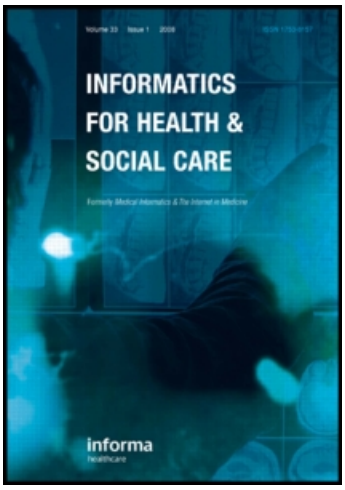
This article was downloaded by: *[Hjollund, Niels Henrik]*
On: *24 March 2009*
Access details: *Access Details: [subscription number 909744547]*
Publisher *Informa Healthcare*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Informatics for Health and Social Care
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713736879

Reliability of short form-36 in an Internet- and a pen-and-paper version

Maja Basnov [a]; Sissel Marie Kongsved [a]; Per Bech [b]; Niels Henrik Hjollund [acd]
[a] Center of Public Health, Region Central Jutland, Aarhus, Denmark [b] Psychiatric Research Unit, Psychiatric Hospital, Hillerod, Denmark [c] Department of Clinical Social Medicine, Institute of Public Health, Aarhus University, Aarhus, Denmark [d] Department of Occupational Medicine, Herning Regional Hospital, Herning, Denmark

Online Publication Date: 01 January 2009

To cite this Article Basnov, Maja, Kongsved, Sissel Marie, Bech, Per and Hjollund, Niels Henrik(2009)'Reliability of short form-36 in an Internet- and a pen-and-paper version',Informatics for Health and Social Care,34:1,53 — 58
To link to this Article: DOI: 10.1080/17538150902779527
URL: http://dx.doi.org/10.1080/17538150902779527

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Informatics for Health & Social Care
January 2009; 34(1): 53–58

informa
healthcare

Reliability of short form-36 in an Internet- and a pen-and-paper version

MAJA BASNOV[1], SISSEL MARIE KONGSVED[1], PER BECH[2] & NIELS HENRIK HJOLLUND[1,3,4]

[1]*Center of Public Health, Region Central Jutland, Aarhus, Denmark,* [2]*Psychiatric Research Unit, Psychiatric Hospital, Hillerod, Denmark,* [3]*Department of Clinical Social Medicine, Institute of Public Health, Aarhus University, Aarhus, Denmark and* [4]*Department of Occupational Medicine, Herning Regional Hospital, Herning, Denmark*

Downloaded By: [Hjollund, Niels Henrik] At: 08:02 24 March 2009

Abstract
Use of Internet versions of questionnaires may have several advantages in clinical and epidemiological research, but we know little about if Internet versions differ with respect to validity and reliability. We aimed to compare Internet- and pen-and-paper versions of short form-36 (SF-36) with respect to test–retest reliability and internal consistency. Women referred to mammography ($n = 782$) were randomised to receive either a paper version with a prepaid return envelope or a guideline on how to fill in the Internet version. A subgroup was asked to answer the questionnaire once again in the alternative version. Test–retest reliability was assessed by the intra-class correlation coefficient. Internal consistency was calculated as Cronbach's alpha. The between-version test–retest reliability for the eight subscales were between 0.63 and 0.92. Cronbach's alpha for the two versions were all between 0.75 and 0.93 with minor differences between the Internet- and the pen-and-paper version. We found little or no evidence of a difference in test–retest reliability and internal consistency when we compared an Internet- and a pen-and-paper version of SF-36.

Keywords: *Internet, internal consistency, questionnaire, reliability, test–retest*

1. Introduction

Internet versions of questionnaires may have several advantages in clinical and epidemiological research [1,2], but we do not know if we can compare answers obtained by an Internet version with answers based on a traditional pen-and-paper questionnaire. Even extensively validated questionnaires are usually only validated in the pen-and-paper version. Only few studies have compared psychometric properties across versions, and due to methodological shortcomings the results cannot be generalised [3,4].

