

<sup>1</sup> Institute for Social Medicine, University Clinics Schleswig-Holstein, Luebeck, Germany

<sup>2</sup> Institute for Community Medicine, Ernst-Moritz-Arndt-University, Greifswald, Germany

## Missing data due to a 'checklist misconception-effect'

Submitted: 4 January 2005

Accepted: 25 August 2005

### Summary

**Objectives:** To analyze the extent and relevance of a postulated "checklist misconception-effect" (a specific response pattern characterized by symptom-free persons not checking the "not at all"-category).

**Methods:** Our data is derived from a survey of blue collar workers ( $n = 228$ ) who previously had filed in applications for medical rehabilitation benefits. We defined the "checklist misconception-effect" by the following response pattern: (1) at least one missing value and (2) at least one valid item response and (3) no 'not at all' responses.

**Results:** 75 % of the responders had complete data, 16.2 % a postulated 'checklist misconception-effect'. Substantial co-variations with socio-demographic characteristics or health status indicators could not be found. Additional imputation of missing values under the assumption of a "checklist misconception-effect" led to a reduction of missing data in the somatisation-subscale score from 12.3 % to 0.4 % compared to a simple manual-based calculation. Correlation with various external criteria (general health perception, level of functioning, depression) remained unchanged.

**Conclusions:** Ignoring the "checklist misconception-effect" would overestimate symptom load. However, the validity of this effect has still to be proven in methodological studies.

**Keywords:** Missing data – Imputation – SCL-90-R – Somatisation.

Missing data due to item non-response are a wide-spread problem in survey, clinical, and social research. There are widely acknowledged standards to report the loss of subjects from the sample (unit non-response) and the reasons for their non-response, e.g. in clinical trials (CONSORT-statement,

Moher et al. 2001) and, to a lesser degree, in survey research (Smith 2002) or epidemiological research (Arbeitsgruppe Epidemiologische Methoden 2004). However, no comparable widely acknowledged standards exist for dealing with *item* non-response. These non-responses, too, lead to a loss of efficiency in estimations, a more extensive data-handling because of incomplete data matrices and the risk of biased parameter estimates.

In practice, missing data are most often dealt with in terms of 'listwise' or 'pairwise deletion'. These imputation procedures assume that the reasons for the missingness of data are completely random (Missing-Completely-At-Random, MCAR) (Schafer & Graham 2002; Wirtz 2004). The resulting reduction of sample size leads to a decrease of statistical power. Also, the assumptions of an MCAR mechanism are often not met. If the missingness mechanism is related to the data collected in the study, and therefore can be explained in part by the available data, this is called Missing-At-Random (MAR). Otherwise, if the missingness is related to variables not observed in the data set, we have to assume a Missing-Not-At-Random (MNAR) mechanism<sup>1</sup>. MAR and MNAR are a major threat to the validity of study results, e.g. due to bias because of systematic exclusion of subjects. Especially the validity of research based on multivariate analyses is called into question.

The selection of appropriate imputation strategies depends on these assumptions regarding the mechanisms of missingness. Therefore, it appears to be essential to develop hypotheses about the emergence of item-nonresponse which could in

<sup>1</sup> MCAR refers to data where the missingness mechanism does not depend on the variable of interest, or any other variable, which is observed in the data set. MAR means that the missingness is conditional on some other variable(s) observed in the data, although not on the variable of interest. MNAR (also referred to as NMAR) occurs when the missingness mechanism depends on the actual value of the missing data and is not conditional on other variables in the data set.

turn guide the selection of imputation strategies. Our study is based on the observation of a particular response pattern in a survey of blue collar workers. In the subscale 'somatisation' of the Symptom Check List 90-R (SCL-90-R) (Derogatis 1977; German version by Franke 2002) some respondents confirmed the presence of a few symptoms but did not check the remaining symptoms (Fig. 1). We offer a possible explanation for this response pattern: these subjects viewed this part of the questionnaire as a 'checklist', in which 'true' items are checked, i.e. symptoms were confirmed, but symptoms that did not apply to the subjects' present state were missed out. In other words, the subjects misconceived the questionnaire as a 'checklist'. Instead of checking the 'not at all'-category, they missed out their answer. Thus, the subject in the example presented in Figure 1 may have suffered from occasional headaches, heart or chest pain, low back pain, sore muscles, and feelings of weaknesses in parts of the body, but never experienced any of the other problems. Therefore, we hypothesize, he or she simply left out the other items. If this response mechanism would prove to be true, an MCAR could not be assumed. Therefore, there is a threat of bias in the analysis. The present study was set up to explore indications for the presence of a 'checklist misconception effect' based on a small-scale survey on blue-collar workers.

