Factors associated with incomplete DASH questionnaires

Arjan G. J. Bot · Steven Ferree · Valentin Neuhaus ·
David Ring

Published online: 28 December 2012
© American Association for Hand Surgery 2012

Abstract

Background Missing data are unavoidable in clinical research. Older age, female gender, and fewer years of education are risk factors for missing items in a questionnaire. This study assessed the differences between patients with complete and incomplete Disabilities of the Arms, Shoulder and Hand (DASH) questionnaires in terms of demographics and psychological factors.

Methods We analyzed a convenience sample of 1,204 patients enrolled in eight prospective studies. The DASH and the Pain Catastrophizing Scale were completed by all patients. The Center for Epidemiologic Studies Depression scale, Patient Health Questionnaire, Pain Anxiety Symptoms Scale, and an ordinal pain scale were completed by 745, 493, 545, and 391 patients, respectively. Bivariate analysis and binary logistic regression were used to determine risk factors for incomplete (one or more unanswered question) or invalid (more than three unanswered questions) DASH questionnaires.

Results Thirty-one percent of patients did not complete all questions on the DASH. Patients with an incomplete DASH were older, had fewer years of education, and had higher levels of catastrophic thinking, depression, and pain anxiety. Age and catastrophic thinking were retained in the best logistic regression models of predictors of both incomplete and invalid DASH questionnaires.

Conclusions The observation that patients who complete disability questionnaires are different from patients who do not may affect the interpretation of clinical research. Computer adaptive testing may be preferable to avoid incomplete questionnaires.

Level of Evidence: Prognostic Level II.

Keywords Research · Incomplete questionnaires · Disability

Introduction

Missing data are commonplace in clinical research that uses paper questionnaires [12, 26]. For example, in a large international study by Gandek and colleagues, up to 13 % of the participants did not complete the SF-36 [11]. Missing answers can compromise the validity of questionnaires and may contribute to inaccuracy or bias [7, 12, 25]. Methods to deal with missing data [5, 6] are helpful, but complete data are ideal.

Prior studies have identified the following predictors of missing items in a questionnaire: age [2, 16, 21, 22, 28], female sex [16, 21, 22, 28], and fewer years of education [16, 22]. The aim of this study was to investigate the differences between patients with an incomplete Disabilities of the Arms, Shoulder and Hand (DASH) and patients with a complete questionnaire. Our null hypothesis was that there are no differences in demographics, catastrophic thinking, and depression between patients with complete and incomplete DASH. Secondary study questions addressed the hypotheses that there would be no differences between patients with a valid DASH and patients with an invalid DASH (more than three unanswered questions) and between completers and non-completers of the most frequent missing questions.

A. G. J. Bot · S. Ferree · V. Neuhaus
Orthopaedic Hand and Upper Extremity Service, Massachusetts
General Hospital, Boston, MA, USA

D. Ring (✉)
Harvard Medical School, Orthopaedic Hand and Upper Extremity
Service, Massachusetts General Hospital, Yawkey Center,
Suite 2100, 55 Fruit Street,
Boston, MA 02114, USA
e-mail: dring@partners.org

Materials and Methods

Study Design

A convenience sample of 1,209 prospectively enrolled patients was collected from a database with subjects from eight different prospective studies performed at our institution from 2006 to 2011. All studies were done under an IRB approved protocol and all patients provided informed consent to participate in the original study. We included studies where the DASH questionnaire [13] and the Pain Catastrophizing Scale (PCS) were administered [20, 27], and either the Center for Epidemiologic Studies Depression Scale (CES-D) [23] or the Patient Health Questionnaire (PHQ-9) [15]. In a portion of the cohort, also the Pain Anxiety Symptoms Scale-20 (PASS-20) [18] was administered and a pain scale. We excluded four patients because we did not have access to the primary data sheets and one additional patient due to a missing DASH questionnaire, leaving a cohort of 1,204 patients in the study. We defined an incomplete DASH as one or more questions left blank. When not all questions were completed, we considered the DASH incomplete. In line with the developers of the DASH, we defined an invalid DASH questionnaire when more than three questions were missing [13].

