Reliability and validity of functional health status and health-related quality of life questionnaires in children with recurrent acute otitis media

Carole N. M. Brouwer · Anne G. M. Schilder ·
Henk F. van Stel · Maroeska M. Rovers ·
Reinier H. Veenhoven · Diederick E. Grobbee ·
Elisabeth A. M. Sanders · A. Rianne Maillé

Abstract In this study the reliability and validity of generic and disease-specific questionnaires has been assessed focusing on responsiveness. This is part of a study on the effects of recurrent acute otitis media (rAOM) on functional health status (FHS) and health-related quality of life (HRQoL) in 383 children with rAOM participating in a randomized clinical trial. The following generic questionnaires were studied: 1. RAND general health rating index, 2. Functional Status Questionnaire (FSQ Generic and FSQ Specific), 3. TNO-AZL Infant Quality of Life (TAIQOL), and the following disease-specific questionnaires: 1. Otitis Media-6 (OM-6), 2. Numerical rating scales (NRS) for child and caregiver (NRS Child and NRS Caregiver), and 3. a new Family Functioning Questionnaire (FFQ). Reliability was good to excellent (Cronbach's α range 0.80–0.90, intraclass correlation coefficient range 0.76–0.93). Moderate to strong correlations were found between the questionnaires as well as between questionnaires and relevant clinical indicators ($r = 0.29$–0.49), demonstrating construct validity. Discriminant validity for children with few versus frequent episodes of acute otitis media per year was good for most questionnaires ($P < 0.004$) but poor for the otitis media-related subscales of the TAIQOL ($P = 0.10$–0.97) and both NRS ($P = 0.22$ and 0.48). Except for the TAIQOL subscales, change scores were significant ($P < 0.003$) for generic and disease-specific questionnaires. Effect sizes were somewhat higher for disease-specific compared to generic questionnaires (0.55–0.95 versus 0.32–0.60) except for the TAIQOL subscales, which showed very poor sensitivity to change. Anchor-based methods resulted in a somewhat larger range of estimates of MCID than distribution-based methods. Combining distribution-based and anchor-based methods resulted in similar ranges for the minimally clinical important differences for generic and disease-specific questionnaires: 2–15 points on a 0–100 scale. Apart from the generic TAIQOL subscales, both generic and disease-specific questionnaires used in this study showed good psychometric qualities and responsiveness for use in clinical studies on children with rAOM.

Keywords Childhood infection · Acute otitis media · Functional health status · Quality of life · Reliability · Validity · Responsiveness

C. N. M. Brouwer · R. H. Veenhoven
Department of Paediatrics, Spaarne Hospital, Postbus 770,
2130 AT Hoofddorp, The Netherlands

C. N. M. Brouwer · H. F. van Stel · M. M. Rovers ·
D. E. Grobbee · A. R. Maillé (✉)
Julius Center for Health Sciences and Primary Care, University
Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The
Netherlands
e-mail: R.Maille@umcutrecht.nl

A. G. M. Schilder
Department of Otorhinolaryngology, Wilhelmina Children's
Hospital/University Medical Center Utrecht, Utrecht, The
Netherlands

E. A. M. Sanders
Department of Paediatric Immunology, Wilhelmina Children's
Hospital/University Medical Center Utrecht, Utrecht, The
Netherlands

Introduction

Acute otitis media (AOM) is a common childhood infection with a peak incidence occurring between 6 and 12 months of age. Five to fifteen percent of all children, depending on their age, suffer from *recurrent* acute infections of the middle ear (4 or more episodes per year) [1–4]. Repetitive

episodes of pain, fever and general illness during acute ear infections [5–8] as well as worries about potential long-term sequelae such as hearing loss and disturbed language development [9–13] may all compromise the quality of life of the child and its family [14–16]. Although several questionnaires have been used in assessing the effects of recurrent acute otitis media (rAOM) in children, lack of true health-related quality of life (HRQoL) questionnaires as well as incomplete data on their reliability and validity mean that our current knowledge on the subject is limited for both research and clinical practice [17].

Assessment of functional health status (FHS) and HRQoL, as defined in Table 1 [18–26], has become increasingly important in clinical trials on the effectiveness of treatment in paediatric chronic conditions. The validation of FHS and HRQoL questionnaires, however, has so far mainly focused on reliability and construct validity. Responsiveness has been assessed for only a few paediatric HRQoL questionnaires for conditions other than otitis media [27–31]. In order to evaluate treatment effects on FHS and HRQoL meaningfully, questionnaires are needed that are not only reliable and valid but also responsive to changes in FHS and HRQoL. In adult studies, various strategies have been used to assess responsiveness, which is defined as the ability to detect clinically important change over time and therefore involves both the assessment of sensitivity to change and the assignment of meaning to that change [32, 33]. Since none of these strategies is without limitations, we will try to assess the responsiveness of FHS and HRQoL questionnaires by using multiple strategies, categorized into distribution-based and anchor-based methods.

Distribution-based methods express the amount of change relative to the amount of random variance of a questionnaire [34, 35], whereas anchor-based methods enhance interpretability of changes in questionnaire scores by linking meaning and clinical relevance to change scores [34, 36].

Both generic and disease-specific questionnaires have been used in studies of paediatric FHS or HRQoL. Generic questionnaires span a wide spectrum of quality of life components, bridging various health states and populations. Disease-specific questionnaires on the other hand, assess health-related issues specific to particular conditions and may be able to detect small changes that are often small but clinically important; these provide a more detailed assessment of HRQoL, but cannot be used for comparisons across health conditions [37–39]. Both questionnaires are often combined in order to profit from the merits of both types. However, there have been few head-to-head comparisons between generic and disease-specific HRQoL measurement questionnaires in the setting of randomized controlled trials (RCT) [40].

The current RCT on the effectiveness of pneumococcal vaccination in children with rAOM will address both the issues of using generic versus disease-specific questionnaires and responsiveness in evaluating treatment effects on HRQoL in RCTs. The results will lead to recommendations regarding the applicability of these questionnaires in clinical studies in children with rAOM.

Methods

Setting and procedure

FHS and HRQoL were assessed in 383 children with rAOM participating in a double-blind randomized, placebo-controlled trial on the effectiveness of pneumococcal conjugate vaccination versus control hepatitis vaccination. The study was conducted at the paediatric outpatient departments of a general hospital (Spaarne Hospital Haarlem) and a tertiary care hospital (University Medical Center Utrecht). Children were recruited for this trial through referral by general practitioners, paediatricians, or otolaryngologists, or were enrolled on the caregiver's own initiative from April 1998 to February 2001.

Study population

Inclusion criteria: children were aged between 12 and 84 months and suffering from rAOM at study entry; defined in this study as having had at least 2 episodes of physician diagnosed AOM in the year prior to study entry. Exclusion criteria were conditions with a known increased risk for AOM such as: known immunodeficiency (other than IgA or IgG2 subclass deficiency), cystic fibrosis, immotile cilia syndrome, cleft palate, chromosomal abnormalities (like Down syndrome) or severe adverse events upon vaccination in the past.

At each scheduled visit, two research physicians (C.N.M.B. and R.H.V.) collected data regarding the

Table 1 Definitions of health-related quality of life and functional health status

Health-related quality of life:	Level of satisfaction a person inputes to those aspects of his or her life that are affected by the effects of illness and its treatment [18–20]. Incorporation of a person's valuation of his life distinguishes HRQoL from other measures of well-being [21, 22].
Functional health status:	Reflection of the (severity of) signs and symptoms and the adequacy of daily functioning across various life-domains in an individual with a certain health condition [23–26].

number of episodes of AOM (based on parental report at baseline and on physician report during follow-up), upper respiratory tract infections, and pneumonia. Information about the medical treatment, and ear, nose, and throat surgery in the preceding 6 months was also collected. The primary caregivers completed questionnaires assessing FHS and HRQoL of their child and family during the clinic visits at baseline and at 7, 14, and 26 months follow-up. Caregivers were requested to have the same person complete the questionnaires each time and to rate their child's FHS and HRQoL with regard to their recurrent episodes of acute otitis media. Informed consent was obtained from caregivers of all children before study entry. Medical ethics committees of both participating hospitals approved the study protocol.

Questionnaires

Four generic questionnaires (RAND, FSQ Generic, FSQ Specific, TAIQOL) and one disease-specific questionnaire (OM-6) were used to assess FHS and HRQoL of the children in the study. Additionally, two disease-specific one-item numerical rating scales (NRS Child and NRS Caregiver) were used to obtain a global rating of HRQoL of the child and of the caregiver, respectively, related to rAOM. For the assessment of the impact of rAOM on family functioning a newly composed disease-specific questionnaire the Family Functioning Questionnaire (FFQ), was used to assess the impact of rAOM on family functioning. Table 2 summarises the characteristics of the questionnaires [14, 41–57].