Reliability is a measure of the extent to which a measurement is free from unsystematic (random) error. In repeated measurements test–retest reliability is defined as the proportion of the total variance in the measurements ($\sigma_s^2 + \sigma_e^2$) which is due to 'true' differences between subjects (σ_s^2) [5]. Internal consistency (Cronbach's alpha) is often misleading reported as the

Correspondence: Niels Henrik Hjollund, Department of Occupational Medicine, Herning Regional Hospital, Gl. Landevej 61, DK-7400 Herning, Denmark. Tel: +45-99-27-24-70. Fax: +45-99-27-24-67. E-mail: hehjo@ringamt.dk

ISSN 1753-8157 print/ISSN 1753-8165 online © 2009 Informa Healthcare USA, Inc.
DOI: 10.1080/17538150902779527

Copyright Informa Healthcare 2009
Not for Sale or Commercial Distribution
Unauthorized users can download,
Authorized users can download,
display, view and print a single
copy for personal use.

reliability of a test [6]. However, because they are based on performance observed in a single setting, there are many sources of variance, which do not enter into the calculation and values are expected to be larger than measures of reliability based on repeated measurements [5].

Short Form-36 (SF-36) is widely used in medical research as a general health measure [7,8]. Three studies have compared an Internet version of SF-36 with a pen-and-paper version and calculated test–retest reliability [9–11]. In these studies the patients filled in the second version of SF-36 within a few hours after the first version. If the interval between the versions is short the answers given the first time may be recalled and affect the answers given the second time, and a time interval from 2 days to 2 weeks between administration of the same questionnaire is often recommended in studies of test–retest reliability [12]. The results from the two studies are therefore hardly sufficient to document the reliability of SF-36 for Internet use. We aimed to compare an Internet version of SF-36 with a pen-and-paper version, with respect to internal consistency and test–retest reliability across versions.

Downloaded By: [Hjollund, Niels Henrik] At: 08:02 24 March 2009

2. Material and methods

2.1. Participants

Participants were women referred for mammography from September 2004 to April 2005 in the Department of Radiology at the public hospital, Randers Regional Hospital, Denmark. Patients were referred by their family doctor. A consultant at the Department of Radiology assigned the referred patients to one of three categories: acute, sub-acute or non-acute. Subsequently, a letter was sent to the woman informing her about the date, location and other details of the mammography. We included women up to retirement age (67 years in Demark) who were referred to their first mammography and did not have a history of breast cancer. Only patients from the acute and the sub-acute group were included in the analysis ($n = 1021$). The women were evenly randomised to answer an Internet- or a pen-and-paper version of the questionnaire. The procedure is more detailed described elsewhere [13].

The women received information of the study together with the letter informing about the date for the mammography. The letter to women randomised to answer the Internet version of the questionnaire included a guideline on how to answer the Web-based questionnaire, whereas those allocated to the pen-and-paper version were asked to fill in and return the questionnaire in a prepaid envelope. If a participant answered the questionnaire at least 10 days before her mammography, she was mailed again and asked to fill out the questionnaire in the alternative version. Only questionnaires filled in before the date of the mammography were included in the analysis.

2.2. Questionnaire

The questionnaire consisted of 17 pages and 119 items and included SF-36 [8] followed by multidimensional fatigue inventory-20 [14] and The Hospital Anxiety and Depression Scale [15]. Items regarding social status, educational level, occupation and access to the Internet were also included.

The layout of the pen-and-paper version of SF-36 was in accordance with the Danish manual of SF-36 [7,8], and the layout of the Internet version was as close to the paper version as possible [13]. In the Internet version the participants were reminded of missing answers if they tried to leave a Web-page with missing answers. However, after clicking an 'OK'-button they were allowed to continue, even if there were still missing answers.

