we found in our study reflect that both coefficients worked very well in this population and that results can be generalised to other groups. Bland-Altman plots are very illustrative in assessing test-retest agreement. They were useful to identify some extreme and outlying differences, to analyse the magnitude of the measurement error, which was small, and to visualise a possible relationship between the difference and the mean of both scores [26].

This study may also have some limitations. Firstly, the sample sizes were relatively small. However, we found that the reliability of the questionnaires was very high, which indicates that the sample sizes were adequate and that a greater number of patients would probably not have influenced the outcomes. Another limitation may be that the majority of adults in this study was female. However, we did not find significant differences in the test-retest reliability outcomes between men and women (data not shown). Therefore, we think that the imbalance between men and women did not influence the generalisability of the results of the FAQLQ-AF. Finally, the significant correlation between the first and second measurement of the FAQLQ-AF (Fig. 1C) and between the mean of both scores and the differences of both scores of the FAQLQ-AF (Fig. 2C) was an unexpected finding. We think this correlation might be due to an outlier. This assumption was supported by a re-analysis excluding this outlier, which showed that the correlation was no longer significant.

In summary, the FAQLQs clearly showed excellent reliability and are thus promising measures in evaluative studies in patients with food allergy, but also in monitoring individual patients. The high test-retest reliability supports the value of the FAQLQs for clinical trials with relatively small sample sizes. We recommend the use of the FAQLQs in clinical trials of current management strategies of food allergy, and they may also be useful when new treatments become available. Currently, the longitudinal validity of the FAQLQs and the validity of several other European language versions of the FAQLQs are being investigated.

Acknowledgement This work was funded by the EU through the EuroPrevall project (FOOD-CT-2005-514000).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Osterballe, M., Hansen, T. K., Mortz, C. G., Host, A., & Bindslev-Jensen, C. (2005). The prevalence of food hypersensitivity in an unselected population of children and adults. *Pediatric Allergy and Immunology*, 16(7), 567–573. doi:10.1111/j.1399-3038.2005.00251.x.
- Sampson, H. A. (2004). Food-induced anaphylaxis. *Novartis Foundation Symposium*, 257, 161–171. doi:10.1002/0470861193.ch13.
- Sampson, H. A. (2004). Update on food allergy. *The Journal of Allergy and Clinical Immunology*, 113(5), 805–819. doi:10.1016/j.jaci.2004.03.014.
- Primeau, M. N., Kagan, R., Joseph, L., Lim, H., Dufresne, C., Duffy, C., et al. (2000). The psychological burden of peanut allergy as perceived by adults with peanut allergy and the parents of peanut-allergic children. *Clinical and Experimental Allergy*, 30(8), 1135–1143. doi:10.1046/j.1365-2222.2000.00889.x.
- Sicherer, S. H., Noone, S. A., & Munoz-Furlong, A. (2001). The impact of childhood food allergy on quality of life. *Annals of Allergy, Asthma & Immunology*, 87(6), 461–464.
- Avery, N. J., King, R. M., Knight, S., & Hourihane, J. O. (2003). Assessment of quality of life in children with peanut allergy. *Pediatric Allergy and Immunology*, 14(5), 378–382. doi:10.1034/j.1399-3038.2003.00072.x.
- Cohen, B. L., Noone, S., Munoz-Furlong, A., & Sicherer, S. H. (2004). Development of a questionnaire to measure quality of life in families with a child with food allergy. *The Journal of Allergy and Clinical Immunology*, 114(5), 1159–1163. doi:10.1016/j.jaci.2004.08.007.
- Marklund, B., Ahlstedt, S., & Nordstrom, G. (2004). Health-related quality of life among adolescents with allergy-like conditions—with emphasis on food hypersensitivity. *Health and Quality of Life Outcomes*, 2, 65. doi:10.1186/1477-7525-2-65.
- Marklund, B., Ahlstedt, S., & Nordstrom, G. (2006). Health-related quality of life in food hypersensitive schoolchildren and their families: parents' perceptions. *Health and Quality of Life Outcomes*, 4, 48. doi:10.1186/1477-7525-4-48.
- Lebovidge, J. S., Stone, K. D., Twarog, F. J., Raiselis, S. W., Kalish, L. A., Bailey, E. P., et al. (2006). Development of a preliminary questionnaire to assess parental response to children's food allergies. *Annals of Allergy, Asthma & Immunology*, 96(3), 472–477.
- Bollinger, M. E., Dahlquist, L. M., Mudd, K., Sonntag, C., Dilling, L., & McKenna, K. (2006). The impact of food allergy on the daily activities of children and their families. *Annals of Allergy, Asthma & Immunology*, 96(3), 415–421.
- Testa, M. A., & Simonson, D. C. (1996). Assessment of quality-of-life outcomes. *The New England Journal of Medicine*, 334(13), 835–840. doi:10.1056/NEJM199603283341306.
- Flokstra-de Blok, B. M. J., DunnGalvin, A., Vlieg-Boersta, B. J., Oude Elberink, J. N. G., Duiverman, E. J., Hourihane, J. O., et al. (2008). Development and validation of a self-administered Food Allergy Quality of Life Questionnaire for children. *Clinical and Experimental Allergy*, 39(1), 127–137.
- Flokstra-de Blok, B. M. J., DunnGalvin, A., Vlieg-Boersta, B. J., Oude Elberink, J. N. G., Duiverman, E. J., Hourihane, J. O., et al. (2008). Development and validation of a self-administered Food Allergy Quality of Life Questionnaire for adolescents. *The Journal of Allergy and Clinical Immunology*, 122(1), 139–144. doi:10.1016/j.jaci.2008.05.008.
- Flokstra-de Blok, B. M. J., DunnGalvin, A., Vlieg-Boersta, B. J., Oude Elberink, J. N. G., Duiverman, E. J., Hourihane, J. O., et al. Development and validation of the first disease-specific quality of life questionnaire for adults; The Food Allergy Quality of Life Questionnaire-Adult Form (FAQLQ-AF). *Allergy* (in press).