Generic questionnaires

The RAND general health-rating index (RAND) and the Functional Status Questionnaire (FSQ) had already been translated and validated for Dutch children by Post et al. [41, 42] (Table 2). The RAND assesses general health perceptions of caregivers regarding their child [43]. The FSQ consists of two parts: one measuring functional limitations in general, not necessarily related to illness (FSQ Generic) and the other (paradoxically named FSQ *Specific*) measuring functional limitations that are attributable to *any* illness [43]. Functional limitations in both versions of the FSQ are mainly expressed as behavioural problems. During the course of the study, a new Dutch questionnaire on generic HRQoL became available: the TNO-AZL Infant Quality of Life (TAIQOL) questionnaire [51, 53]. For this reason, from July 1999 the TAIQOL was added to the previously selected set of questionnaires. Although the full, original version of the TAIQOL has been applied during the study, only those subscales from the TAIQOL are discussed that, based on their content, were assumed to be

Table 2 Characteristics of FHS and HRQoL questionnaires used in this study

Questionnaires	Type; number of items; scale	Construct(s) measured	Application in other studies
Generic			
RAND	FHS; 7; Likert	General health: current health; previous health; resistance to illness	Low-birth-weight children; survivors of childhood cancer; asthmatic children [42, 44, 47, 48]
FSQ generic	FHS; 14; Likert	Age appropriate functioning and emotional behaviour	Low-birth-weight children; survivors of childhood cancer; asthmatic children [41, 45–47, 49–51]
FSQ specific	*Idem* to FSQ Generic, measuring general impact of illness on functioning and behaviour		
TAIQOL	HRQoL; 35/46*; Likert	Sleeping, appetite, lung problems, stomach problems, skin problems, motor functioning, problem behaviour, social functioning, communication, positive mood, anxiety, liveliness	Low-birth-weight children, childen with chronic illness, children with chronic OME [51, 54]
Disease-specific			
OM-6	FHS; 6; Likert	Physical suffering; hearing loss; speech impairment; emotional distress; activity limitations; caregiver concerns	Children with recurrent AOM; children with chronic OME [14, 55–57]
NRS Child	HRQoL; 1; index 0–100	Global well-being of child related to AOM episodes	Children with recurrent AOM or chronic OME
NRS Caregiver	HRQoL; 1; index 0–100	Global well-being of parent related to child's AOM episodes	none [14]
Family Functioning Questionnaire (FFQ)	FHS; 7; Likert	Parents: sleep deprivation; change of daily or social activities; emotional distress. Family: cancelling family plans or trips. Siblings: feeling neglected; demanding extra attention.	None

* 46 items when age > 15 months

sensitive to the consequences of AOM. The following subscales tap functional items that are often affected by AOM (OM-related): 'Sleeping', 'Appetite', 'Liveliness', 'Problem behaviour', 'Positive mood', and 'Communication' (items about speech and language capacity) which are 6 of the 12 subscales in the TAIQOL. Although the TAI-QOL has been developed for children aged up to 5 years, we also used the questionnaire in children aged 6–7 years, as no appropriate alternative was available during the study.

Disease-specific questionnaires

To measure disease-specific FHS, the Otitis Media-6 (OM-6) [14, 55] was translated into Dutch according to principles of backward–forward translation [58–61]. This six-item questionnaire covers both acute and long-term functional effects of otitis media in children on FHS.

A new questionnaire has been developed to assess the impact of rAOM in children on their caregivers and siblings: the FFQ. The content of the FFQ was based on previous work by Asmussen et al. [15, 62] on the impact of rAOM on family well-being. A panel of paediatric otorhinolaryngologists and paediatricians from our study sites selected the items most relevant according to their clinical experience. The FFQ is composed of six questions covering effects of the child's rAOM on caregiver and family activities and two questions assessing these effects on emotional behaviour of the other siblings. The Likert-scale was used as a response format and was analogous to that of the RAND and OM-6 in our study, ranging from score 1–4.

Furthermore, two numerical rating scales (NRS) (0–100) were used, the NRS Child and the NRS Caregiver (see Table 2). The NRS Child [14] was translated into Dutch using the same principles of backward–forward translation that have been applied to translation of the OM-6. The newly created NRS Caregiver was modelled upon the NRS Child and added to the previously selected set of questionnaires from July 1999. The NRS caregiver has been created in this study, following the example of the NRS child which was created by Rosenfeld et al. [14].

Finally, the Dutch version of the OM Functional Status Questionnaire specific (OMFSQ [52]) was included as an anchor for responsiveness (instrument description in section on responsiveness).

Questionnaire application

Questionnaires were completed in a randomly selected, but fixed order during the follow-up assessments to prevent possible influence of order effects [63, 64]: RAND, FSQ Generic and Specific, OM-6, NRS Child, FFQ, TAIQOL,

OMFSQ, NRS Caregiver. For all questionnaires higher scores indicate the presence of a better HRQoL or FHS. To allow comparisons between scores on the questionnaires, all scores were linearly transformed into 0–100 Scales. For each questionnaire, the evaluation period was the 6 weeks before completion.

Statistical analyses

Floor and ceiling effects

Floor and ceiling effects were estimated for the baseline-assessment of each questionnaire by calculating percentages of respondents that had minimum and maximum scores, respectively. Questionnaires should exhibit minimal floor and ceiling effects to be optimally able to detect difference and change.

Reliability

First, internal consistency was assessed by calculating Cronbach's alpha, which should be above 0.70 for each questionnaire or subscale [65]. Inter-item correlations of questionnaires were assessed to reveal item redundancy or 'hidden' subscales that may erroneously yield a high overall Cronbach's alpha.

For the assessment of test–retest reliability, a subset of caregivers attending the outpatient ward from February 2000 to June 2001 (n = 160) was given a second set of the same questionnaires (retest) to complete at home. The time frame for completion was 2 weeks after the first set of questionnaires was filled out during the outpatient visit at 14 months (first test). Children with AOM at the first test were excluded, since differences in their scores could be due to real change and interfere with the assessment of reliability.

For the assessment of test–retest reliability, a time-interval of 2–14 days is often considered long enough to prevent recall bias and too short for relevant change to occur in chronic disease [66]. Test–retest reliability was computed as the intraclass correlation coefficients (ICC) between the two sets of questionnaires. An ICC of 0.80 was considered the required minimum for good reliability [65, 67].

Construct and discriminant validity

In order to demonstrate construct validity, hypotheses were formulated about the strength of correlations between questionnaires. A higher percentage of correct predictions indicates stronger support for construct validity. A correlation of 0.10–0.30 was defined as weak, 0.30–0.50 as moderate, and >0.50 as strong [68]. The correlation between FSQ Generic and NRS Caregiver was predicted

to be weak since they were expected to assess two different constructs. Moderate to strong correlations ($r > 0.40$) were predicted between RAND and NRS Caregiver. Moderate to strong correlations were also expected between OM-6 and FSQ Specific, NRS Child, NRS Caregiver and FFQ, as all assess otitis media-related HRQoL or FHS. The correlation between FSQ Generic and FSQ Specific was expected to be strong ($r > 0.50$). The remaining correlations among the questionnaires were expected to be moderate (Table 5). Additionally, correlations between questionnaire scores and frequency of physician visits for upper respiratory tract infections as well as frequency of AOM episodes in the preceding 6 months were calculated. Since distributions of questionnaire scores were skewed, correlations were assessed using Spearman's rho.

Discriminant validity was assessed by dichotomizing the study participants in children with 2–3 versus 4 or more episodes of otitis media per year. Based on clinical and immunological data, children with 4 or more AOM episodes per year are considered as 'otitis prone' [2, 69–71], reflecting a sub-group with an increased rate of upper respiratory tract infections, related medical interventions and compromised child functioning [72, 73]. It was assumed that this group would perform significantly poorer than children with 2–3 otitis media episodes per year on all questionnaires, which was assessed by independent sample Mann–Whitney tests.

Responsiveness

Since pneumococcal conjugate vaccination showed no clinical effectiveness when compared to the control vaccine [74], the intervention could not be used as an external criterion of change. Data of both vaccine groups were pooled instead for the assessment of responsiveness to spontaneous remission. The clinical experience of a panel of 5 experts in the field of otitis media, formed the basis for defining a reduction of 2 or more episodes of AOM per child per year as the external criterion for change while a reduction of 1 episode or less identified no change. Responsiveness was evaluated for two intervals: from 0 to 7 months and from 7 to 14 months follow-up. The observed change in these episodes was multiplied by 12/7 (1,714) to get the estimated change per year.

The first step in the assessment of responsiveness was to explore the ability of questionnaires to detect change at all, i.e., its sensitivity to change. Secondly, meaning and clinical relevance of the change score were determined in accordance with recent recommendations, using both distribution- and anchor-based methods [36, 75–77]. Distribution-based methods express the amount of change relative to the amount of random variance of a questionnaire [34, 35]. Some ratios of change to random variance have, often empirically, been found to represent a minimally clinical important difference. Anchor-based methods enhance interpretability of changes in questionnaire scores by linking meaningful and clinically relevant indicators to change scores [34, 36].

The assessment of responsiveness will be described in further detail below.

Sensitivity to change

Sensitivity to change was assessed by calculating both the statistical significance of change scores using a paired t-test or Wilcoxon matched pairs test (for skewed distributions), and effect sizes (ES) using Guyatt's responsiveness statistic [78] for changed subjects. In this statistic, the observed change that occurred in changed subjects is related to the observed random change or random error, in unchanged subjects. A parametric effect size was computed as: mean change score changed group/SD (change score unchanged group); a nonparametric effect size was computed as: median change score changed group/interquartile range (change score unchanged group)).

According to the benchmarks of Cohen [79], an effect size of 0.2 represents a small change, 0.5 a moderate change and 0.8 or higher represents a large change. For skewed distributions Wilcoxon matched pairs test was used to calculate the significance of change.