Patient Characteristics

The 1,204 patients included in the cohort had a mean age of 53 years and 69 % of the patients completed the DASH (Table 1). Education information was assessed in 848 patients (70 % of the total cohort).

Evaluation

Upper extremity-specific disability was measured with the 30-item DASH questionnaire [13]. A DASH score is valid for a maximum of three missing questions [13].

The PCS was used to measure catastrophic thinking [20, 27]. This questionnaire contains 13 items [20, 27]. Patients are asked to score how often a certain feeling occurred in the last week. A higher score indicates more catastrophic thinking [20, 27]. There are three subscales in the PCS: rumination, magnification, and helplessness [20, 27]. Thirteen patients had an invalid or missing PCS score, so the group mean was imputed. There were 56 missing questions in 50 patients; we imputed the mean of the patients' other questions for those missing values.

The CES-D was used to measure the level of depressive thoughts [23]. The CES-D was obtained from 745 patients in four different studies. This questionnaire has 20 items each answered on a four-point Likert scale [23]. The scores range from 0 to 60, with a higher score indicating more

depressive symptoms [23]. Seven hundred and forty-five patients were enrolled in a study where the CES-D was administered [23]. Twenty-nine patients did not complete the questionnaire and we imputed the group mean. Forty-seven patients had one missing question in the CES-D and we imputed the mean of the patients' other answers for those missing questions.

A group of 493 patients enrolled in five different studies where the PHQ-9 was evaluated [15]. The PHQ-9 is a questionnaire that measures symptoms of depression [15]. It consists of nine items scored on a four-point scale from 0 "not at all" to 3 "nearly every day" [15]. The scores range from 0 to 27 and higher scores indicate more depressive symptoms [15]. One patient did not complete this questionnaire and we imputed the group mean. There were 17 missing answers in 15 patients; we imputed the mean of the patients' other questions for those missing questions.

A group of 545 patients in three different studies completed the PASS questionnaire in addition to a depression questionnaire [18]. The PASS-20 is a 20-item questionnaire that assesses pain anxiety [18]. The questionnaire has four subscales: cognitive anxiety, fear of pain, escape and avoidance, and physiological anxiety [18]. Patients are asked how often they are engaged in certain activities or thoughts, and the items are scored on a six-point scale from 0 "never" to 5 "always" [18]. Total scores of the PASS-20 range from 0–100, and scores of 0–25 on each of the subscales are possible [18]. Higher scores indicate a higher level of pain anxiety [18]. Thirty-two questions were missing in 30 patients; we inserted the mean of the patients' other questions.

Pain was measured on an 11-point ordinal scale in 391 patients in four different studies. We imputed the mean pain score in 15 patients.

Statistical Evaluation

A power analysis indicated that 831 patients in the group with complete DASH and 373 with missing DASH questions provided 94 % power to detect an effect size of 0.22 in differences in the PCS, with a two-tailed *t* test.

We used a Pearson chi-square test to assess differences in the dichotomous and categorical variables (gender and diagnosis) between patients with a complete and incomplete DASH (yes/no). Independent samples *t* test were executed to detect differences in age, education, and the different questionnaires.

Pearson chi-square test and Fisher's exact test were used to assess categorical variables. Unequal variance *t* tests were used to detect differences in age, education, and the different questionnaires. The most frequent missing question in the DASH was assessed and an unequal variance independent samples *t* test was conducted to assess the differences in the

