

Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version

Hein Raat · Resiti T. Mangunkusumo ·
Jeanne M. Landgraf · Gitte Kloek · Johannes Brug

Received: 14 July 2006 / Accepted: 9 December 2006 / Published online: 8 February 2007
© Springer Science+Business Media B.V. 2007

Abstract

Aims In this study we evaluated indicators of the feasibility, reliability, and validity of the Child Health Questionnaire-Child Form (CHQ-CF). We compared the results in a subgroup of adolescents who completed the standard paper version of the CHQ-CF with the results in another subgroup of adolescents who completed an internet version, i.e., an online, web-based CHQ-CF questionnaire.

Methods Under supervision at school, 1,071 adolescents were randomized to complete the CHQ-CF and items on chronic conditions by a paper questionnaire or by an internet administered questionnaire.

Results The participation rate was 87%; age range 13–17 years. The internet administration resulted in fewer missing answers. All but one multi-item scale showed internal consistency reliability (Cronbach's $\alpha > 0.70$). All scales clearly discriminated between adolescents with no, a few, or many self-reported chronic conditions. The paper administration resulted in statistically significant, higher scores on 4 of 10 CHQ-CF scales compared with the internet administration ($P < 0.05$), but Cohen's effect sizes d were ≤ 0.21 . Mode of administration interacted significantly with age ($P < 0.05$) on four CHQ-CF scales, but Cohen's effect sizes for these differences were also ≤ 0.21 .

Conclusion This study supports the feasibility, internal consistency reliability of the scales, and construct validity of the CHQ-CF administered by either a paper questionnaire or online questionnaire. Given Cohen's suggested guidelines for the interpretation of effect sizes, i.e., 0.20–0.50 indicates a small effect, differences in CHQ-CF scale scores between paper and internet administration can be considered as negligible or small.

Keywords Health status measurement · Health-related quality of life · Adolescents · Feasibility · Reliability · Validity · Online questionnaire · Internet questionnaire · Web-based questionnaire · Child Health Questionnaire Child Form 87 items (CHQ-CF87) · Reference / norm scores

Introduction

During the past two decades, several measures have become available to describe generic health-related quality of life in pediatrics, but adolescent self-report questionnaires received relatively little attention until now [1, 2]. The Child Health Questionnaire (CHQ) is one of the most widely used pediatric health-related quality of life measures and has been translated into 21 languages (32 countries). There is a form for parents and also a self-report form for adolescents, the Child Health Questionnaire-Child Form (CHQ-CF) [2–7]. The CHQ covers physical and psychosocial aspects of health, and includes the impact of child health problems or handicaps on family life [3]. This study focuses on the evaluation of missing answers at the item level, distribution of the scale scores, reliability, and validity of the CHQ-CF in an adolescent population.

H. Raat (✉) · R. T. Mangunkusumo · G. Kloek · J. Brug
Department of Public Health, Erasmus MC – University
Medical Center Rotterdam, P.O. Box 2040, 3000 CA
Rotterdam, The Netherlands
e-mail: h.raat@erasmusmc.nl

J. M. Landgraf
HealthActCHQ, Boston, MA, USA

We expect the commonly used paper format of the CHQ and other health questionnaires to be increasingly replaced by internet versions, especially in adolescent populations that are accustomed to the use of computers and the internet [8]. From the perspective of clinicians and researchers, the advantages of using the internet include avoiding paper work, on-line data-entry, and procedures designed to reduce the number of missing answers and the length of questionnaires [9, 10].

In general, the mode of questionnaire administration (e.g., written questionnaire, face to face interview, telephone interview, computer questionnaire) may affect the participation rate, number of missing answers, psychometric properties, and actual scores [11–14]. With regard to health questionnaires, several studies demonstrated some differences between the commonly used paper versions and computer versions of the same questionnaires [15–17]. Especially in studies comparing paper and computer questionnaires on sensitive topics, administration via computer was found to increase reporting of e.g., drug use or unsafe sexual behaviors, as this medium is apparently perceived as providing more privacy than a paper form [18–20].

With regard to online, i.e., internet or web-based administration of health questionnaires, several studies have demonstrated that online health questionnaires are feasible in various settings, especially among adolescents [21, 22]. However, very few randomized studies have evaluated whether psychometric properties and scores differ between the paper and the internet mode of administration of the same health questionnaire [23–26].

In this study, we compared indicators of the feasibility, reliability, and validity of the CHQ-CF in a subgroup of adolescents who completed the standard paper version of the CHQ-CF with the same indicators in another subgroup of adolescents who completed a newly developed internet version of the questionnaire. Additionally, we compared the mean CHQ-CF scores and distributions of the scale scores between both subgroups. A randomized parallel group design was applied in a large adolescent population (13–17 years old), ensuring that both subgroups were comparable.