Clinical relevance of change scores

The interpretation of change is often assessed by calculating the minimally clinical important difference (MCID), which is the smallest difference in a questionnaire total or domain score that patients perceive as beneficial [80]. The MCID can be computed from both distribution-based and anchor-based methods. Several estimates of the MCID from both methods are reported, to assess the likely range of the MCID for each questionnaire.

Interpretation of change—distribution-based methods (ES-MCID and SEM-MCID)

The main distribution-based methods for assessing the MCID are the Effect Size and the Standard Error of Measurement. A change in questionnaire scores corresponding to the effect size of Guyatt's Responsiveness Statistic with values of 0.3–0.5 has been found to be consistent with other (empirical) estimates of the MCID [36, 81–83]. In this study the change in questionnaire scores corresponding with an effect size of 0.3 is used as benchmark of MCID (ES-MCID). A change of one Standard Error of Measurement (1-SEM) has empirically been found to correspond with the MCID of a questionnaire [77,

84–86]. The 1-SEM of a questionnaire links reliability of an instrument to the variance of scores in a population as reflected in its formula: 1-SEM = SD (change scores unchanged subjects) * $\sqrt{(1\text{-}ICC)}$. It is an estimate of what part of the observed change may be due to random measurement error by including distribution of scores (SD) and instrument reliability (ICC). Change larger than the SEM therefore is considered 'real' change. The SEM is here used as an estimate of the MCID (SEM-MCID). The ES-MCID and SEM-MCID support the interpretation of measured change, as they reflect the smallest change that is substantially larger than the random variability in the study population which is based on the standard deviation of the unchanged subjects.

Interpretation of change—anchor-based methods

Anchor-based methods require an independent standard, the anchor, that in itself is easily interpretable and that is at least moderately correlated (>0.3) with the questionnaire being assessed. Changes in questionnaire scores were compared with change in two clinically relevant anchors: the AOM *frequency* (incidence of AOM episodes per child) and the AOM *severity* assessed with the Dutch version of the OM-Functional Status Questionnaire specific (OM-FSQ) [52]. The OM-FSQ was used as an anchor for responsiveness. It consists of three questions assessing clinical AOM severity: earache, sleeping problems, and other signs and symptoms (irritability, fussiness, fever) that may indicate the presence of an ear infection. In our population, the OM-FSQ demonstrated high internal consistency (Cronbach's α = 0.88) and good test–retest reliability (ICC = 0.94). The OM-FSQ correlated weakly with the NRS Child (Spearman's rho = 0.18), but moderately with the RAND (0.36), FSQ Generic (0.37), and NRS Caregiver (0.34), and strongly with the FSQ Specific (0.52), OM-6 (0.73) and FFQ (0.61).

In relation to the AOM frequency, an expert panel in the field of otitis media considered a reduction of 2 episodes per year as a small or minimal clinically important change, whereas a change of 3 to 4 episodes per year was considered moderate to large. In the study of Alsarraf et al. [52], the OM-FSQ total score was about 62 on a scale of 0–100 during an episode of AOM, increasing to 92 at 6 weeks and to 90 at 12 weeks after an episode of AOM with higher scores reflecting less severe ear-related symptoms. Therefore, a score change of 10–20 on the 0–100 scale of the OM-FSQ in the current population was considered to be a small clinically relevant change in AOM severity, a score change of 30–50 as moderate to large. Anchor-based estimates of the MCID were computed as the change in questionnaire scores associated with small changes in AOM frequency and OM-FSQ.

For all analyses the Statistical Package for the Social Sciences (SPSS) version 10.1 was used.

Results

Population

The population characteristics summarized in Table 3 show that the majority of children suffered from 4 or more AOM episodes per year, and half of them suffered from chronic airway problems or atopic symptoms. Most children had undergone one or more ENT surgeries. Overall they seemed to suffer from more severe disease than the average child with 2–3 middle ear infections, as stated earlier.

Floor and ceiling effects

Generally, the questionnaires demonstrated no floor-effects. However, Table 4 shows that some questionnaires (FSQ Specific and FFQ) and most TAIQOL subscales showed moderate to large ceiling effects, which indicates that measurement of improvement may be limited while it may actually be present.

Table 3 Characteristics of study population*

	Mean or % (n = 383)	SD or 95% CI
Age (months)	34	(19.7)
Male gender	62%	(57–67)
In the year prior to inclusion		
Number of AOM episodes/year	5.0	(2.7)
2–3	37%	(32–42)
4–5	31%	(26–36)
6 or more	32%	(27–37)
Impaired hearing**	35%	(30–40)
Language or speech problems**	22%	(18–26)
History of		
Chronic airway problems or atopic symptoms ***	51%	(46–56)
Adenoidectomy	47%	(42–52)
Tympanostomy tubes	51%	(46–56)
Other ear-, nose-, and throat surgeries	2%	(0.6–3)
Antibiotic prophylaxis	15%	(11–19)
Ever had speech-therapy	9%	(6–12)

* at inclusion in the study

** reported by the caregiver

*** asthma, wheezing, hayfever, or eczema

Table 4 Floor and ceiling effects*, internal consistency and test–retest reliability of the questionnaires

	Minimum score (%)	Maximum score (%)	Internal consistency Cronbach's α $n = 383$**	Test–retest reliability ICC*** $n = 106$
Generic				
RAND	0	0	0.81	0.89
FSQ generic	0	2	0.80	0.92
FSQ specific	0	21	0.86	0.89
TAIQOL	N.A.	N.A.	0.72–0.90	0.76–0.90
Sleeping	2	12	0.90	0.83
Appetite	0	22	0.86	0.82
Positive mood	0	80	0.90	0.81
Liveliness	0.6	81	0.88	0.76
Problem behaviour	1	4	0.86	0.85
Communication	0.4	53	0.88	0.82
Disease-specific				
OM-6	0	14	0.85	0.89
NRS child	2	3	N.A.	0.83
FFQ	0.5	27	0.90	0.93
NRS caregiver	0	0	N.A.	0.81

* percentage of respondents with minimum (floor effect) and maximum (ceiling effect) scores

** $n = 169$ for the TAIQOL subscales and NRS Caregiver

*** Intra-class Correlation Coefficient

Reliability

Cronbach alpha coefficients were adequate to high (range 0.72–0.90) for the TAIQOL subscales and high (range 0.80–0.90) for all other questionnaires. The calculation of inter-item correlations revealed no 'hidden' subscales or item redundancy (i.e., individual correlations are too high, with possible loss of content validity) (Table 4).

In order to assess test–retest reliability, 126 (79%) of 160 approached caregivers completed a second set of questionnaires of which 113 (71%) were completed within 2 weeks. Seven children with AOM at the time of the outpatient visit (test 1) were excluded, resulting in 106 sets for analysis (Table 4). ICCs were moderate to high for all questionnaires (range 0.81–0.93) and most TAIQOL subscales (range 0.76–0.90), but in the borderline range for the TAIQOL subscale 'Liveliness' (0.76).

Construct and discriminant validity

Table 5 reflects the calculated correlations between the questionnaires, which ranged from moderate to strong for the RAND, FSQ Generic, FSQ Specific, OM-6, and FFQ. These outcomes show that 14 (67%) of the hypothesized correlations were correct. False predictions were mainly made about the NRS Child and NRS Caregiver, as the correlations with other questionnaires were generally expected to be at least moderate, but were found to be weak. Disease-specific questionnaires (OM-6, NRS Child, FFQ and NRS Caregiver), showed moderate correlations (Spearmans' rho 0.39–0.49) with the frequency of AOM episodes in the preceding 6 months. Moderate correlations (Spearmans' rho 0.29–0.48) were also found between global FHS (RAND) and the disease-specific questionnaires on the one hand and the number of physician visits

Table 5 Construct validity: calculated correlations * between the questionnaires**

	RAND	FSQ generic	FSQ specific	OM-6	NRS child	FFQ	NRS caregiver
RAND	1.00	**0.52**	**0.49**	**0.34**	**0.33**	**0.43**	**0.49**
FSQ generic		1.00	**0.80**	**0.37**	0.25	**0.43**	**0.24**
FSQ specific			1.00	**0.49**	0.26	**0.52**	0.24
OM-6				1.00	0.23	**0.74**	0.28
NRS child					1.00	0.22	**0.47**
FFQ						1.00	0.39
NRS caregiver							1.00

* Spearman correlation coefficients were calculated

** appropriately à priori predicted correlations are bold-printed

for all upper respiratory tract infections (URTIs), a more global indicator of illness, on the other hand (Table 6).

The RAND, FSQ Generic, FSQ Specific, OM-6 and FFQ were able to discriminate between children with moderately recurrent AOM (2–3 episodes per year) and "otitis-prone" children with severe, recurrent AOM (4 or more episodes per year) (Table 7). However, neither the two numerical rating scales (NRS Child and NRS Caregiver) nor the otitis media-related subscales of the TAI-QOL discriminated between these two groups.

Responsiveness

According to our external criterion of change (a reduction of 2 or more episodes of AOM per year), 270 children (70%) of 383 were classified as 'changed' for the first interval (0–7 months) and 126 children (33%) for the second interval (7–14 months). The two intervals differed considerably regarding the reduction of AOM incidence; during the 0–7 months follow-up the mean incidence per child decreased by 1.8 AOM episodes, whereas during 7–14 months follow-up the mean decrease was 0.35 episodes [74].