Table 1 Patient demographic, $n=1,204$

	N valid	Value	DASH complete Numbers (%)	DASH incomplete Numbers (%)	p	t	Valid DASH Numbers (%)	Invalid DASH Numbers (%)	p	t
		Total cohort								
Patients		1,204	831 (69 %)	373 (31 %)			1,152 (96 %)	52 (4 %)		
Age (years)	1,204									
Mean		53	50	58			52	61		
SD		16	16	16			16	16		
Range		18–90	18–86	19–90	<0.001	8.4	18–90	31–88	<0.001	3.7
Education (years)	848									
Mean		16	16	15			16	14		
SD		2.8	2.7	2.9			2.7	3.8		
Range		4–22	4–22	4–20	0.001	–3.3	4–22	4–20	0.021	–2.5
Gender	1,204					Chi				
Male		511 (42 %)	367 (72 %)	144 (28 %)			491 (96 %)	20 (4 %)		
Female		693 (58 %)	464 (67 %)	229 (33 %)	0.071		661 (95 %)	32 (5 %)	0.55	
Diagnosis	1,204									
Fracture		326 (27 %)	227 (70 %)	99 (30 %)	0.78		316 (97 %)	10 (3 %)	0.19	
Trigger finger		202 (17 %)	126 (62 %)	76 (38 %)	0.025	5.0	191 (95 %)	11 (5 %)	0.39	
Carpal tunnel syndrome		144 (12 %)	94 (65 %)	50 (35 %)	0.30		135 (94 %)	9 (6 %)	0.22	
Osteoarthritis		100 (8.3 %)	61 (61 %)	39 (39 %)	0.070		96 (96 %)	4 (4 %)	1.0	
Non-specific arm pain		46 (3.8 %)	35 (76 %)	11 (24 %)	0.29		44 (96 %)	2 (4 %)	1.0	
Other		386 (32 %)	288 (75 %)	98 (26 %)	0.004	8.3	370 (96 %)	16 (4 %)	0.84	

DASH Disabilities of the Arm, Shoulder and Hand

PCS score and age between the completers and non-completers of the most frequent missing question. A Pearson chi-square test was used to assess differences in gender between completers and non-completers of the most frequent missing question in the DASH.

Due to adequate sample size and to account for possible confounding, all variables that were available in all 1,204 patients and had a p value of less than 0.10 in the bivariate analysis were entered in a backwards, stepwise (likelihood) binary logistic regression analysis to assess the correlation of these variables with the outcome variable DASH complete (yes/no). We coded the categorical variable diagnosis with dummy variables with “other diagnosis” as the reference category and investigated the interaction terms of age with all separate diagnoses. To determine the predictors of an invalid DASH score, we inserted all variables with $p < 0.10$ in bivariate analysis in a backwards, stepwise (likelihood) binary logistic regression analysis.

Results

Evaluation Scores

Patients with an incomplete DASH were older, had fewer years of education, and had higher PCS, CES-D, and PASS

scores (Table 2). Patients with an invalid DASH (more than three missing questions) were older, had fewer years of education, and had higher PCS and CES-D scores (Table 2).

The most frequent missing question of the DASH was question 21 regarding difficulties with sexual activities. One hundred fifty patients (12 %) did not complete this question. The non-completers of question 21 were significantly older than the completers, had a significantly higher PCS score, and were more often female. (Table 3)

The five most frequent not-completed questions were: question 21 (“Sexual activities”)—150 times (12 % of the cohort); question 8 (“Garden or do yard work”)—54 times (4.5 %); question 19 (“Recreational activities in which you take some force or impact through your arm, shoulder or hand (golf, hammering, tennis, etc.)”)—49 times (4.1 %), question 27 (“Weakness in your arm, shoulder or hand”)—48 times (4.0 %), and question 5 (“Push open a heavy door”)—40 times (3.3 %).