It aims at

- quantifying the prevalence of the postulated 'checklist misconception-effect'<sup>2</sup> in the survey;

- identifying this effect in other scales of the same questionnaire;
- analyzing a possible dependence of the 'checklist misconception-effect' on subjects' characteristics, and finally
- comparing different imputation procedures.

### Methods

The analyses were based on a data set of a larger project in rehabilitation research. This study was set up as a postal survey of n = 312 blue collar workers who previously had filed applications for medical rehabilitation benefit primarily due to back pain. Participation in the survey was voluntary. The responses to the survey were not related to obtaining any kind of rehabilitative or social benefits.

Besides the twelve items of the somatisation subscale of the SCL-90-R (Franke 2002), the questionnaire included a 1-item-question on general health status (5-step response format), questions on intensity, duration, and treatment of back pain, the assessment of pain within the last seven days in accordance with the national health survey (Robert Koch Institut 1998), health insurance status and an appraisal of personal work capacity, questions on health-related behaviours, a questionnaire on functional capacities (Funktionsfragebogen Hannover (FFbH)) (Kohlmann & Raspe 1996), rumination on health-related complaints (PRSS) (Flor & Turk 1988), the subscale 'vitality' of the short-form 36 questionnaire (SF-36) (Bullinger & Kirchberger 1998), the CES-D for the assessment of the degree of depressive symptoms (Kohlmann & Gerbershagen 1998), and socio-demographic characteristics. The items of the somatisation subscale were localized in the beginning section of the questionnaire. There, a 'checklist misconception-effect' could be assumed because of other response formats used in this section. In order to harmonize the response formats of the whole questionnaire the 'not at all'-category was coded as '1' instead of '0', which deviates from the specification of the original SCL-90-R questionnaire (see Fig. 1).

The 'checklist misconception-effect' was defined here by a response pattern with (1) at least one missing value, (2) at least one valid response, and (3) no 'not at all'-responses in any item of the subscale.

In Table 1, the number of all valid and positive responses (i.e. all responses with the 'not at all' category) is cross-tabled with the number of missing data. This points towards the relevance of the 'checklist misconception-effect' compared to other forms of missing data. Most persons with missing data are found in the highlighted diagonal which comprises all persons with a possible 'checklist misconception-effect'. In persons with complete missing data (last column of Tab. 1) it is impossible to distinguish those who have not answered this

In the last 7 days, how much have you been distressed by...?					
	Not at all	A little bit	Moderately	Quite a bit	Extremely
headaches.....	[1]	[2]	<del>[3]</del>	[4]	[5]
faintness or dizziness.....	[1]	[2]	[3]	[4]	[5]
pains in the heart or chest.....	[1]	<del>[2]</del>	[3]	[4]	[5]
pains in the lower back.....	[1]	[2]	[3]	<del>[4]</del>	[5]
nausea or upset stomach.....	[1]	[2]	[3]	[4]	[5]
soreness of your muscles.....	[1]	[2]	<del>[3]</del>	[4]	[5]
trouble getting your breath.....	[1]	[2]	[3]	[4]	[5]
hot or cold spells.....	[1]	[2]	[3]	[4]	[5]
numbness or tingling in parts of your body..	[1]	[2]	[3]	[4]	[5]
a lump in your throat.....	[1]	[2]	[3]	[4]	[5]
feeling weak in parts of your body.....	[1]	[2]	<del>[3]</del>	[4]	[5]
heavy feelings in your arms or legs.....	[1]	[2]	[3]	[4]	[5]

**Figure 1** An example of a response pattern classified as a 'checklist misconception-effect'

<sup>2</sup> At first we had named the effect 'checklist-effect' (Mittag et al. 2005). One of the reviewers pointed out that the term was already in use in the medical informatics literature (Wyatt 1997). Therefore, we decided to rename the effect as 'checklist misconception-effect' which might capture the nature of the effect also more adequately.