Sensitivity to change

Sensitivity to change, expressed as significant mean change and effect size, is presented in Table 8. Except for most TAIQOL subscales, generic as well as disease-specific questionnaires yielded significant change scores during both follow-up periods, ranging from 4.9 to 28.3 on a 0–100 scale. Absolute change scores for the first follow-up

period generally were larger (range 0.4–28.3) than for the second period (range –2.8–14.2).

The effect sizes for the generic FHS questionnaires ranged from small to moderate (0.29–0.60). For the generic TAIQOL subscales however, the effect sizes were lower, ranging from almost zero for the subscales 'Appetite' (0,0), 'Problem behaviour' (0.02) and 'Positive mood' (0.06) to small for 'Sleeping'(0.37) and 'Liveliness' (0.22). Effect sizes for the disease-specific questionnaires were moderate to large (0.55–0.95). For the questionnaires the ES were quite similar for the first (0–7 months) and second intervals (7–14 months), whereas for the second interval absolute change scores were smaller.

The TAIQOL was excluded from further analyses on the interpretation of change, due to its poor sensitivity to change.

Interpretation of change—distribution-based methods

Minimally clinical important differences (MCIDs) calculated with distribution-based methods are presented in Table 9. During the first interval, ES-MCIDs using an effect size of 0.3 as benchmark were somewhat smaller for generic questionnaires, ranging from 5.0 to 7.4 on a 0–100 scale, than those for disease-specific questionnaires ranging from 6.1 to 9.4. During the second interval, however, ES-MCIDs for generic and disease-specific questionnaires

Table 6 Construct validity—'correlations* between questionnaire scores and frequency of physician visits for URTI** and of AOM** episodes'

	Frequency of physician visits for URTI	Frequency of AOM episodes***
Generic		
RAND	–0.48	–0.31
FSQ generic	–0.20	–0.07[#]
FSQ specific	–0.27	–0.12[##]
Disease-specific		
OM-6	–0.32	–0.41
NRS child	–0.41	–0.49
FFQ	–0.29	–0.39
NRS caregiver	–0.41	–0.40

* Spearmans'rho correlation coefficients were calculated

** URTI: upper respiratory tract infection; AOM: acute otitis media

*** All correlations $P < 0.001$, except for [#] ($P = 0.16$) and [##] ($P = 0.02$)

Table 7 Discriminant validity: scores of children with 2–3 vs. 4 or more AOM episodes in the preceding year*

	2–3 AOM episodes	≥4 AOM episodes	Mann–Whitney P-value
Generic			
RAND	21.1	19.6	0.004
FSQ generic	76.5	72.2	0.002
FSQ specific	83.9	78.4	0.001
TAIQOL			
Sleeping	66.2	60.7	0.10
Appetite	74.7	73.2	0.44
Liveliness	93.2	91.3	0.81
Positive mood	92.0	92.5	0.97
Problem behaviour	64.8	60.9	0.24
Communication	83.8	84.5	0.69
Disease-specific			
OM-6	18.9	17.0	<0.001
NRS child	5.2	5.4	0.48
FFQ	84.9	78.5	<0.001
NRS caregiver	6.6	6.2	0.22

Calculated by Mann–Whitney test

* 2–3 episodes means moderate and >4 episodes means serious AOM

Table 8 Sensitivity to change: mean change-scores* and effect sizes** for changed subjects

	Mean change-score				Effect size—GRS	
	0–7 months[#]		7–14 months		0–7 months	7–14 months
	$n = 270$***	P-value	$n = 126$****	P-value	$n = 270$***	$n = 126$****
Generic						
RAND	10.2	<0.001	7.7	<0.001	0.60	0.54
FSQ Generic	7.0	<0.001	4.9	0.001	0.37	0.29
FSQ specific	9.1	<0.001	6.0	<0.001	0.37	0.32
TAIQOL						
Sleeping	9.9	<0.001	7.1	0.03	0.37	0.36
Appetite	6.8	0.001	0.0	1.0	0.28	0.00
Problem behaviour	0.4	0.80	−2.8	0.33	0.02	0.13
Positive mood	1.5	0.30	3.9	0.11	0.06	0.25
Liveliness	2.3	0.19	1.6	0.51	0.22	0.11
Communication	2.9	0.12	1.7	0.32	0.16	0.11
Disease-specific						
OM-6	16.6	<0.001	11.5	<0.001	0.60	0.73
NRS child	28.3	<0.001	14.2	<0.001	0.91	0.64
FFQ	13.6	<0.001	8.0	<0.001	0.55	0.60
NRS caregiver	19.2	0.003	9.1	0.003	0.95	0.57

* calculated with paired *t*-test

**calculated with Guyatt's responsiveness statistic (GRS)

*** $n = 114$ for TAIQOL subscales and NRS Caregiver; [#] follow-up interval

**** $n = 51$ for TAIQOL subscales and NRS Caregiver

were comparable (range 4.0–6.7), indicating that for both types of questionnaires similar change scores are needed in order to be clinically relevant.

Except for the NRS Child and NRS Caregiver, the SEM-MCIDs were quite comparable with the ES-MCIDs for both generic and disease-specific questionnaires. Assuming that the estimated MCIDs using either an effect size of 0.3 or a one-SEM as benchmark are correct, our results suggest that the range for the distribution-based MCID for generic as well as disease-specific questionnaires corresponds with a change of 3 - 9 points on a 0–100 scale (see Table 9).

Interpretation of change—anchor-based methods

Changes in AOM *frequency* (AOM incidence per child per year) were compared to the magnitude of change scores on the FHS and HRQoL questionnaires. A *small* change of 2 AOM episodes, which is considered a MCID, in AOM frequency corresponded with 3–10 points change on a 0–100 scale for the generic questionnaires (Graph 1a), and with 5–15 points change for disease-specific questionnaires, except for the NRS Child during the 0–7 months interval with 29 points change.

Likewise, a *small* improvement in AOM *severity* corresponded with change scores ranging from 2–10 points on a 0–100 scale for the generic questionnaires and with change scores from 4–8 points for the disease-specific questionnaires, except again for the NRS Child with 16 and 17 points change (Graph 1b).

Change scores corresponding with *moderate to large* changes in AOM frequency and severity are also presented in Graph 1a, b. Comparing small change with moderate to large change shows that, overall, the larger the change in AOM severity or frequency, the larger the magnitude of the change score on the questionnaires. However, this trend was not true for the FSQ Generic and the disease-specific NRS Child (e.g., a small change in AOM severity equalized a change score of 17 on the NRS Child, whereas a moderate-large change equalized a change score of 13).

Comparison of anchor- and distribution-based methods

Comparing the results of the anchor-based methods with those of the distribution-based methods (Graph 2) showed that generic questionnaires (RAND, FSQ Generic, and FSQ Specific), disease-specific questionnaires (OM-6 and FFQ) and the NRS Caregiver yielded quite similar estimates of the MCID for both methods (3–9 points on a 0–100 scale for distribution and 2–15 points for anchor-based methods)

Table 9 Responsiveness—distribution-based indices for minimally clinical important difference (MCID) using 0.3 Effect Size (ES) and one standard error of measurement (SEM)

	ES—MCID*		SEM—MCID**	
	0–7 months[#]	7–14 months	0–7 months	7–14 months
Generic				
RAND	5.0	4.3	5.3	4.5
FSQ generic	5.7	5.1	5.4	4.8
FSQ specific	7.4	5.6	7.8	5.9
Disease-specific				
OM-6	8.3	4.7	8.8	5.0
NRS child	9.4	6.7	12.5	8.9
FFQ	7.4	4.0	6.1	3.3
NRS caregiver	6.1	4.8	8.3	6.6

* MCID using 0.3 effect size as benchmark; [#] follow-up interval

** MCID using one-SEM as benchmark

as well as for both follow-up periods (4–15 points for 0–7 months interval, 2–8 points for 7–14 months interval). Averaging these distribution-based and anchor-based estimates of MCID yields a point-estimate MCID for generic questionnaires of 6.0 (range 2–10) and for disease-specific questionnaires of 7.3 (range 3–15) on a 0–100 scale (excluding the NRS Child, as it had much larger estimates for the MCID).

Discussion

In this study, the reliability and validity of generic as well as disease-specific FHS and HRQoL questionnaires have been assessed in the setting of a RCT concerning children with recurrent AOM. Most generic (RAND, FSQ-Generic and FSQ-Specific) and disease-specific (OM-6 and FFQ) questionnaires showed similar, good to excellent reliability and adequate construct and discriminant validity. Construct validity was poor for the numerical rating scales (NRS Child and NRS Caregiver), and discriminant validity was low to moderate for both NRS and the subscales of the TAIQOL considered to be otitis media-related (Tables 4, 5, 6 and 7).

Generic as well as disease-specific questionnaires proved to be sensitive to change in the incidence of AOM (Table 8). The effect sizes were found to be ranging from small to moderate for both generic and disease-specific questionnaires (Table 8). The MCIDs for generic and disease-specific questionnaires were quite similar in terms of responsiveness (Table 9 and Figure 1 & 2). However, most otitis media-related subscales of the TAIQOL, the only true HRQoL questionnaire, proved insensitive to change.