The best logistic regression model for predictors of an incomplete DASH questionnaire included age ($\beta=0.035$, $p < 0.001$, OR 1.04, CI 1.03–1.05) and PCS ($\beta=0.034$, $p < 0.001$, OR 1.04, CI 1.02–1.05, constant $\beta_0=-3.5$) (Model fit; Hosmer and Lemeshow test: chi-square=8.4, $p=0.40$, Nagelkerke R-square=0.10) (Table 4). These findings mean that with each year of increased age or 1 point increase of the PCS, the odds to have missing questions in the DASH

Table 2 Outcome differences in patients with complete and valid/invalid DASH

	Total cohort	DASH complete	DASH incomplete	<i>p</i>	<i>t</i>	Valid DASH	Invalid DASH	<i>p</i>	<i>t</i>
PCS									
<i>N</i>	1,204	831	373			1,152	52		
Mean	6.8	6.2	8.2			6.7	9.7		
SD	8.4	8.0	9.1			8.2	11		
Range	0–39	0–39	0–39	<0.001	3.6	0–39	0–37	0.049	2.0
CES-D									
<i>N</i>	745	455	290			697	48		
Mean	10	9.4	12			10	12		
SD	9.3	9.1	9.5			9.4	6.8		
Range	0–55	0–55	0–46	0.002	3.1	0–55	1–36	0.045	2.0
PHQ-9									
<i>N</i>	493	395	98			486	7		
Mean	3.6	3.6	3.7			3.6	2.9		
SD	4.5	4.5	4.3			4.5	1.7		
Range	0–26	0–26	0–24	0.83		0–26	0–5	0.36	
PASS									
<i>N</i>	545	375	170			523	22		
Mean	22	20	24			22	23		
SD	18	18	19			18	18		
Range	0–89	0–83	0–89	0.018	2.4	0–89	0–61	0.66	
Pain									
<i>N</i>	391	299	92			384	7		
Mean	4.2	4.1	4.5			4.2	3.4		
SD	2.7	2.7	2.6			2.7	3.0		
Range	0–10	0–10	0–9	0.19		0–10	0–8	0.52	

DASH Disabilities of the Arm, Shoulder and Hand, *PCS* Pain Catastrophizing Scale, *CES-D* Center for Epidemiological Studies Depression Scale, *PHQ-9* Patient Health Scale-9, *PASS* Pain Anxiety Symptoms Scale

increases with 4 % (more or less comparable to a relative risk of 1.04 or 4 % more chance to be a non-completer).

The best logistic regression model for predictors of an invalid DASH questionnaire included age ($\beta=0.036$, $p<0.001$, OR 1.04, CI 1.02–1.06) and PCS ($\beta=0.038$, $p=0.007$, OR 1.04, CI 1.01–1.07, constant $\beta_0=-5.9$) (Model fit; Hosmer and Lemeshow test: chi-square=4.6, $p=0.80$, Nagelkerke R-square=0.055) (Table 5). In other words, the

odds to be a non-completer increase with 4 % per year or point on the PCS.

Discussion

Both age and catastrophic thinking were associated with significantly higher probability to have an incomplete or

Table 3 Most frequent missing question analysis of the DASH, $n=1,204$

Question 21 ^a	Completers of question 21 Mean $\pm$ SD	<i>N</i> =1,054 Range	Non-completers of question 21 Mean $\pm$ SD	<i>N</i> =150 Range	<i>p</i>	<i>t</i>
<i>T</i> test						
Age	51 $\pm$ 16	18–90	62 $\pm$ 15	20–89	<0.001	8.1
PCS	6.5 $\pm$ 8.1	0–39	9.0 $\pm$ 9.6	0–37	0.003	3.0
Chi-squared						Chi
Sex	Males (%)	44	31		0.003	8.7

DASH Disabilities of the Arm, Shoulder and Hand, *PCS* Pain Catastrophizing Scale

^a Question 21: sexual activities

Table 4 Logistic regression for an incomplete DASH questionnaire

Predictors of a incomplete DASH total cohort ^a				<i>n</i> =1,204	
	β	<i>p</i>	Odds ratio	95 % CI for odds ratio	
				Lower	Upper
Age	0.035	<0.001	1.04	1.03	1.05
PCS	0.034	<0.001	1.04	1.02	1.05
Age $\times$ fracture	0.005	0.076	1.01	1.00	1.01
Constant	$\beta_0 = -3.5$				

DASH Disabilities of the Arm, Shoulder and Hand, *PCS* Pain Catastrophizing Scale, *n* number of patients in the cohort

^a Variables included in the regression: age, sex, PCS, diagnoses (with dummy coding), age by all diagnoses

invalid DASH questionnaire. Our findings that patients with incomplete DASH score were older and had less years of education is in agreement with findings in literature [2, 16, 21, 22, 28]. Patients with incomplete questionnaires had higher scores on measures of depression and pain anxiety but did not have higher levels of pain.