The study assessed and compared the paper and internet mode of CHQ-CF administration with regard to the following indicators:

- (a) the number of missing answers (indicator of feasibility),
- (b) the distribution of the scale scores including mean scale scores in the whole sample and in gender and age specific subgroups,
- (c) the internal consistency reliability of multi-item scales (indicator of reliability),

- (d) the ability of the CHQ-CF to discriminate between subgroups with and without self-reported chronic conditions (indicator of construct validity).

Methods

Study population

In 2003, 1,071 students in 55 classes of various educational levels in the 3rd year of seven secondary schools (13–17 years old) in the area of Vlaardingen (metropolitan area) and Harderwijk (rural area), The Netherlands, were invited to complete the Child Health Questionnaire Child Form (CHQ-CF). The parents and students each received written information about the study several weeks before data collection; parents could refuse their child's participation, and participation by the students was voluntary.

Data collection

The CHQ-CF consists of 87 items with 4, 5, or 6 response options divided over 10 multi-item scales and two single-item scales (Table 1) [3]. To reduce respondent burden, the item “change-in-health” was not fielded in this study, and the CHQ-CF scales “role functioning-emotional” and “role functioning-behavioral” were combined into a single scale. The combination of the two role functioning scales is a departure from the CHQ-CF instructions that makes the test analogous to the parent form of the CHQ in this regard [3]. For each scale, items were summed up (some recoded/recalibrated) and transformed into a 0 (worst possible score) to 100 (best possible score) scale [3]. Items on standard socio-demographic variables and the prevalence of seven chronic conditions were included in the questionnaire. From the conventional paper format, using the same wording of the items and instructions, an internet version of the questionnaire was developed through a generic internet tool using PHP (4.0.1), MySQL (3.22), and JavaScript (1.3) [27]. The internet version of the questionnaire listed the items of each CHQ-CF scale on a separate web-page. The internet version did not allow the respondent to select more than one answer to each item of the CHQ-CF and it checked the questionnaire for missing answers before the respondent could “logout”. If one or more of the items were not answered, the internet version prompted the respondent to go back to complete those items; but, if the user failed to “logout” properly, missing answers would remain.

Table 1 CHQ-CF scales, items per scale, and interpretation of low and high scores^a

CHQ-CF Scales	Number of items	Description low score	Description high score
Physical functioning (PF)	9	Child is greatly limited in performing all physical activities, including self-care, due to health	Child performs all types of physical activities, including the most vigorous without limitations due to health
Role functioning: Emotional (RE) ^b	3	Child is greatly limited in school work or activities with friends as a result of emotional problems	Child has no limitations in schoolwork or activities with friends as a result of emotional problem
Role functioning: Behavioral (RB) ^b	3	Child is greatly limited in school work or activities with friends as a result of behavior problems	Child has no limitations in schoolwork or activities with friends as a result of behavior problems
Role functioning: Physical (RP)	3	Child is greatly limited in school work or activities with friends as a result of physical health	Child has no limitations in schoolwork or activities with friends as a result of physical health
Bodily pain (BP)	2	Child has extremely severe, frequent and limiting bodily pain	Child has no pain or limitations due to pain
General behavior (BE)	17	Child very often exhibits aggressive, immature, delinquent behavior	Child never exhibits aggressive, immature, delinquent behavior
Mental health (MH)	16	Child has feelings of anxiety and depression all of the time	Child feels peaceful, happy, and calm all of the time
Self esteem (SE)	14	Child is very dissatisfied with abilities, looks, family/peer relationships and life overall	Child is very satisfied with looks, family/peer relationships and life abilities, overall
General health perceptions (GH)	12	Child believes his/her health is poor and likely to get worse	Child believes his/her health is excellent and will continue to be so
Change in health (CH) ^c	1	Child's health is much worse now than 1 year ago	Child's health is much better now than 1 year ago
Family activities (FA)	6	The child's health very often limits and interrupts family activities or is a source of family tension.	The child's health never limits or interrupts family activities nor is family a source of tension
Family cohesion (FC)	1	Family's ability to get along is rated "poor"	Family's ability to get along is rated "excellent"

^a Reproduced with permission [3]

^b The CHQ-CF scales "Role functioning-emotional" and "Role functioning-behavioral" were merged into a single scale "Role functioning-emotional/behavioral" (REB) in this study

^c This single-item scale was not fielded in this study

Randomization

Within each school class, students were randomly assigned to either the paper or the internet mode of administration using SPSS-generated random numbers. Students completed the questionnaires, either on paper or online in a classroom with computers linked to the internet, under the supervision of a research assistant; the students were allowed adequate privacy.