Reliability and validity

Results on internal consistency and test–retest reliability of the RAND, FSQ Generic, FSQ Specific, TAIQOL and OM-6 found in this study, were comparable with those of previous studies using these questionnaires [14, 41, 42, 51, 52]. The consistency of results across different paediatric populations supports the reliability of these questionnaires. Similar to the poor discriminant validity in this study of the otitis media-related TAIQOL subscales, Fekkes et al. [51] found the TAIQOL subscales 'Problem behaviour', 'Positive mood', and 'Liveliness' discriminated neither between healthy and preterm children nor between healthy and chronically ill children. The ability of the RAND, FSQ Generic and FSQ Specific to discriminate between children who differed in AOM frequency, on the other hand, supported their discriminant validity previously found in children with asthma and healthy children [41, 42]. However, the heterogeneity of methods used limits the comparability of results regarding validity of this study with those from previous studies.

The FFQ and NRS Caregiver are newly composed questionnaires to assess the influence of recurrent AOM on the caregiver and family. The FFQ demonstrated excellent reliability and validity, meeting the minimal required reliability coefficients of 0.90 for individual assessment [65, 87]. The strong correlation with the OM-6 supports its complementary usefulness in FHS and HRQoL assessment in children with rAOM. Results of the NRS Caregiver, however, were similarly poor as those observed for the NRS Child, which needs further exploration. Their global, single-item assessment of HRQoL may be too crude to reflect subtle differences in HRQoL [88, 89]. On the other hand, comments of the caregivers indicated that some of
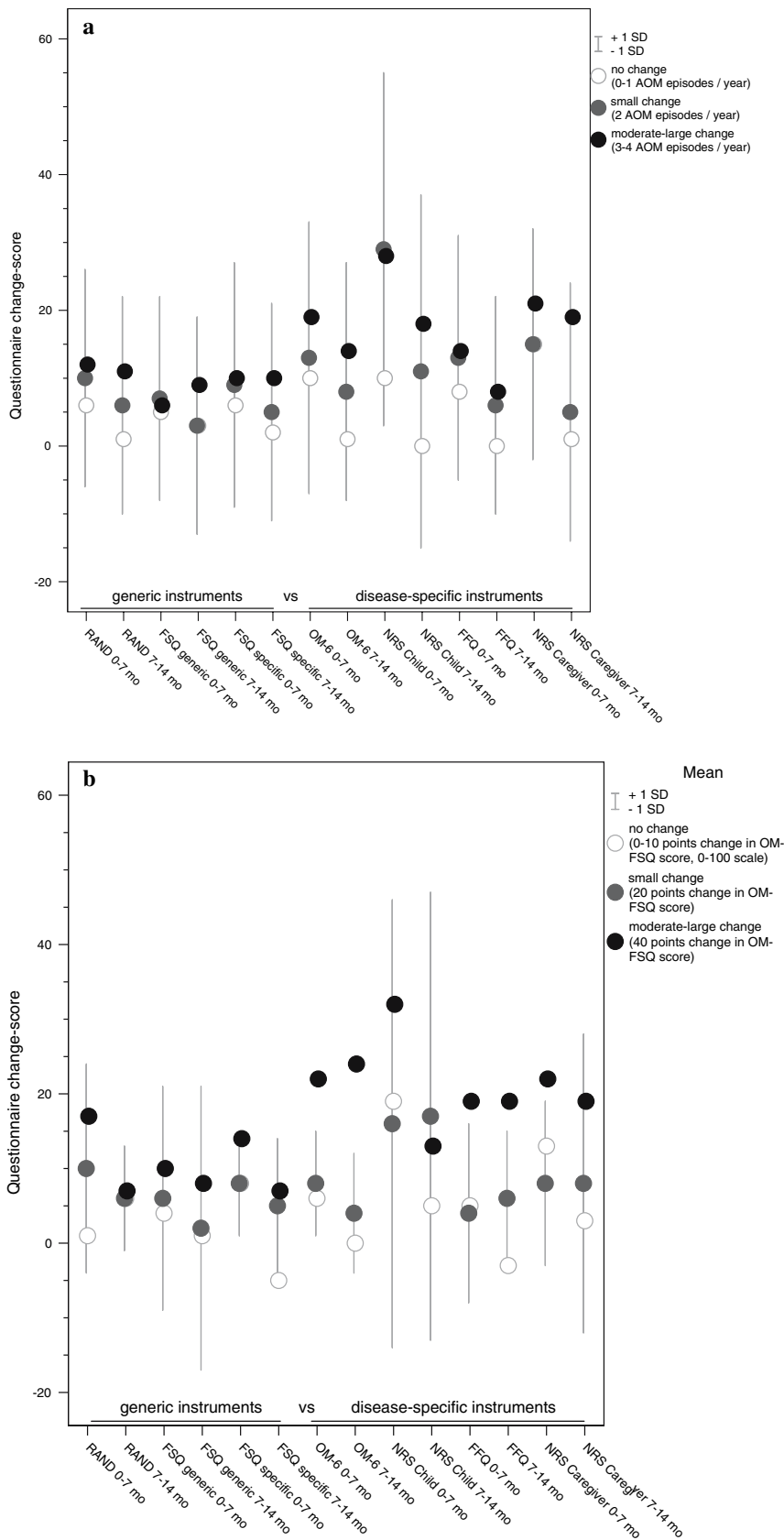
Graph 1 Responsiveness—change-scores per questionnaire corresponding with an anchor-based responsiveness index: (**a**) AOM frequency;
(**b**) AOM severity (OM-FSQ score)
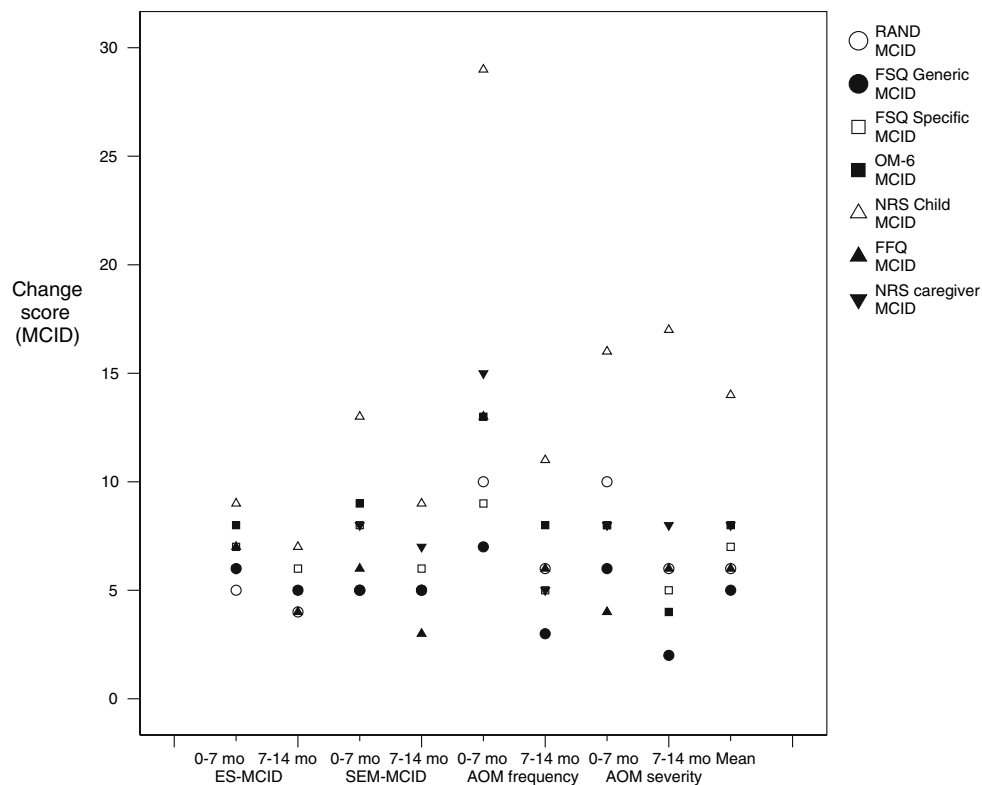
Graph 2 Minimally clinical important difference (MCID) per questionnaire according to distribution-based (ES-MCID and SEM-MCID) and anchor-based (AOM frequency and AOM severity) methods

them may have misunderstood the NRS test-instructions. This is supported by the fact that improvement of construct validity occurred during follow-up assessments, presumably due to learning effects after reading the instructions a second time.

Responsiveness

So far, little attention has been given to the responsiveness of the questionnaires used in our study. Only Rosenfeld et al. [55] assessed effect sizes for the OM-6 (using a standardized response mean) that were much larger (1.1–1.7) than the ones found in this study. This may be explained by the use of different identifiers of change. Rosenfeld et al. [55] used an intervention with expected clinical effectiveness, for which proxies were not blinded, as indicator of change. Since pneumococcal vaccination proved to be clinically ineffective [74], treatment could not be used as an external criterion for change. Instead, a change of 2 or more AOM episodes per year was used as criterion to identify changed subjects. In addition, social desirability and expectancy bias may have influenced the outcome of the study of Rosenfeld et al. [55]

Although clinical criteria such as change in the incidence of AOM episodes have been suggested as adequate

alternative criteria to identify change [34], the choice for any external criterion for change remains somewhat arbitrary. It is a surrogate measure that often only reflects one aspect of the QoL construct. The poor responsiveness of the TAIQOL subscales 'Behavioural problems', 'Positive mood' and 'Liveliness', for example, may indicate that our clinical indicator is less suitable as external criterion for change in emotional and behavioural functioning. However, considering the overall poor responsiveness of the twelve TAIQOL subscales (results not shown), it seems more obvious that poor responsiveness in itself mainly applies for these three subscales as well.