The finding that the majority of patients only missed a few questions (e.g., only 52 patients had more than three missing questions on the DASH) is in line with the findings of Liu and colleagues [16]. The authors also found a similar percentage of missing questions [16]. The percentage of missing questions per study and questionnaire vary. For instance, Burns and colleagues found 0.5–95 % of items missing on the MMSE [2]. Other studies found 13 and 41 % missing items on the SF-36 [11, 19], 11 % missing items on the Expanded Prostate Cancer Index Composite-26 [4], 24 % on the MSKCC Bowel Function instrument [29], and 10 % on the SF-12 [21].

The question “difficulties with sexual activities” was the most frequent missing question in the DASH and patients who did not complete this question were older, female, and had more catastrophic thinking. The finding that older female patients are at risk for not completing the question about sexual activities was also reported in the study by Fayers and colleagues [7]. The authors reported that 19 % of patients left this question blank in the Rotterdam Symptom Checklist [7]. In our study, 150 patients (12 %) did not

complete this question, which was lower than the 29 % that Beaton and colleagues found in their study [1].

The majority of the most frequent unanswered questions are things that not everyone participates in, e.g., sex, gardening, and forceful recreational activities. This reinforces one of the advantages of outcome measures based on item response theory (IRT). This approach uses computerized adaptive tests, with which the next question depends on the answer of the previous question [3, 24]. The questions are ranked so that once a patient reports difficulty with a specific task, difficulty with all more difficult tasks is assumed. IRT-based questionnaires can arrive at a patient’s disability level with fewer questions [9, 10, 14, 24]. Questions that are not relevant to the patient can be bypassed [14].

Using IRT for the QuickDASH was described by Lyren and Atroshi, where a minimum number of six questions was required; in contrast to a minimum of ten questions in a traditional administration of the questionnaire [17]. This technique has also been used for the PHQ-9 [8] and the CTS-6 questionnaire [17].

This study should be viewed in the light of several shortcomings. Some comparisons might be underpowered. For instance, only 52 patients had invalid questionnaires. Our conclusions concerning education, depression (based on CES-D and PHQ-9), pain anxiety, and pain are based on subpopulations of 862, 745, 493, 545, and 391 patients, respectively, so are less strong than the conclusions for age

Table 5 Logistic regression for predictors of an invalid DASH questionnaire

Predictors of an invalid DASH total cohort ^a				<i>n</i> =1,204	
	β	<i>p</i>	Odds ratio	95 % CI for odds ratio	
				Lower	Upper
Age	0.036	<0.001	1.04	1.02	1.06
PCS	0.038	0.007	1.04	1.01	1.07
Constant	$\beta_0 = -5.9$				

DASH Disabilities of the Arm, Shoulder and Hand, *PCS* Pain Catastrophizing Scale, *n* number of patients in the cohort

^a Variables included in the regression: age and PCS

and catastrophic thinking. That is also the reason why we did not include those variables in the logistic regression model.

In conclusion, this study found that catastrophic thinking and age were risk factors for incomplete or invalid DASH scores. The finding that patient characteristics affect how questionnaires are completed is important for future research. The fact that 31 % of the patients did not complete the DASH score may be a reason to consider moving to computer adaptive testing, where the same information is obtained with fewer questions, subjects can bypass questions that are not relevant to them, and incomplete questionnaires are not possible.

Conflict of interest The authors declare that there are no conflicts of interest.