Analysis

Preparatory secondary vocational education was labelled as "lower secondary education"; secondary schools that prepare students for higher professional

training as "intermediate secondary education", and university preparatory secondary education as "higher secondary education". Differences between the characteristics of the participants allocated to the paper versus the internet versions of the questionnaires were tested by Student's *t*-test and the χ^2 test. We assessed the frequency of missing answers to CHQ-CF items; the difference in the number of missing answers between the two formats was assessed by the Mann-Whitney *U* test. We assessed the distributions of the CHQ-CF scale scores to evaluate floor and ceiling effects ($\geq 25\%$ of the respondents having the lowest/highest score) for both modes of administration. Differences between CHQ-CF scale scores by format in the total sample were assessed by Mann-Whitney *U* tests. Additionally, after

transforming the scale scores into ranks, ANOVA was applied to test whether the mode of questionnaire administration interacted with the variables gender (male $n = 432$; female $n = 501$) and age (13–14 year olds, $n = 399$; 15–17 year olds, $n = 534$). Cohen's effect sizes, defined as $d = [\text{Mean}(a) - \text{Mean}(b)] / \text{SD}$, where the denominator was the square root of $[(n_a - 1)SD_a^2 + (n_b - 1)SD_b^2] / [(n_a - 1) + (n_b - 1)]$, were applied to indicate the relative magnitude of score differences between modes of administration. Here, the letters "a" and "b" refer to the subgroups administered the paper and internet forms of the test, respectively [28]. Following Cohen's suggested guidelines, $0.20 \leq d < 0.50$ indicated a "small effect", $0.50 \leq d < 0.80$ a "medium effect", and $d \geq 0.80$ a "large effect" [28]; Norman et al. have suggested that, in general, $d = 0.50$ can be considered as threshold for a "minimally important difference" (MID) [29]. Cronbach's α was applied to evaluate the internal consistency reliability of CHQ-CF multi-item scales by format; α of 0.70 or higher was considered to indicate sufficient internal consistency reliability [30]. We applied statistical tests of the hypothesis that the Cronbach's α reliability coefficients of CHQ multi-item scales in the sample administered the test on paper were equal to those administered the test online [31]. We applied item-level discriminant tests to evaluate whether the CHQ-CF items represent separate domains. For each mode of questionnaire administration, we evaluated whether (on average) correlation coefficients (Pearson- r correlation coefficients) between the items and their own scale score (without the item under consideration) were higher than the correlation coefficients between these items and any other scale. The average Pearson- r correlation coefficients were calculated by applying Fisher's z transformations [32]; we tested whether the differences between the average Pearson- r correlation coefficients in the subgroup administered the paper form and in the subgroup administered the test online were statistically significant [33]. We assessed the CHQ-CF's ability to discriminate between subgroups with 0, 1 or 2, and 3 or more chronic conditions, after having transformed the CHQ-CF scale scores into ranks, by ANOVA with the independent variables "number of chronic conditions", "mode of questionnaire administration", and the interaction term "number of chronic conditions"/"mode of questionnaire administration"; Cohen's effect sizes $d = [\text{Mean}(a) - \text{Mean}(b)] / \text{SD}$ in the condition subgroup were calculated for 1 or 2 versus 0 conditions, and for ≥ 3 versus 0 conditions. The designations "a" and "b" refer to the subgroups without chronic conditions and those with chronic conditions, respectively [28].

All analyses were done using SPSS, Version 11.0.1. The medical ethical committee of the Erasmus MC-University Medical Center Rotterdam, approved the study.

Results

Participants and randomization

The participation rate was 87%. The age range of the participants was 13–17 years (mean age 14.7 years; SD 0.68), 54% were female, 93% were born in the Netherlands, and the majority attended lower secondary education (Table 2). The prevalence of self-reported chronic conditions was as follows: asthma, 8%; allergies, 25%; hearing problems, 7%; visual problems, 8%; headaches or migraine, 17%; chronic lower back pain, 17%; and depression or anxiety attacks, 8% (Table 2). These characteristics were equally distributed in the groups assigned to the paper and internet versions of the questionnaires ($P \geq 0.05$; Table 2). The demographic characteristics of the participants (age, gender, country of birth, and educational level) reflected those of the general population of Dutch adolescents [34].

Difference in the number of missing answers between different modes of CHQ-CF administration

At the item level, use of the paper version of the CHQ-CF resulted in more missing answers (0–1.89% per item; mean 0.54%) compared with the internet version (0–0.22% per item; mean 0.04%; $P < 0.01$).