2.3. Assessment of reliability

Test–retest reliability was assessed as the intra-class correlation coefficient (ICC) ICC(3,1) [5]. In case of only two measurements and fixed methods, the ICC(3,1) coefficient can be calculated as the product–moment correlation [16]. Data must be normally distributed, which is not the case with SF-36 data. We therefore also calculated Spearman's rank correlation coefficient. Internal consistency was calculated as Cronbach's alpha, and confidence intervals for Cronbach's alpha were calculated using the macro published by Duhacheck and Lacobucci [17].

3. Results

A total of 782 women were enrolled and randomised. Of the 385 women randomised to answer the Internet version, 91 (23.6%) responded, compared with 301 (75.8%) of the 397 women randomised to answer the pen-and-paper version. Determinants for response rates are described elsewhere [13]. Characteristics of the women are shown in Table I. There were no differences between respondents to the Internet- and the pen-and-paper version with respect to age and education.

Downloaded By: [Hjollund, Niels Henrik] At: 08:02 24 March 2009

3.1. Test–retest reliability between versions

Forty-one women answered both versions of the questionnaire. Among these, 51% first answered the pen-and-paper version and later the Internet version. The mean time range between the two versions was 14.7 days (standard deviation (SD) 8.0 days). Estimates of test–retest reliability between the two versions are presented in Table II. Spearman's rank

Table I. Characteristics of participants included in analyses of reliability of pen-and-paper and Internet versions of short form-36.

Analysis of internal consistency		
Internet version ($n = 91$)	Pen-and-paper version ($n = 301$)	Analysis of test–retest between Internet and pen-and-paper versions ($n = 41$)
45.9 (9.7)	46.5 (10.6)	47.2 (9.0)
13.0 (1.9)	12.7 (2.4)	12.7 (2.1)

Mean values (SD).

Table II. Test–retest reliability between Internet and pen-and-paper versions of short form-36.

Short form-36 subscales	Mean (SD)	ICC (95% CI)	Rank correlation (95% CI)
Physical function	89.9 (16.2)	0.92 (0.86–0.98)	0.80 (0.64–0.95)
Role physical	81.0 (32.1)	0.73 (0.52–0.94)	0.71 (0.47–0.94)
Bodily pain	76.2 (22.7)	0.88 (0.82–0.93)	0.89 (0.82–0.95)
General health	75.7 (20.0)	0.70 (0.51–0.90)	0.66 (0.43–0.89)
Vitality	63.8 (23.0)	0.82 (0.69–0.96)	0.75 (0.56–0.94)
Social function	86.1 (20.6)	0.73 (0.48–0.97)	0.54 (0.24–0.84)
Role emotional	79.3 (32.8)	0.63 (0.38–0.88)	0.54 (0.21–0.86)
Mental health	70.3 (20.2)	0.75 (0.61–0.89)	0.74 (0.58–0.90)

ICC, intra class correlation.

correlation coefficient was between 0.54 and 0.89 for the eight subscales, whereas the product–moment correlation was between 0.63 and 0.92. Mean differences (pen-and-paper minus Internet) ranged from −4.82 to +2.31 for the eight subscales (paired *t*-test for difference = 0: $p > 0.2$ for any subscale).

3.2. Internal consistency of the two versions

Internal consistency for the Internet version and the pen-and-paper version is shown in Figure 1. Cronbach's alpha was in the range of 0.75–0.92 for the Internet version and 0.79–0.93 for the pen-and-paper version. Only the scale for *physical function* did not have overlapping 95% confidence intervals.