16. DunnGalvin, A., Flokstra-de Blok, B. M. J., Burks, A., Dubois, A. E. J., & Hourihane, J. O. (2008). Food Allergy Quality of Life Questionnaire (FAQLQ-PF) for children aged 0–12 years: Content, construct, and cross-cultural validity. *Clinical and Experimental Allergy*, 38(6), 977–986. doi:10.1111/j.1365-2222.2008.02978.x.
17. Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases*, 40(2), 171–178. doi:10.1016/0021-9681(87)90069-5.
18. Oude Elberink, J. N., de Monchy, J. G., Golden, D. B., Brouwer, J. L., Guyatt, G. H., & Dubois, A. E. (2002). Development and validation of a health-related quality-of-life questionnaire in patients with yellow jacket allergy. *The Journal of Allergy and Clinical Immunology*, 109(1), 162–170. doi:10.1067/mai.2002.120552.
19. Juniper, E. F., Guyatt, G. H., & Jaeschke, R. (1996). Quality of life and pharmacoeconomics in clinical trials. In B. Spilker (Ed.), *How to develop and validate a new health-related quality of life instrument* (2nd ed., pp. 49–56). Philadelphia: Lippincott-Raven Publishers.
20. Schuck, P. (2004). Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Quality of Life Research*, 13(3), 571–586. doi:10.1023/B:QURE.0000021318.92272.2a.
21. Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials*, 12(4, Suppl), 142S–158S. doi:10.1016/S0197-2456(05)80019-4.
22. Mueller, H. L. (1966). Diagnosis and treatment of insect sensitivity. *The Journal of Asthma Research*, 3(4), 331–333. doi:10.3109/02770906609106941.
23. Terwee, C. B., Gerding, M. N., Dekker, F. W., Prummel, M. F., van der Pol, J. P., & Wiersinga, W. M. (1999). Test–retest reliability of the GO-QOL: A disease-specific quality of life questionnaire for patients with Graves’ ophthalmopathy. *Journal of Clinical Epidemiology*, 52(9), 875–884. doi:10.1016/S0895-4356(99)00069-4.
24. Scientific Advisory Committee of the Medical Outcome Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, 11(3), 193–205. doi:10.1023/A:1015291021312.
25. Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews of the Cambridge Philosophical Society*, 81(2), 259–291. doi:10.1017/S1464793106007007.
26. Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135–160. doi:10.1191/096228099673819272.
27. Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307–310.
28. Bland, J. M., & Altman, D. G. (1996). Measurement error and correlation coefficients. *BMJ (Clinical Research Ed.)*, 313(7048), 41–42.
29. Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1), 255–268. doi:10.2307/2532051.