CHQ-CF scores by mode of administration

A ceiling effect was observed for four CHQ-CF scales in the subgroup that completed the paper questionnaire, and 3 CHQ-CF scales in the subgroup that completed the internet questionnaire (Table 3). Four CHQ-CF scales, i.e., "general behavior", "role functioning-physical", "mental health", and "family activities", resulted in statistically significant, higher scores for paper versus internet administration ($P < 0.05$), but the effect sizes (d) were ≤ 0.21 (Table 3). The mode of questionnaire administration did not interact significantly with gender ($P \geq 0.05$ regarding all scales), nor with age ($P \geq 0.05$ regarding six scales), except for the CHQ-CF scales "role

Table 2 Characteristics of study participants (total sample: $n = 933$; participants assigned to the paper questionnaire: $n = 475$; participants assigned to the internet questionnaire: $n = 458$)

	Total study group ($n = 933$)			Group with paper mode of administration ($n = 458$)			Group with Internet administration ($n = 475$)			P-value Internet versus paper mode
	Mean (SD) or Range	n	% of Participants	Mean (SD) or Range	n	% of Participants	Mean (SD) or Range	n	% of Participants	
Demographic characteristics										
Age (years)										
Mean (SD)	14.7 (0.68)			14.7 (0.68)			14.7 (0.68)			0.61 ^a
Range	13–17			13–17			13–17			
Gender										
Women		501	54%		244	51%		257	56%	0.17 ^b
Born in the Netherlands										
Yes		866	93%		441	93%		425	93%	0.90 ^b
Educational level of the school										
Lower secondary education		545	58%		274	58%		271	59%	
Intermediate secondary education		179	19%		94	20%		85	19%	0.88 ^c
Higher secondary education		209	22%		107	23%		102	22%	
Chronic conditions:										
Asthma										
Yes		76	8%		40	8%		36	8%	0.81 ^b
Allergies										
Yes		229	25%	118	25%		111	24%		0.88 ^b
Problems with hearing										
Yes		62	7%	30	6%		32	7%		0.70 ^b
Problems with seeing										
Yes		78	8%	39	8%		39	9%		0.91 ^b
Headaches or migraine										
Yes		159	17%	81	17%		78	17%		1.00 ^b
Chronic low back pain										
Yes		159	17%	88	19%		71	16%		0.22 ^b
Depression and/or anxiety attacks										
Yes		74	8%	36	8%		38	8%		0.72 ^b

^a Student's *t*-test^b Chi square test $df = 1$ ^c Chi square test $df = 2$

functioning-emotional/behavioral" ($P < 0.05$), "mental health" ($P < 0.05$), "self esteem" ($P < 0.05$), and "general health" ($P < 0.01$). Regarding these 4 CHQ-CF scales, administration of the paper version resulted in lower scores than online administration (or nearly equal scores in the case of

"mental health") in the subgroup of 13–14 year olds, while in the subgroup of 15–17 year olds, paper administration resulted in higher scores compared with internet administration; the Cohen's effect sizes (d) for these differences, regardless of sign, were ≤ 0.21 (data not shown).

Table 3 Comparison of mean scores, distributions of the scale scores, and other psychometric properties of CHO-CF scales in subgroups with paper ($n=475$) and internet modes ($n=458$) of questionnaire administration

CHO-CF scales ^a (range 0–100)	Mode of administration	Mean (SD)	Paper versus internet mode of administration	Range of scores	% max ^d	% min ^e	25 th %tile	50 th %tile	75 th %tile	Cronbach's alpha	Average item-own scale correlation ^f	Average item-other scale correlation
Physical Functioning	Paper	96.0 (6.9)	0.37	0.01	56	0	96	100	100	0.69	0.40	0.19
	Internet	95.8 (7.2)			53	0	93	100	100	0.72	0.44	0.18
Role funct.-Emo/beh	Paper	89.4 (17.2)	0.14	0.00	60	0	78	100	100	0.81	0.65	0.35
	Internet	89.5 (15.1)			54	0	86	100	100	0.70 ^g	0.51 ^g	0.28
Role funct.-Physical	Paper	95.0 (12.9)	0.02	0.05	81	0	100	100	100	0.86	0.74	0.30
	Internet	94.4 (11.7)			75	0	89	100	100	0.76 ^g	0.60 ^g	0.25
Bodily Pain	Paper	73.5 (22.7)	0.56	-0.05	25	0	60	80	100	0.88	0.80	0.35
	Internet	74.7 (21.4)			24	1	60	80	90	0.89	0.81	0.30
General Behavior	Paper	80.9 (10.6)	0.00	0.21	1	0	75	82	88	0.83	0.44	0.26
	Internet	78.6 (11.4)			0	0	72	79	87	0.85	0.46	0.26
Mental Health	Paper	76.5 (15.4)	0.01	0.12	2	0	69	80	89	0.92	0.63	0.35
	Internet	74.6 (14.8)			0	0	67	77	84	0.90 ^h	0.60	0.31
Self Esteem	Paper	74.7 (12.2)	0.99	0.02	1	0	68	75	82	0.88	0.56	0.30
	Internet	74.4 (11.6)			0	0	70	76	82	0.87	0.55	0.28
General Health	Paper	73.5 (16.5)	0.59	0.05	3	0	63	76	86	0.82	0.50	0.28
	Internet	72.6 (16.7)			1	0	63	75	86	0.81	0.58	0.28
Family Activities	Paper	80.0 (17.7)	0.03	0.11	17	0	67	83	96	0.81	0.58	0.24
	Internet	78.1 (16.9)			13	0	67	79	92	0.78	0.54	0.24
Family Cohesion	Paper	70.6 (23.5)	0.91	0.00	18	2	60	85	85	na	na	0.36
	Internet	70.7 (23.0)			16	2	60	85	85	na	na	0.33