4. Discussion

We found little or no evidence of a difference in reliability between an Internet- and a pen-and-paper version of SF-36 in a group of 392 women referred to mammography from their family doctor. The mean and SD of the eight subscales were in the same range as a Danish general population sample (Table III). The sample size was limited, but to our knowledge, this is the first study to report on test–retest reliability between Internet- and pen-and-paper versions of SF-36 with a relevant time range between the versions. The response rate was low in the Internet group and bias must be considered. However, test–retest analyses are based on intra-individual comparisons, which are less vulnerable to selection bias than cross-sectional

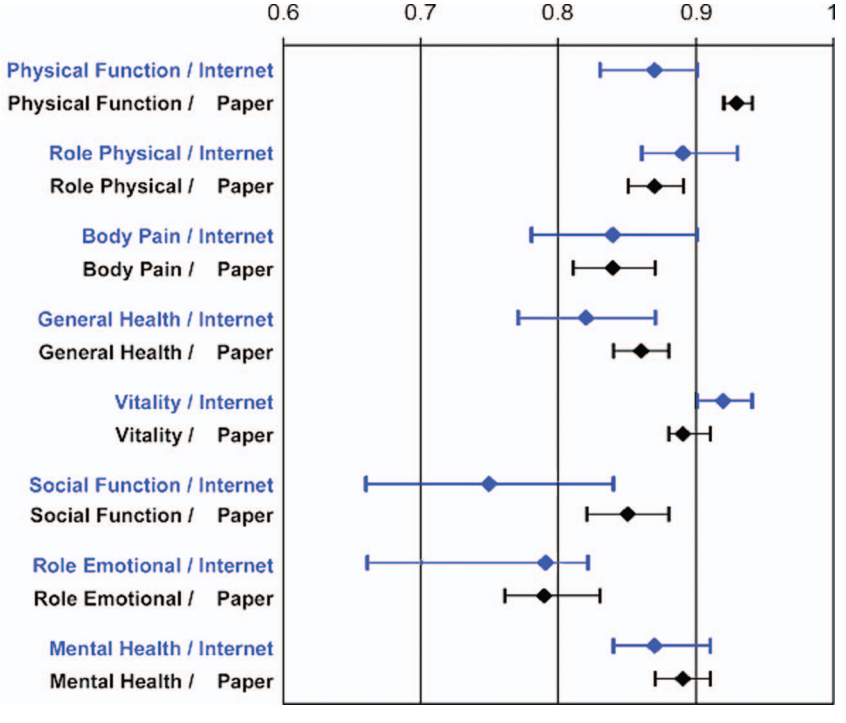


Figure 1. Internal consistency (Cronbach's alpha) for the Internet version and the pen-and-paper version of short form-36 subscales.

Downloaded By: [Hjollund, Niels Henrik] At: 08:02 24 March 2009

Table III. Characteristics of short form-36 in a Danish population sample (women 16+ years, $n = 2141$) [7].

Scale	Mean (SD)	Items (n)	Internal consistency (Cronbach's alpha)
Physical function	86.4 (21.2)	10	0.92
Role physical	80.3 (33.6)	4	0.87
Bodily pain	76.3 (24.3)	2	0.84
General health	75.6 (20.9)	5	0.80
Vitality	67.4 (20.6)	4	0.85
Social function	90.2 (18.6)	2	0.80
Role emotional	84.5 (29.2)	3	0.76
Mental health	80.0 (16.0)	5	0.81

analyses. Furthermore, the respondents in the two groups ($n = 248$) did not differ with respect to age, education and scores for subscales of SF-36 [13].

Downloaded By: [Hjollund, Niels Henrik] At: 08:02 24 March 2009

4.1. Test–retest reliability within and between versions

Although reliability studies of Internet versions of SF-36 are lacking, a number of studies have explored the test–retest reliability of pen-and-paper versions of SF-36. A study of the Basque version of SF-36 found ICCs ranging between 0.57 and 0.89 for the eight subscales [18], whereas a study in Turkish cancer patients found ICCs ranging between 0.81 and 0.94 [19]. These two studies compared the same pen-and-paper version of SF-36 at two occasions with a test–retest interval of 2 weeks. Our results are comparable with their results, which suggests that test–retest reliability does not depend on which version of SF-36 is used.