Several studies have supported the empirically found link between one SEM and the MCID for HRQoL questionnaires [75, 81, 85, 86]. In this study the MCIDS based on the value of one-SEM largely corresponded with a MCID that was estimated using 0.3 ES as a benchmark, which is in further support of the one-SEM as an indicator of MCID (Table 9). However, it should be realized that the SEM as well as the ES are both only statistical indicators, which relate change to random (error) variance. Interestingly, the anchor-based methods yielded similar estimates for the MCIDs (Graphs 1a, b, 2), which is in agreement with recent observations that one-SEM equals anchor-based MCID in patients with moderately severe illness

[90]. By applying and comparing multiple methods as well as two evaluation periods, we have not only been able to demonstrate consistency in responsiveness but also to give ranges for minimally clinical important changes instead of point-estimates. As there is no 'golden standard' for the assessment of responsiveness in FHS and HRQoL measurement, a range of scores gives a more realistic reflection of responsiveness than a point-estimate. Point estimates can be misapplied by users who are either unaware of the limited precision of data used for estimating the MCID or who are unaware of the intrinsic limitations of dichotomising what is actually a continuum.

Generic versus disease-specific questionnaires

Although generic questionnaires are generally expected to be less sensitive to differences in FHS or HRQoL than disease-specific questionnaires [19, 37, 91, 92], in this study most disease-specific questionnaires performed only marginally better than the generic questionnaires on the discriminant validity test. Likewise, the responsiveness of generic questionnaires, and their usefulness as measures of outcome in randomized trials has been questioned [21]. Although in some studies generic measures indeed were found to be less responsive to treatment effects than specific measures [93–96], other studies did find comparable responsiveness [97–99]. In this study, only the smaller effect sizes for the FSQ Generic and FSQ Specific may indicate that sensitivity to responsiveness of generic questionnaires is somewhat poorer than that of disease-specific questionnaires. Possibly, this higher sensitivity at the start of the study reflects the higher incidence of symptoms and functional limitations that are specific to AOM, whereas during the study AOM incidence decreases and consequently AOM symptoms become less prominent compared to other health problems. Overall, the generic questionnaires appeared to be as sensitive to clinical change as disease-specific questionnaires, except for the TAIQOL.

For the FSQ Generic and FSQ Specific, but not for the RAND which assesses general health perceptions, sensitivity to differences and change in FHS could be explained by their content, as they include many physical and emotional behaviour items that may be affected by rAOM. The more relevant a questionnaire is to a particular condition, the more sensitive it is likely to be. The sensitivity of the RAND, assessing general health and resistance to illness, may indicate that it meets the perceptions of the caregivers of children with rAOM in thinking that their overall health is worse compared with other children. It also may reflect the significant comorbidity like chronic airway problems and atopic symptoms in the study population (Table 3).

The reasons for the poor performance of the TAIQOL with regard to both discriminant validity and sensitivity to change are not obvious. Possibly the subscale scores represent each an aspect of HRQoL that is too limited to be sensitive to differences or change. Combining the subscales to more comprehensive constructs may then improve sensitivity. In addition, each item of the TAIQOL consists of two questions; a question about FHS is followed by the request to rate the child's well-being in relation to this health status. Response shift bias may have modified the caregivers' expectations about how their child feels in line with the child's changing health, that is caregivers may rate their child's well-being as better than it actually is as they adapt to the situation. Studies on factors that may influence sensitivity to change or responsiveness besides the type of questionnaire (generic versus disease-specific), such as questionnaire structure and content, disease severity, comorbidity and other population characteristics, are needed.

Bias and generalisibility

There are several issues that need to be considered when interpreting the current results. First, frequency of AOM episodes at enrolment was based on proxy report, whereas during the trial only physician-diagnosed episodes were counted. The number of AOM episodes in the year prior to inclusion is likely to be overestimated by proxies [100], resulting in the underestimation of HRQoL change scores because they may have evaluated the situation as worse than it objectively was in the first place. However, if such a recall-bias regarding AOM frequency was in fact present, it may also have influenced caregivers' reflection on subjective measures such as FHS and HRQoL, which results in realistic or even overestimated change scores. However, estimating responsiveness for the interval of 7–14 months, in which AOM frequency was not affected by recall bias since al episodes were physician diagnosed, yielded similar results. This indicates that recall bias appears not to have influenced responsiveness substantially.

Secondly, in assessing test–retest reliability, two different modes of questionnaire administration were used: completion at the clinic versus home completion. The possible intention to give more socially desirable answers at the clinic as well as other effects such as being more distracted when filling in the questionnaires at home, may have caused differences in questionnaire scores between the first (test) and second (retest) assessment. Although this impact may be larger for single item questionnaires such as the NRSs compared to multiple item questionnaires, and might explain their somewhat smaller ICCs, the impact on the ICCs appears to be small.

Thirdly, during the trial, 8 children (4.2%) in the pneumococcal vaccine group and 13 (6.7%) in the control

vaccine group were lost to follow-up. One child switched from the control to the pneumococcal vaccine group. It is unlikely that these small numbers of dropouts and crossovers influenced the trial results.

Furthermore, indices of validity and reliability are not fixed characteristics of FHS and HRQoL questionnaires but are influenced by the study design, intervention, and study population in particular. Our study population had significantly severe ear disease with frequent episodes and was older than the average child with AOM. Assessment of reliability and validity of the questionnaires in populations with less severe disease may present more ceiling effects and lack of discriminant validity. Therefore, the results of this study should only be generalized to paediatric populations with moderately to seriously severe recurrent acute ear-infections at an older age (approximately 14–54 months).

Finally, of all questionnaires in this study, only the FFQ demonstrated a reliability that meets the minimal required reliability coefficients for individual assessment of HRQoL. Although some authors suggest to use FHS and HRQoL questionnaires for individual assessment in clinical practice as well [31], we do not support this approach. It is suggested that routine use of these questionnaires would facilitate detection and discussion of psychological issues and help guide decisions regarding, for example, referral. However, considering the complexity and many pitfalls of reproducibility and responsiveness assessment, individual use of HRQoL and FHS questionnaires as part of the follow-up of individuals is not reliable nor valid.

Recommendations for clinical use

In conclusion, generic (RAND, FSQ Generic and FSQ Specific) as well as disease-specific (OM-6, FFQ, and, to a lesser extent, NRS Caregiver) questionnaires demonstrated similar and high reliability and adequate construct and discriminant validity as well as responsiveness to justify use in clinical studies of children with rAOM. However, NRS as used in this study may be less adequate for assessment of HRQoL in this population. The TAIQOL, the only true generic HRQoL questionnaire, unfortunately showed a poor discriminant validity and sensitivity to change, needing extensive revision before further use in clinical outcome studies in children with otitis media. Using both a generic questionnaire (RAND or FSQ) and the OM-6 in clinical studies regarding FHS in children with rAOM is recommended, as it would combine the merits of both generalisability and sensitivity in outcome assessment and facilitate head-to-head comparisons of their performance in various paediatric populations with OM.

More studies are needed assessing responsiveness of paediatric QoL questionnaires by multiple, distribution as well as anchor-based, methods to increase our appreciation of minimal clinically important changes in various paediatric conditions. Further studies on factors such as questionnaire structure and content, disease severity, comorbidity and other population characteristics that may influence sensitivity to change or responsiveness besides the type of questionnaire (generic versus disease-specific) may increase our appreciation of the complex dynamics in HRQoL and FHS assessment.

References

1. Teele, D. W., Klein, J. O., & Rosner, B. (1989). Epidemiology of otitis media during the first seven years of life in children in greater Boston: A prospective, cohort study. *The Journal of Infectious Diseases, 160*, 83–94.
2. Alho, O. P., Koivu, M., & Sorri, M. (1991). What is an 'otitis-prone' child? *International Journal of Pediatric Otorhinolaryngology, 21*, 201–209.
3. Alho, O. P. (1997). How common is recurrent acute otitis media? *Acta Oto-Laryngologica. Supplementum, 529*, 8–10.
4. Kilpi, T., Herva, E., Kaijalainen, T., Syrjanen, R., & Takala, A. K. (2001). Bacteriology of acute otitis media in a cohort of Finnish children followed for the first two years of life. *The Pediatric Infectious Disease Journal, 20*, 654–662.
5. Niemela, M., Uhari, M., Jounio-Ervasti, K., Luotonen, J., Alho, O. P., & Vierimaa, E. (1994). Lack of specific symptomatology in children with acute otitis media. *Pediatric Infectious Disease Journal, 13*, 765–768.
6. Ruuskanen, O., & Heikkinen, T. (1994). Otitis media: Etiology and diagnosis. *Pediatric Infectious Disease Journal, 13*, S23–S26.
7. Heikkinen, T., & Ruuskanen, O. (1995). Signs and symptoms predicting acute otitis media. *Archives of Pediatrics & Adolescent Medicine, 149*, 26–29.
8. Kontiokari, T., Koivunen, P., Niemela, M., Pokka, T., & Uhari, M. (1998). Symptoms of acute otitis media. *Pediatric Infectious Disease Journal, 17*, 676–679.
9. Gravel, J. S., & Wallace, I. F. (1998). Language, speech, and educational outcomes of otitis media. *The Journal of Otolaryngology, 27*(Suppl. 2), 17–25.
10. Paradise, J. L. (1998). Otitis media and child development: Should we worry? *Pediatric Infectious Disease Journal, 17*, 1076–1083.
11. Bennett, K. E., & Haggard, M. P. (1999). Behaviour and cognitive outcomes from middle ear disease. *Archives of Disease in Childhood, 80*, 28–35.
12. Johnson, D. L., Swank, P. R., Owen, M. J., Baldwin, C. D., Howie, V. M., McCormick, D. P. (2000). Effects of early middle ear effusion on child intelligence at three, five, and seven years of age. *Journal of Pediatric Psychology, 25*, 5–13.
13. Paradise, J. L., Dollaghan, C. A., Campbell, T. F., Feldman, H. M., Bernard, B. S., Colborn, D. K., et al. (2000). Language, speech sound production, and cognition in three-year-old children in relation to otitis media in their first three years of life. *Pediatrics, 105*, 1119–1130.
14. Rosenfeld, R. M., Goldsmith, A. J., Tetlus, L., & Balzano, A. (1997). Quality of life for children with otitis media. *Archives of Otolaryngology—Head & Neck Surgery, 123*, 1049–1054.
15. Asmussen, L., Olson, L. M., & Sullivan, S. A. (1999). 'You have to live it to understand it...'—Family experiences with chronic otitis media in children. *Ambulatory Child Health, 5*, 303–312.