^a The CHO-CF scales 'Role functioning-emotional' and 'Role functioning-behavioral' were merged into a single scale in this study
^b Two-sided Mann-Whitney U test of scale scores of the group using paper, and the group using the internet version of the questionnaire
^c Difference of mean scores divided by weighted average of SDs in groups given the paper and internet version of the questionnaires [28, 29]
^{d/e} % of respondents with best/worst possible score (ceiling/floor)
^f Each item was correlated with the applicable ad-hoc scale without the item under consideration
^{g/h} Statistically significant differences of Cronbach's alpha/average Pearson-r correlation coefficients in subgroups given paper or internet questionnaires (^g $P \leq 0.01$; ^h $P \leq 0.05$) [31, 33].
 na Not applicable; Role funct.-Emo/beh - Role functioning-emotional behavioral; Role funct.-Physical - Role functioning-physical

Internal consistency reliability of scales by mode of administration

Cronbach's α s for the two formats were adequate for all CHQ-CF scales, except "physical functioning" in the subgroup administered the paper version of the questionnaire ($\alpha = 0.69$). The two "role functioning" scales and "mental health" showed statistically significant, higher Cronbach's α s in the subgroup administered the paper version of the questionnaire compared with the alphas in the subgroup administered the internet version ($P < 0.01$, respectively $P < 0.05$) (Table 3). All multi-item scales, regarding both modes of administration, showed higher average (corrected) item-own scale correlation coefficients than average item-other-scale correlation coefficients. The two "role functioning" scales showed statistically significant, higher average item-own scale correlation coefficients in the subgroup administered the paper version of the questionnaire compared with the item-own scale correlation coefficients in the subgroup administered the internet version ($P < 0.01$) (Table 3).

Construct validity by mode of administration

All mean CHQ-CF scale scores were lower in the subgroup with one or two reported conditions and in the subgroup with three or more reported conditions when either was compared with the subgroup with no reported conditions. For both modes of questionnaire administration, and for all CHQ-CF scales, the more chronic conditions that were reported, the higher the effect sizes compared with the subgroup with no chronic conditions. ANOVA showed statistically significant CHQ-CF score differences by "number of chronic conditions" for all scales ($P < 0.01$) (Table 4). The mode of questionnaire administration did not interact significantly with the variable "number of chronic conditions" ($P \geq 0.05$ for all scales) (Table 4).

Discussion and conclusions

In this study we applied a randomized design to compare the results of the Child Health Questionnaire-Child Form (CHQ-CF) administered by a paper questionnaire and by an online questionnaire. The results provided support for the feasibility, internal consistency reliability, and construct validity of the CHQ-CF scales. Both modes of questionnaire administration yielded comparable scale scores and showed comparable psychometric properties. Additionally, the

study provided reference/norm scores for clinical studies (general population of 13–17 year olds).

Strengths of the current study

The participation rate was high. Study group characteristics (age, gender, country of birth, and educational levels) were representative of those of the general population of Dutch adolescents [34]. Randomization to either the paper or internet mode of administration of the CHQ-CF was successful with respect to the evaluated characteristics.

Limitations

We applied a randomized parallel group design that allows for the comparison of indicators of feasibility, reliability, and validity at the group level between a subgroup that completed a paper version and a subgroup that completed an internet version of the CHQ-CF. However, this did not allow an evaluation of whether the same person would provide equivalent or different answers to the same CHQ-CF questionnaire administered by the alternative mode, which would require a randomized crossover design [25, 35]. Such an evaluation at the individual level requires the respondent to forget all previously provided answers at the second assessment, e.g., by waiting 1 or 2 weeks between the two measurements. It also requires that there is no effect from having previously completed a CHQ-CF questionnaire by any mode, paper or internet, at the second assessment, and that scores by the same mode of administration after a relatively short interval, in the absence of changes in health status, are exactly the same. However, in an evaluation of retesting with the same paper version of the CHQ-CF after 2 weeks, 5 out of 10 CHQ-CF scales showed statistically significant, higher scores at the second measurement with Cohen's effect sizes ranging from 0.25 to 0.40, while intraclass correlation coefficients between the first and second measurement ranged from 0.06 thru 0.84 [7]. Furthermore, in a randomized crossover design, "carry-over" effects may be present, i.e., completing an internet version before a paper version may have a different effect on the second assessment, than does completing a paper version before an internet version [35]. Despite the logistical and the above-mentioned methodological challenges, we recommend future studies comparing the paper and internet versions of the CHQ-CF applying a randomized crossover design to evaluate congruency of answers to CHQ-CF items at the individual level.