4.2. Internal consistency

The subscale *Physical function* had lower internal consistency in the Internet version. However, the value of 0.87 is still acceptable and within the range of values published for the pen-and-paper version (Table III). Cronbach's alpha is often referred to as a measure of reliability [5], although it is calculated from only one measurement. As expected, Cronbach's alpha values were generally higher than the correspondent ICCs, because the ICC includes other sources of variation than Cronbach's alpha [5]. In a population sample of Danish women Cronbach's alpha ranged between 0.76 and 0.92 for the eight subscales of SF-36 [7] (Table III). The subscales with the highest and lowest Cronbach's alpha in that study tend to be the same as in our study.

5. Conclusion

We found little or no differences with respect to test–retest reliability or internal consistency when we compared an Internet- and a pen-and-paper version of SF-36. Although the sample size was limited, our results suggest that answers obtained from Internet questionnaires may be compared with results obtained by traditional pen-and-paper questionnaires.

Acknowledgement

This work was supported by a grant from the Danish Cancer Society.

References

1. Buchanan T, Smith JL. Using the Internet for psychological research: personality testing on the World wide Web. British Journal of Psychology 1999;90(Pt 1):125–144.
2. Cramer JA, Silberstein SD, Winner P. Development and validation of the Headache Needs Assessment (HANA) survey. Headache 2001;41:402–409.
3. Andersson G, Kaldo-Sandstrom V, Strom L, Stromgren T. Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. Journal of Psychosomatic Research 2003;55:259–262.
4. Athale N, Sturley A, Skoczen S, Kavanaugh A, Lenert L. A web-compatible instrument for measuring self-reported disease activity in arthritis. Journal of Rheumatology 2004;31:223–228.
5. Streiner DL, Norman GR. Health measurement scales, a practical guide to their development and use, New York: Oxford University Press; 2003.
6. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
7. Bue Bjørner J, Trab Damsgaard M, Watt T, Bech P, Rasmussen N, Søndergaard Kristensen T, Modvig J, Thunedborg K. Danish manual for short form-36 [Dansk manual til SF-36], Copenhagen: The Danish Association of Pharmaceutical Industry; 1997.
8. Ware JE Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care 1992;30:473–483.
9. Wilson AS, Kitas GD, Carruthers DM, Reay C, Skan J, Harris S, Treharne GJ, Young SP, Bacon PA. Computerized information-gathering in specialist rheumatology clinics: an initial evaluation of an electronic version of the short form 36. Rheumatology (Oxford) 2002;41:268–273.
10. Bliven BD, Kaufman SE, Spertus JA. Electronic collection of health-related quality of life data: validity, time benefits, and patient preference. Quality of Life Research 2001;10:15–22.
11. Chen TH, Li L, Sigle JM, Du YP, Wang HM, Lei J. Crossover randomized controlled trial of the electronic version of the Chinese SF-36. Journal of Zhejiang University Science B 2007;8:604–608.
12. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. Journal of Clinical Epidemiology 2003;56:730–735.
13. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet *versus* paper-and-pencil versions. Journal of Medical Internet Research 2007;9:e25.
14. Smets EM, Garssen B, Bonke B, de Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. Journal of Psychosomatic Research 1995;39:315–325.
15. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica 1983;67:361–370.
16. Shrout PE. Reliability. In: Tsuang MT, Tohen M, editors. Textbook in psychiatric epidemiology, 2nd edn, New York: John Wiley & Sons; 2002. p 131–147.
17. Duhachek A, Lacobucci D. Alpha's standard error (ASE): an accurate and precise confidence interval estimate. Journal of Applied Psychology 2004;89:792–808.
18. Gonzalez N, Quintana JM, Arostegui I, Padierna A, Martinez E, Crespo I, Vesga MA. Translation and psychometric testing of the Basque version of the SF-36 health survey. Quality of Life Research 2005;14:549–554.
19. Pinar R. Reliability and construct validity of the SF-36 in Turkish cancer patients. Quality of Life Research 2005;14:259–264.

Downloaded By: [Hjollund, Niels Henrik] At: 08:02 24 March 2009