16. Curry, M. D., Mathews, H. F., Daniel, H. J. III, Johnson, J. C., & Mansfield, C. J. (2002). Beliefs about and responses to childhood ear infections: a study of parents in eastern North Carolina. *Social Science & Medicine, 54*, 1153–1165.

17. Brouwer, C. N. M., Maillé, A. R., Rovers, M. M., Grobbee, D. E., Sanders, E. A. M., & Schilder, A. G. M. (2005). Health-related quality of life in children with otitis media. *International Journal of Pediatric Otorhinolaryngology, 69.8.*, 1031–1041.

18. Eiser, C. (1997). Children's quality of life measures. *Archives of Disease in Childhood, 77*, 350–354.

19. Jenney, M. E., & Campbell, S. (1997). Measuring quality of life. *Archives of Disease in Childhood, 77*, 347–350.

20. Theunissen, N. C., Vogels, T. G., Koopman, H. M., Verrips, G. H., Zwinderman, K. A., Verloove-Vanhorick, S. P., et al. (1998). The proxy problem: Child report versus parent report in health-related quality of life research. *Quality of Life Research , 7*, 387–397.

21. Gill, T. M., & Feinstein, A. R. (1994). A critical appraisal of the quality of quality-of-life measurements. *The Journal of the American Medical Association, 272*, 619–626.

22. Schipper, H., Clinch, J. J., & Olweny, C. L. M. (1996). Quality of life studies: Definitions and conceptual issues. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 11–23). Philadelphia, USA: Lippincot-Raven Publishers.

23. Bergner, M. (1989). Quality of life, health status, and clinical research. *Medical care, 27*, S148–S156.

24. Bullinger, M., & Ravens-Sieberer, U. (1995). Health related quality of life assessment in children: A review of the literature. *Revue Européenne de Psychologie Appliquée, 45*(4), 245–254.

25. Muldoon, M. F., Barger, S. D., Flory, J. D., & Manuck, S. B. (1998). What are quality of life measurements measuring? *British Medical Journal, 316*, 542–545.

26. Feldman, B. M., Grundland, B., McCullough, L., & Wright, V. (2000). Distinction of quality of life, health related quality of life, and health status in children referred for rheumatologic care. *The Journal of Rheumatology, 27*, 226–233.

27. Eiser, C., & Morse, R. (2001). Quality-of-life measures in chronic diseases of childhood. *Health Technology Assessment, 5*(4), 1–157.

28. Schmidt, L. J., Garrat, A. M., & Fitzpatrick, R. (2002). Child/parent population health outcomes: A structural review. *Child: Care, Health and Development, 28*(3), 227–237.

29. Cremeens, J., Eiser, C., & Blaas, M. (2006). Characteristics of health-related self-report measures for children aged three to eight year: A review of the literature. *Quality of Life Research , 15*, 739–754.

30. Eiser, C., & Jenney, M. (2007). Measuring quality of life. *Archives of Disease in Childhood, 92*, 348–350.

31. de Wit, M., Delemarre-van de Waal, H. A., Pouwer, F., Gemke, R. J. B. J., & Snoek, F. J. (2007). Monitoring health-related quality of life in adolescents with diabetes: A review of measures. *Archives of Disease in Childhood, 92*, 434–439.

32. Guyatt, G. H., Kirshner, B., & Jaeschke, R. (1992). Measuring health status: What are the necessary measurement properties? *Journal of Clinical Epidemiology, 45*, 1341–1345.

33. Patrick, D. L., & Chiang, Y. P. (2000). Measurement of health outcomes in treatment effectiveness evaluations: Conceptual and methodological challenges. *Medical Care, 38*, II14–II25.

34. Guyatt, G. H., Osoba, D., Wu, A. W., Wyrwich, K. W., & Norman, G. R. (2002). Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings, 77*, 371–383.

35. Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life questionnaires: Attributes and review criteria. *Quality of Life Research, 11*, 193–205.

36. Samsa, G., Edelman, D., Rothman, M. L., Williams, G. R., Lipscomb, J., & Matchar, D. (1999). Determining clinically important differences in health status measures: A general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics, 15*, 141–155.

37. Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical care, 27*, S217–S232.

38. Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine, 118*, 622–629.

39. Haggard, M. P., & Smith, S. C. (1999). Impact of otitis media on child quality of life. In R. M. Rosenfeld & C. D.Bluestone (Ed.), *Evidence-based otitis media* (pp. 375–399). Hamilton, Ontario: B.C. Becker Inc.

40. Guyatt, G. H., King, D. R., Feeny, D. H., Stubbing, D., & Goldstein, R. S. (1999). Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. *Journal of Clinical Epidemiology, 52*, 187–192.

41. Post, M. W., Kuyvenhoven, M. M., Verheij, M. J., de Melker, R. A., & Hoes, A. W. (1998). The Dutch version of 'Functional Status II(R)': A questionnaire measuring the functional health status of children. *Nederlands Tijdschrift Voor Geneeskunde, 142*, 2675–2679.

42. Post, M. W., Kuyvenhoven, M. M., Verheij, M. J., de Melker, R. A., & Hoes, A. W. (1998). The Dutch 'Rand General Health Rating Index for Children': A questionnaire measuring the general health status of children. *Nederlands Tijdschrift Voor Geneeskunde, 142*, 2680–2683.

43. Lewis, C. C., Pantell, R. H., & Kieckhefer, G. M. (1989). Assessment of children's health status. Field test of new approaches. *Medical Care, 27*, S54–S65.

44. Tebbi, C. K., Bromberg, C., & Piedmonte, M. (1989). Long-term vocational adjustment of cancer patients diagnosed during adolescence. *Cancer, 63*, 213–218.

45. Olson, A. L., Boyle, W. E., Evans, M. W., & Zug, L. A. (1993). Overall function in rural childhood cancer survivors. The role of social competence and emotional health. *Clinical Pediatrics (Philadelphia), 32*, 334–342.

46. Rosier, M. J., Bishop, J., Nolan, T., Robertson, C. F., Carlin, J. B., & Phelan, P. D. (1994). Measurement of functional severity of asthma in children. *American Journal of Respiratory and Critical Care Medicine, 149*, 1434–1441.

47. Scholle, S. H., Whiteside, L., Kelleher, K., Bradley, R, & Casey, P. (1995). Health status of preterm low-birthweight infants. Comparison of maternal reports. *Archives of Pediatrics & Adolescent Medicine, 149*, 1351–1357.

48. McCormick, M. C., Workman-Daniels, K., & Brooks-Gunn, J. (1996). The behavioral and emotional wellbeing of school-age children with different birth weights. *Pediatrics, 97*, 18–25.

49. Mahajan, P., Pearlman, D., & Okamoto, L. (1998). The effect of fluticasone propionate on functional status and sleep in children with asthma and on the quality of life of their parents. *The Journal of Allergy and Clinical Immunology, 102*, 19–23.

50. Sawyer, M., Antoniou, G., Toogood, I., & Rice, M. (1999). A comparison of parent and adolescent reports describing the health-related quality of life of adolescents treated for cancer. *International Journal of Cancer. Supplement, 12*, 39–45.

51. Fekkes, M., Theunissen, N. C., Brugman, E., Veen, S., Verrips, E. G., & Koopman, H. M., et al. (2000). Development and psychometric evaluation of the TAPQOL: A health-related quality of life questionnaire for 1–5-year-old children. *Quality of Life Research , 9*, 961–972.

52. Alsarraf, R., Jung, C. J., Perkins, J., Crowley, C., & Gates, G. A. (1998). Otitis media health status evaluation: A pilot study for the investigation of cost-effective outcomes of recurrent acute otitis media treatment. *The Annals of Otology, Rhinology, and Laryngology, 107*, 120–128.

53. TNO—Prevention and Health/LUMC. *TAIQOL—Questionnaire for parents of children aged 1—5 years.* (1997). Leiden, The Netherlands : Leiden University Medical Center.

54. Rovers, M. M., Krabbe, P. F., Straatman, H, Ingels, K, van der Wilt, G. J., & Zielhuis, G. A. (2001) Randomised controlled trial of the effect of ventilation tubes (grommets) on quality of life at age 1–2 years. *Archives of Disease in Childhood, 84*, 45–49.