**Table 1** Number of missing item responses in the somatisation subscale of the SCL-90-R in relation to the number of valid positive responses ('checklist misconception-effect' highlighted in the diagonal)

		number of missing item responses											
		1	2	3	4	5	6	7	8	9	10	11	12
number of valid positive responses <sup>a)</sup>	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	4	
	2	0	0	0	0	0	0	0	0	0	4		
	3	0	1	1	0	0	0	0	1	6			
	4	3	2	0	0	0	0	0	4				
	5	2	0	0	0	0	0	6					
	6	1	0	0	0	0	3						
	7	1	3	0	0	0							
	8	2	0	1	1								
	9	2	0	1									
	10	0	3										
	11	5											

<sup>a)</sup> all valid responses without the 'not at all'-responses

part of the questionnaire at all ('true' missing values) from those without symptoms in terms of a 'checklist misconception-effect'. Therefore, the imperative presence of at least one valid answer might slightly underestimate the true prevalence of the checklist misconception-effect. However, this problem is of no relevance for the present analysis because there is at least one valid answer in the items of the somatisation subscale from each participant.

Differences between persons with divergent response patterns were tested depending on scale level and distribution characteristics by means of a  $\chi^2$ -test, an analysis of variance or a Kruskal-Wallis-test. The following imputation procedures were compared: (a) manual-based (i. e. computation of scale mean if no more than four item responses were missing), (b) manual-based (see (a)) with additional imputation of the remaining missing subscale data by the group mean, (c) a maximum likelihood estimation (EM-algorithm, Wirtz 2004) of the somatisation subscale score based on the score values of the CES-D, FFbH, PRSS, the vitality subscale of the SF-36, and age; and (d) imputation of missing data in the presence of a 'checklist misconception-effect' through a 'not at all' response and a subsequent computation of the scale mean according to (a).

In order to analyse the bivariate relationships between the somatisation subscale scores computed on the basis of the different imputation models and external criteria Pearson correlation coefficients were computed. Because of the mutual dependence of these correlation coefficients, significance levels are not presented. All computations were done with SPSS for Windows (Version 11.0.1) except for the maximum-likelihood estimation to impute missing values (imputation

model c) which was done with NORM for Windows (Version 2.03) by J.L. Schafer (<http://www.stat.psu.edu/~jls/missoftwa.html#win>).

**Table 2** Characteristics of the sample

	M (SD) or frequency(%)
<b>age</b>	M = 50.06 (SD = 6.48)
<b>sex</b>	
men	167 (73.2 %)
women	61 (26.8 %)
<b>family status</b>	
single	25 (11.0 %)
married	172 (75.4 %)
divorced/ separated	26 (11.4 %)
widowed	5 ( 2.2 %)
<b>education</b>	
secondary school ('Hauptschule')	178 (78.1 %)
secondary school ('Realschule')	26 (11.4 %)
university or technical college entrance diploma	11 ( 4.8 %)
other kinds of graduation	2 ( 0.9 %)
no school graduation	11 ( 4.8 %)
<b>employment</b>	
full-time	153 (68.9 %)
at least half-time	16 ( 7.2 %)
less than half-time	5 ( 2.3 %)
housewife / house husband	4 ( 1.8 %)
unemployed	14 ( 6.3 %)
others	30 (13.5 %)
not specified	6
<b>work status</b>	
blue-collar	194 (85.8 %)
white-collar	28 (12.4 %)
self-employed	3 ( 1.3 %)
others	1 ( 0.4 %)
not specified	2

M = mean, SD = standard deviation

**Table 3** Prevalence of different types of missing data in three different scales (n = 228)

type of missing data	SCL-90-R somatisation subscale		pain items		CES-D depression	
	n	%	n	%	n	%
no missing data	171	75.0	188	82.5	198	86.8
complete missing data <sup>1)</sup>	0	0	1	0.4	3	1.3
„checklist misconception-effect“	37	16.2	26	11.4	4	1.8
other types of missing data	20	8.8	13	5.7	23	10.1