Table 4 Ability of the CHQ-CF scales to discriminate between subgroups differing in the participants' number of chronic conditions, for the group that was assigned to complete the paper version ($n = 475$) and for the group that was assigned to complete the internet version ($n = 458$)

CHQ-CF scales ^a (range 0–100):	Number of chronic conditions per participant						
	0 conditions		1 or 2 conditions		≥3 conditions		ANOVA P -value Interaction term
	Mode of administration	$n = 224$ (Paper) $n = 206$ (Internet) Mean (SD)	$n = 210$ (Paper) $n = 217$ (Internet) Mean (SD)	$n = 41$ (Paper) $n = 35$ (Internet) Mean (SD)	Effect size d^b	ANOVA P -value Number of chronic conditions	
Physical functioning	Paper	97 (5)	96 (7)	90 (10)	0.24	0.80*	
	Internet	97 (5)	95 (7)	89 (12)	0.28	0.64*	0.52
Role funct.-emo/behav	Paper	93 (13)	88 (18)	77 (24)	0.25	0.64*	
	Internet	91 (14)	89 (15)	79 (18)	0.15	0.66*	0.72
Role funct.-physical	Paper	97 (11)	94 (13)	91 (17)	0.19	0.34	
	Internet	96 (10)	94 (13)	91 (13)	0.14	0.34	0.69
Bodily pain	Paper	82 (18)	70 (22)	47 (24)	0.52*	1.45*	
	Internet	80 (18)	72 (21)	55 (24)	0.39	1.03*	0.32
General behavior	Paper	84 (9)	80 (10)	71 (13)	0.40	0.98*	
	Internet	81 (10)	78 (11)	69 (14)	0.26	0.90*	0.65
Mental health	Paper	83 (10)	74 (16)	57 (16)	0.56*	1.58*	
	Internet	80 (11)	72 (15)	59 (18)	0.50	1.20*	0.29
Self esteem	Paper	79 (10)	73 (12)	61 (12)	0.52*	1.44*	
	Internet	77 (9)	73 (12)	66 (14)	0.33	0.80*	0.06
General health	Paper	80 (14)	71 (15)	54 (14)	0.57*	1.82*	
	Internet	79 (13)	70 (17)	55 (18)	0.54*	1.29*	0.58
Family activities	Paper	83 (16)	79 (18)	70 (20)	0.26	0.67*	
	Internet	81 (16)	78 (17)	66 (15)	0.17	0.97*	0.51
Family cohesion	Paper	77 (21)	68 (23)	53 (26)	0.38	0.89*	
	Internet	74 (20)	68 (24)	64 (28)	0.24	0.38	0.05

^a The CHQ-CF scales 'Role functioning-emotional' and 'Role functioning-behavioral' were merged into a single scale in this study

^b Difference of the means divided by SD in the subgroup with chronic condition(s), where * indicates at least a "minimally important difference" ($d \geq 0.50$) [28, 29]

Role funct.-Emo/beh - Role functioning-emotional behavioral; Role funct.-Physical - Role functioning-physical

In this study, internet and paper questionnaires were completed in a controlled environment with adequate privacy and supervision. This may not be the case during future applications. We are unaware of the impact less privacy during completion of the questionnaires may have, but this would apply to both the paper and the internet versions of the questionnaire.

For both modes of questionnaire administration, we did not evaluate correlations between CHQ-CF scores and a relevant parent-rated questionnaire such as the CHQ-PF50 [2, 3]. Test-retest reliability of the CHQ-CF and its responsiveness and sensitivity to changes in health were not evaluated in the current study. The CHQ-CF has a relatively large number of items; therefore, we recommend developing a shorter version in the future.

Psychometric properties

The psychometric properties, with only a few exceptions, were equal between the two modes of questionnaire administration. The Cronbach's α of the scale "physical functioning" in the subgroup administered the paper version of the questionnaire was just under 0.70, and the difference with the alpha in the subgroup administered the internet version was not statistically significant.