55. Rosenfeld, R. M., Bhaya M. H., Bower C. M., Brookhouser P. E., Casselbrant M. L., Chan K. H., et al. (2000). Impact of tympanostomy tubes on child quality of life. *Archives of Otolaryngology—Head & Neck Surgery, 126*, 585–592.

56. Timmerman, A. A., Anteunis, L. J., & Meesters, C. M. (2003). Response-shift bias and parent-reported quality of life in children with otitis media. *Archives of Otolaryngology—Head & Neck Surgery, 129*, 987–991.

57. Kubba, H., Swan, I. R., & Gatehouse, S. (2004). How appropriate is the OM6 as a discriminative questionnaire in children with otitis media? *Archives of Otolaryngology—Head & Neck Surgery, 130*, 705–709.

58. Bullinger, M., Anderson, R., Cella, D., & Aaronson, N. (1993). Developing and evaluating cross-cultural questionnaires from minimum requirements to optimal models. *Quality of Life Research , 2*, 451–459.

59. Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*, 1417–1432.

60. Guyatt, G. H. (1993). The philosophy of health-related quality of life translation. *Quality of Life Research , 2*, 461–465.

61. Bullinger, M., Alonso, J., Apolone, G., Leplege, A., Sullivan, M., & Wood-Dauphinee, S., et al. (1998). Translating health status questionnaires and evaluating their quality: The IQOLA Project approach. International Quality of Life Assessment. *Journal of Clinical Epidemiology, 51*, 913–923.

62. Asmussen, L., Sullivan, S. A., Olson, L. M., & Flemming, G. V. (1996). The "Ear Infection Survey": A condition-specific functional outcomes measure for families of children with chronic otitis media. AHSR FHSR Annu Meet Abstr Book. 1996;13:14. [confer. proceeding].

63. McColl, E, Eccles, M. P., Rousseau, N. S., Steen, I. N., Parkin, D. W., & Grimshaw, J. M. (2003). From the generic to the condition-specific? Questionnaire order effects in quality of life assessment. *Medical Care, 41*(7), 777–790.

64. Cheung, Y.-B., Wong, L.-C., Tay, M.-H., Toh, C.-K., Koo, W.-H., Epstein, R, et al. (2004). Order effects in the assessment of quality of life in cancer patients. *Quality of Life Research , 13*, 1217–1223.

65. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric testing.* New York, U.S.A.: McGraw-Hill.

66. Streiner, D. L., & Norman, G. R. (1995). Health measurement scales—A practical guide to their development and use. New York, U.S.A.: Oxford University Press.

67. Pedhazur, S. (1991). *Measurement, design, and analysis.* Hillsdale, New Jersey: Erlbaum.

68. Cohen, J. (1988). The significance of a product moment r. In *Statistical power analysis for the behavioral sciences* (pp. 75–107). Hillsdale, New Jersey: Erblaum.

69. Jero, J., & Karma, P. (1997). Prognosis of acute otitis media. Factors associated with the development of recurrent acute otitis media. *Acta oto-Laryngologica. Supplementum, 529*, 30–33.

70. Hotomi, M, Yamanaka, N, Saito, T, Shimada, J., Suzumoto, M., Suetake, M., et al. (1999). Antibody responses to the outer membrane protein P6 of non-typeable Haemophilus influenzae and pneumococcal capsular polysaccharides in otitis-prone children. *Acta Oto-Laryngologica, 119*, 703–707.

71. Dhooge, I. J., van Kempen, M. J., Sanders, L. A., & Rijkers, G. T. (2002). Deficient IgA and IgG2 antipneumococcal antibody levels and response to vaccination in otitis prone children. *International Journal of Pediatric Otorhinolaryngology, 64*, 133–141.

72. Stenstrom, C., & Ingvarsson, L. (1994). General illness and need of medical care in otitis prone children. *International Journal of Pediatric Otorhinolaryngology, 29*, 23–32.

73. Stenstrom, C., & Ingvarsson, L. (1997). Otitis-prone children and controls: A study of possible predisposing factors. 2. Physical findings, frequency of illness, allergy, day care and parental smoking. *Acta Oto-Laryngologica, 117*, 696–703.

74. Veenhoven, R., Bogaert, D., Uiterwaal, C., Brouwer, C., Kiezebrink, H., Bruin, J., Ijzerman, E., Hermans, P., de Groot, R., Zegers, B., Kuis, W., Rijkers, G., Schilder, A., & Sanders, E. (2003). Effect of conjugate pneumococcal vaccine followed by polysaccharide pneumococcal vaccine on recurrent acute otitis media: A randomised study. *Lancet, 361*(9376), 2189–2195.

75. Norman, G. R., Sridhar, F. G., Guyatt, G. H., & Walter, S. D. (2001). Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Medical Care, 39*, 1039–1047.

76. Terwee, C., Dekker, F., & Bossuyt, P. (2002). A taxonomy for responsiveness? *Journal of Clinical Epidemiology, 55*, 1156.

77. Van Stel, H. F., Maillé, A. R., Colland, V. T., & Everaerd, W. T. A. M. (2003). Interpretation of change and longitudinal validity of the Quality of Life for Respiratory Illness Questionnaire (QoLRIQ) in inpatient pulmonary rehabilitation. *Quality of Life Research, 12*, 133–145.

78. 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*, 142S–58S.

79. Cohen, J. (1988) The *t*-test for means. In: *Statistical power analysis for the behavioural sciences* (pp. 19–74). Hillsdale, New Jersey: Erblaum.

80. Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10*, 407–415.

81. Wyrwich, K. W, Tierney, W. M., & Wolinsky, F. D. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology, 52*, 861–873.

82. Cella, D., Eton, D. T., Fairclough, D. L., Bonomi, P., Heyes, A. E., Silberman, C., et al. (2002). What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACTL) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *Journal of Clinical Epidemiology, 55*, 285–295.

83. Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41*(5), 582–92.

84. Pfennings, L. E., van der Ploeg, H. M., Cohen, L., & Polman, C. H. (1999). A comparison of responsiveness indices in multiple sclerosis patients. *Quality of Life Research , 8*, 481–489.

85. Wyrwich, K. W., Nienaber, N. A., Tierney, W. M., & Wolinsky, F. D. (1999). Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Medical Care, 37*, 469–478.

86. Wyrwich, K. W., Tierney, W. M., & Wolinsky, F. D. (2002). Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. *Quality of life research , 11*, 1–7.

87. Weiner, E. A., & Stewart, B. J. (1984). Correlation and reliability. In *Assessing individuals—psychological and educational tests and measurements* (pp. 47–70). Toronto, Canada: Little, Brown and Company.

88. Bowling, A. (1995) Comments on measurement issues and sources of information. In *Measuring disease—A review of disease-specific quality of life measurement scales* (pp. 286–297). Buckingham, U.K.: Open University Press.

89. Wu, A. W., Jacobson, K. L., Frick, K. D., Clark, R., Revicki, D. A., Freedberg, K. A., et al. (2002). Validity and responsiveness of the euroqol as a measure of health-related quality of life in people enrolled in an AIDS clinical trial. *Quality of Life Research, 11*, 273–282.

90. de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes, 4*, 54.

91. Wolinsky, F. D., Wyrwich, K. W., Nienaber, N. A., & Tierney, W. M. (1998). Generic versus disease-specific health status measures. An example using coronary artery disease and congestive heart failure patients. *Evaluation & the Health Professions, 21*, 216–243.

92. Guyatt, G. H., Naylor, C. D., Juniper, E., Heyland, D. K, Jaeschke, R., & Cook, D. J. (1997). Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-based medicine working group. *The Journal Of The American Medical Association, 277*, 1232–1237.

93. Wright, J. G., & Young, N. L. (1997). A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology, 50*, 239–246.

94. Ware, J. E. Jr., Kemp, J. P., Buchner, D. A., Singer, A. E., Nolop, K. B., & Goss, T. F. (1998). The responsiveness of disease-specific and generic health measures to changes in the severity of asthma among adults. *Quality of Life Research , 7*, 235–244.

95. Bessette, L., Sangha, O., Kuntz, K. M., Keller, R. B., Lew, R. A., Fossel, A. H., et al. (1998). Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. *Medical Care, 36*, 491–502.

96. Salaffi, F., Stancati, A., & Carotti, M. (2002). Responsiveness of health status measures and utility-based methods in patients with rheumatoid arthritis. *Clinical Rheumatology, 21*, 478–487.

97. Varni, J. W., Seid, M., Smith, K. T., Burwinkle, T., Brown, J., & Szer, I. S. (2002). The PedsQL in pediatric rheumatology: Reliability, validity, and responsiveness of the pediatric quality of life inventory generic core scales and rheumatology module. *Arthritis and Rheumatism, 46*, 714–725.

98. Eberhardt, K., Duckberg, S., Larsson, B. M., Johnson, P. M., & Nived, K. (2002). Measuring health related quality of life in patients with rheumatoid arthritis-reliability, validity, and responsiveness of a Swedish version of RAQoL. *Scandinavian Journal of Rheumatology, 31*, 6–12.

99. Tsukino, M., Nishimura, K., McKenna, S. P., Ikeda, A., Hajiro, T., Zhang, M., et al. (2002). Change in generic and disease-specific health-related quality of life during a one-year period in patients with newly detected chronic obstructive pulmonary disease. *Respiration, 69*, 513–520.

100. Alho, O. P. (1990). The validity of questionnaire reports of a history of acute otitis media. *American Journal of Epidemiology, 132*, 1164–1170.