<sup>1)</sup> no valid responses

**Table 4** Socio-demographic and clinical differences between persons with different types of missing data in the somatisation subscale of the SCL-90-R

	type of missing data			test on difference between groups
	„checklist misconception-effect“ (n = 37)	other missing data (n = 20)	no missing data (n = 171)	
<b>Sex male</b>	30 (81.1 %)	12 (60.0 %)	125 (73.1 %)	$\chi^2 = 2.9$ ; df = 2 p = .229; $V_c = .11$
<b>age</b>	M = 49.8 SD = 6.3	M = 50.3 SD = 6.7	M = 52.0 SD = 7.6	F = 0.27; $df_{betw} = 2$ ; $df_{within} = 220$ ; p = .36 $\eta^2 = 0.009$
<b>school education</b>				$\chi^2 = 3.7$ ; df = 2 p = .154; $V_c = .13$
lower <sup>a)</sup>	28 (75.7 %)	15 (75.0 %)	148 <sup>b)</sup> (86.5 %)	
medium or higher <sup>c)</sup>	9 (24.3 %)	5 (25.0 %)	23 (13.5 %)	
<b>health status</b>				$\chi^2 = 3.395$ ; df = 6; p = .758; $V_c = .09$
good	0	1 (5.0 %)	2 (1.2 %)	
satisfactory	7 (20.6 %)	4 (20.0 %)	30 (17.6 %)	
not good	20 (58.8 %)	11 (55.0 %)	110 (64.7 %)	
bad	7 (20.6 %)	4 (20.0 %)	28 (16.5 %)	
<b>vitality</b> (SF-36, 0-100)	M = 36.6 SD = 19.9	M = 37.8 SD = 21.2	M = 39.0 SD = 17.9	F = 0.27; $df_{betw} = 2$ ; $df_{within} = 212$ ; p = .76; $\eta^2 = 0.002$
<b>depression</b> (CES-D, 0-60)	M = 19.9 SD = 10.6	M = 16.3 SD = 7.7	M = 17.1 SD = 8.8	F = 1.44; $df_{betw} = 2$ ; $df_{within} = 218$ ; p = .239; $\eta^2 = 0.013$

M = mean, SD = standard deviation,  $V_c$  = Cramér's V

<sup>a)</sup> no or secondary school education ('Hauptschule')

<sup>b)</sup> including two subjects with 'other type of education'

<sup>c)</sup> secondary school education ('Realschule') or higher

## Results

Between June 26, 2002 and March 25, 2003 questionnaires were sent to a total of n = 312 subjects who previously had filed applications for medical rehabilitation benefits primarily due to back pain. The response-rate was 73 % (n = 228). Major demographic characteristics of the sample are presented in Table 2. There were no indications of a non-response bias with regard to age, sex, need for rehabilitation and the type of legal institution that reviewed the application.

The items of the somatisation subscale were completely filled in by 75.0% of the 228 study participants, i. e. without any

missing data. According to the 'checklist misconception-effect' response pattern, 16.2% of the respondents had missing data; an additional 8.8% had other types of missing data (Tab. 3; see also Tab. 1).

In order to test how stable the 'checklist misconception-effect' is within the questionnaire we compared the 'checklist misconception-effect' in the somatisation subscale and in eleven pain items with an identical response format which directly succeeded the somatisation subscale items (see Tab. 3). Within these pain items we found a 'checklist misconception-effect' in 11.4% of the respondents. All of those had a 'check-

**Table 5** Scores of the somatisation subscale of the SCL-90-R according to different imputation models (cf. text) and correlation with external criteria (Pearson coefficient)

	(a) manual-based	(b) manual-based + group mean	(c) maximum likelihood imputation <sup>a)</sup>	(d) ,checklist misconception-effect' <sup>b)</sup>
<b>N</b>				
valid	200	228	228	227
missing	28	0	0	1
<b>mean</b>	1.043	1.043	0.963	0.979
<b>median</b>	0.833	0.917	0.833	0.833
<b>standard deviation</b>	0.665	0.623	0.581	0.605
<b>correlation with external criteria</b>				
<b>depression (CES-D)</b>	.531; n = 190	.504; n = 215	.536; n = 215	.496; n = 215
<b>ruminatation (PRSS)</b>	.437; n = 193	.425; n = 216	.436; n = 216	.446; n = 215
<b>functional capacity (FFbH-R)</b>	-.402; n = 200	-.373; n = 226	-.413; n = 226	-.391; n = 225
<b>vitality (SF-36)</b>	-.335; n = 195	-.317; n = 222	-.370; n = 221	-.304; n = 221

<sup>a)</sup> EM-algorithm

<sup>b)</sup> imputation procedure: 1. in every individual who fulfilled the criteria for a 'checklist misconception-effect' response pattern all missing data were recoded into 'not at all' responses, 2. computation of scale score according to the algorithm provided by the manual (i. e. allowing for four missing values)

list misconception-effect' response pattern in the somatisation subscale, too. The CES-D, which was located at a different position of the questionnaire, has a similar but not identical response format<sup>3</sup>. Here, we found a 'checklist misconception-effect' response pattern in four respondents only (1.8%). Two of these four subjects had a 'checklist' response pattern in the somatisation subscale, too.