Missing values

Compared with the paper version, the internet version was successful in reducing the quantity of missing data. Theoretically, differences in selective partial non-response between formats might have contributed to differences in scores that were reported in this study. In our study, in the subgroup ($n = 86$) that had at least one missing answer to a paper CHQ-CF item, all scale means were somewhat lower than in the subgroup ($n = 389$) with no missing answers, but these differences were not significant ($P \geq 0.05$). Thus, missing answers are not a source of the observed score differences.

Score differences between modes of questionnaire administration

Recently, Ritter et al. found no statistically significant score differences between internet and paper modes of administration for 16 health-related measures, but the study was conducted in an opportunity sample retrieved from the internet, which limits its generalizability [23]. In a randomized internet-paper comparison among adolescents concerning various health measures other than the CHQ-CF, only one statistically significant

score difference was reported among 21 topics [24]. In another randomized adolescent study, a medical consumption index and 11 indicators of fruit consumption and determinants of fruit consumption did not show statistically significant score differences between internet and paper administration of the questionnaire, except for one measure that showed small score differences between modes of administration [25]. The International Study of Asthma and Allergies in Childhood (ISAAC) questionnaire did not show statistically significant score differences between internet and paper administration in two randomized adolescent studies [25, 26].

In our study, in the whole sample, the paper version resulted in slightly, yet statistically significant, higher scores for 4 of 10 CHQ-CF scales compared with the internet version. One plausible explanation is chance, since it should be considered that given multiple comparisons, there is a 1-in-20 chance of a false association for each comparison (Type I error at $\alpha = 0.05$) [36]. A commonly used Bonferroni correction for 10 comparisons would imply an adjusted $\alpha = 0.05/10 = 0.005$ [36]; at $\alpha = 0.005$, only one score difference (regarding the scale "general behavior") was significant. Furthermore, given Cohen's suggested guidelines for the interpretation of effect sizes, three of the four statistically significant differences between modes of administration can be considered as negligible ($d \leq 0.12$), and one difference regarding the CHQ-scale "general behavior" ($d = 0.21$) can be considered as small [28]; all effect sizes were far below $d = 0.50$ that was suggested as an approximate threshold for "minimally important differences" by Norman et al. [29]. This study provides no explanations for the established small score differences between paper and internet administration, or for the established statistically significant, but small interaction effects of administration mode with age in the case of four CHQ-CF scales.

Conclusions

With increasing application of online health questionnaires rather than questionnaires on paper, especially in adolescent populations, it should be noted that comparison of results requires that the scores between these modes of administration do not show meaningful statistically significant differences. This study showed that, overall, paper and internet versions of the CHQ-CF yielded only a few, negligible or small, differences. Paper and internet modes of CHQ-CF administration may be combined in a single study, although researchers should consider the possibility of minor

score differences depending on the mode of administration for some scales. We recommend repeated studies in other populations, including clinical populations, to confirm or reject our results.

Acknowledgements The Municipal Health Services in Vlaardingen and Harderwijk, The Netherlands, were responsible for data collection. We are grateful to the school physicians, nurses, doctor's assistants, and epidemiologists of the participating Municipal Health Services for facilitating this project in collaboration with the related schools. We thank Ineke Vogel for expert statistical advice. This study was funded by the Netherlands Organization for Health Research and Development (ZonMw) Prevention Research Program Grant #2100.0066.