References

1. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87:1038–46.
2. Burns RA, Butterworth P, Kiely KM, et al. Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data. *J Clin Epidemiol*. 2011;64:787–93.
3. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63:1179–94.
4. Chang P, Szymanski KM, Dunn RL, et al. Expanded prostate cancer index composite for clinical practice: development and validation of a practical health related quality of life instrument for use in the routine clinical care of patients with prostate cancer. *J Urol*. 2011;186:865–72.
5. Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol*. 2003;56:968–76.
6. Faris PD, Ghali WA, Brant R, et al. Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *J Clin Epidemiol*. 2002;55:184–91.
7. Fayers PM, Curran D, Machin D. Incomplete quality of life data in randomized trials: missing items. *Stat Med*. 1998;17:679–96.
8. Fischer HF, Tritt K, Klapp BF, et al. How to compare scores from different depression scales: equating the Patient Health Questionnaire (PHQ) and the ICD-10-Symptom Rating (ISR) using Item Response Theory. *Int J Methods Psychiatr Res*. 2011;20:203–14.
9. Fries JF, Cella D, Rose M, et al. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol*. 2009;36:2061–6.
10. Fries JF, Krishnan E, Rose M, et al. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Res Ther*. 2011;13:R147.
11. Gandek B, Ware Jr JE, Aaronson NK, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. *International Quality of Life Assessment*. *J Clin Epidemiol*. 1998;51:1149–58.
12. Haukoos JS, Newgard CD. Advanced statistics: missing data in clinical research—part 1: an introduction and conceptual framework. *Acad Emerg Med*. 2007;14:662–8.
13. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med*. 1996;29:602–8.
14. Jenkinson C, Fitzpatrick R, Garratt A, et al. Can item response theory reduce patient burden when measuring health status in neurological disorders? Results from Rasch analysis of the SF-36 physical functioning scale (PF-10). *J Neurol Neurosurg Psychiatry*. 2001;71:220–4.
15. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606–13.
16. Liu H, Hays RD, Adams JL, et al. Imputation of SF-12 health scores for respondents with partially missing data. *Health Serv Res*. 2005;40:905–21.
17. Lyren PE, Atroshi I. Using item response theory improved responsiveness of patient-reported outcomes measures in carpal tunnel syndrome. *J Clin Epidemiol*. 2012;65:325–34.
18. McCracken LM, Zayfert C, Gross RT. The Pain Anxiety Symptoms Scale: development and validation of a scale to measure fear of pain. *Pain*. 1992;50:67–73.
19. Muller-Buhl U, Franke B, Hermann K, et al. Lowering missing item values in quality-of-life questionnaires: an interventional study. *Int J Public Heal*. 2011;56:63–9.
20. Osman A, Barrios FX, Gutierrez PM, et al. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. *J Behav Med*. 2000;23:351–65.
21. Perneger TV, Burnand B. A simple imputation algorithm reduced missing data in SF-12 health surveys. *J Clin Epidemiol*. 2005;58:142–9.
22. Peyre H, Coste J, Leplege A. Identifying type and determinants of missing items in quality of life questionnaires: application to the SF-36 French version of the 2003 Decennial Health Survey. *Health Qual Life Outcomes*. 2010;8:16.
23. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1:385–401.
24. Rose M, Bjorner JB, Becker J, et al. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2008;61:17–33.
25. Ross L, Thomsen BL, Boesen EH, et al. In a randomized controlled trial, missing data led to biased results regarding anxiety. *J Clin Epidemiol*. 2004;57:1131–7.
26. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
27. Sullivan MPJ. The Pain Catastrophizing Scale: development and validation. *Psychol Assess*. 1995;7:525–32.
28. Vercherin P, Gutknecht C, Guillemin F, et al. [Missing data mechanisms of the questionnaire SF-36's items in the SU.VI.MAX study]. *Rev Epidemiol Sante Publique*. 2003;51:513–25.
29. Zotti P, Del Bianco P, Serpentin S, et al. Validity and reliability of the MSKCC Bowel Function instrument in a sample of Italian rectal cancer patients. *Eur J Surg Oncol*. 2011;37:589–96.