References

1. Connolly, M. A., & Johnson, J. A. (1999). Measuring quality of life in paediatric patients. *Pharmacoeconomics*, *16*(6):605–625.
2. Raat, H., Mohangoo, A. D., & Grootenhuis, M. A. (2006). Pediatric health-related quality of life questionnaires in clinical trials. *Current Opinion in Allergy and Clinical Immunology*, *6*(3):180–185.
3. Landgraf, J. M., Abetz, L., & Ware, J. E. (1996). *The CHQ user's manual*. Boston: The Health Institute, New England Medical Center.
4. Landgraf, J. M., & Abetz, L. N. (1997). Functional status and well-being of children representing three cultural groups: Initial self-reports using the CHQ-CF87. *Psychology and Health*, *12*:839–854.
5. Ruperto N., Ravelli A., Pistorio A., Malattia C., Cavuto S., Gado-West L. & et al. (2001). Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire in 32 countries. Review of the general methodology. *Clinical and Experimental Rheumatology*, *19*(4(supplement 23)):S1–S9.
6. Waters, E. B., Salmon, L. A., Wake, M., Wright, M., & Hesketh, K. D. (2001). The health and well-being of adolescents: A school-based population study of the self-report Child Health Questionnaire. *Journal of Adolescent Health*, *29*(2):140–149.
7. Raat, H., Landgraf, J. M., Bonsel, G. J., Gemke, R. J., & Essink-Bot, M. L. (2002). Reliability and validity of the child health questionnaire-child form (CHQ-CF87) in a Dutch adolescent population. *Quality of Life Research*, *11*(6):575–581.
8. Schonlau, M. (2004). Will web surveys ever become part of mainstream research? *Journal of Medical Internet Research*, *6*(3):e31.
9. Bloom, D. E. (1998). Technology, experimentation, and the quality of survey data. *Science*, *280*(5365):847–848.
10. Bayliss, M. S., Dewey, J. E., Dunlap, I., Batenhorst A. S., Cady, R., Diamond, M. L. & et al. (2003). A study of the feasibility of Internet administration of a computerized health survey: The headache impact test (HIT). *Quality of Life Research*, *12*(8):953–961.
11. O'Toole, B. I., Battistutta, D., Long, A., & Crouch, K. (1986). A comparison of costs and data quality of three health survey methods: Mail, telephone and personal home interview. *American Journal of Epidemiology*, *124*(2):317–328.
12. McHorney, C. A., Kosinski, M., & Ware, J. E. Jr. (1994). Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: Results from a national survey. *Medical Care*, *32*(6):551–567.
13. Brewer, N. T., Hallman, W. K., Fiedler, N., & Kipen H. M. (2004). Why do people report better health by phone than by mail? *Medical Care*, *42*(9):875–883.
14. Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health (Oxford, England)*, *27*(3):281–291.
15. Kleinman, L., Leidy, N. K., Crawley, J., Bonomi, A., & Schoenfeld P. (2001). A comparative trial of paper-and-pencil versus computer administration of the Quality of Life in Reflux and Dyspepsia (QOLRAD) questionnaire. *Medical Care*, *39*(2):181–189.
16. Ryan, J. M., Corry, J. R., Attewell, R., & Smithson, M. J. (2002). A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Quality of Life Research*, *11*(1):19–26.
17. Litaker, D. (2003). New technology in quality of life research: Are all computer-assisted approaches created equal? *Quality of Life Research*, *12*(4):387–393.
18. Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, *280*(5365):867–873.
19. Webb, P. M., Zimet, G. D., Fortenberry, J. D., & Blythe, M. J. (1999). Comparability of a computer-assisted versus written method for collecting health behavior information from adolescent patients. *Journal of Adolescent Health*, *24*(6):383–388.
20. MacMillan, H. L. (1999). Computer survey technology: A window on sensitive issues. *CMAJ: Canadian Medical Association Journal*, *161*(9):1142.
21. Pealer, L. N., Weiler, R. M., Pigg, R. M. Jr., Miller, D., & Dorman, S. M. (2001). The feasibility of a web-based surveillance system to collect health risk behavior data from college students. *Health Education & Behavior*, *28*(5):547–559.
22. Balter, K. A., Balter, O., Fondell, E., & Lagerros, Y. T. (2005). Web-based and mailed questionnaires: A comparison of response rates and compliance. *Epidemiology*, *16*(4):577–579.
23. Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical and Internet Research*, *6*(3):e29.
24. Mangunkusumo, R. T., Moorman, P. W., van den Berg-de Ruiter, A. E., van der Lei, J., de Koning H. J., & Raat, H. (2005). Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *Journal of Adolescent Health*, *36*(1):70.e1–6.
25. Mangunkusumo, R. T., Duisterhout, J. S., de Graaff, N., Maarsingh, E. J., de Koning, H. J., & Raat, H. (2006). Internet versus paper mode of health and health behavior questionnaires in elementary schools: Asthma and fruit as examples. *The Journal of School Health*, *76*(2):80–86.
26. Raat, H., Mangunkusumo, R. T., Mohangoo A. D., Juniper E. F., & Van Der Lei J. (2007). Internet and written respiratory questionnaires yield equivalent results for adolescents. *Pediatric Pulmonology*, In press.
27. Gosselin, D. (2005) *PHP Programming with MySQL*. Course Technology, (1st ed.).
28. Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, Academic Press
29. 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):582–592.
30. Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *British Medical Journal*, 314(7080):572.
 31. Feldt, L. S., & Kim, S. (2006). Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement*, 66(4):589–600.
 32. Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined Pearson r s and Fisher's z transformations. *The Journal of General Psychology*, 125(3):245–261.
 33. Hays, H.L. (1994). *Statistics*. (5th ed.). Fort Worth: Harcourt Brace College Publishers.
 34. Statistics Netherlands - Statline. In. Heerlen, The Netherlands; 2005.
 35. Garcia, R., Benet, M., Arnau, C., & Cobo, E. (2004). Efficiency of the cross-over design: An empirical estimation. *Statistics in Medicine*, 23(24):3773–3780.
 36. Wit, E., & McClure, J. (2004). *Statistics for microarrays: design, analysis and inference*. Chichester, John Wiley & Sons, Ltd.