Persons with a 'checklist misconception-effect' response pattern did not differ from subjects with complete data or data missing due to other reasons with regard to age, gender, years of schooling, general health status, functional status, depression and vitality (Tab. 4). However, this comparison had low statistical power due to small sample sizes in the subgroups with missing data. Therefore, an effect size estimator was calculated. No substantial differences seemed to be present (Tab. 4). A low association between education and the presence of missing data might be existent, substantial covariations indicative of a MAR-mechanism based on these sociodemographic and health status variables had to be discarded.

Results of the sensitivity analysis are shown in Table 5. The mean of the somatisation subscale of the SCL-90-R is presented for the different imputation models. The manual of the SCL-90-R requires the exclusion of cases with more than four missing values in the somatisation subscale (listwise deletion)

(Franke 2002). Up to four missing values are imputed by the individual mean (column a).

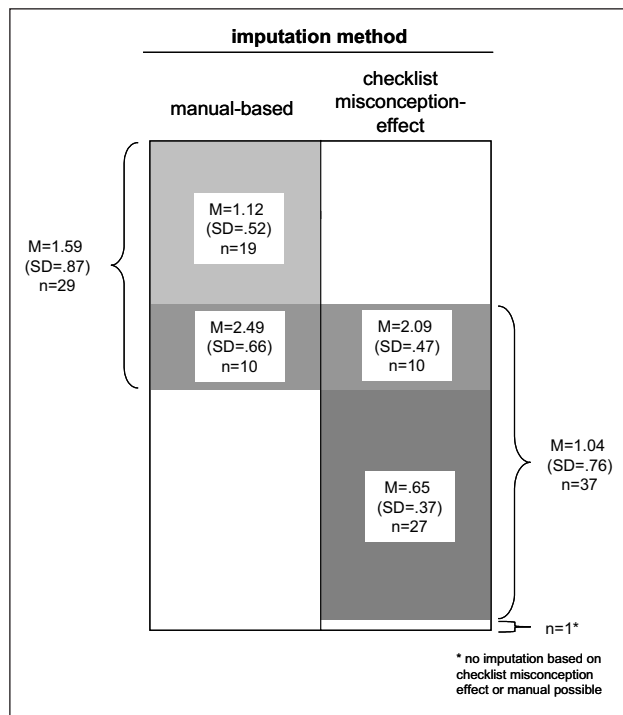
In column d, imputation of missing data was done according to the 'checklist misconception-effect'. In a first step, in every case fulfilling the criteria for a 'checklist misconception-effect' response pattern, all missing data were recoded into 'not at all' responses. In a second step the scale score was computed on the basis of the algorithm provided by the manual (i. e. allowing for addition four missing values).

Two additional imputation procedures are shown in columns b and c: imputation of the subscale score by a constant group mean if more than four missing values were present (column b) and a maximum likelihood estimation by an EM-algorithm (column c).

Manual-based imputation (column a) discarded 28 cases with more than four missing values. The subscale's mean score was 1.04. It is obvious that this mean score remained unchanged when it was imputed by the group mean in those 28 subjects with more than four missing responses, and that the standard deviation was lower (column b).

The imputation based on the 'checklist misconception-effect' according to model d resulted in a reduction of missing scale scores from 12.3% (as in the manual-based method in column 1) to 0.4% (column d). It should be noted that treating missing data only in terms of the 'checklist misconception-effect' would leave n = 20 persons (8.8%) without valid score on the somatisation subscale (Fig. 2). In comparison with the imputation models in columns a and b the 'checklist misconception-effect' imputation yielded the lowest subscale score.

<sup>3</sup> 1 – none or almost none of the time, 2 – some of the time, 3 – most of the time, 4 – all / almost all of the time; reference time of responses are the last seven days, the instruction relates the response categories to a specific number of days within the last week, e. g. some of the time = on 1 to 2 days.



**Figure 2** Comparison of imputation according to manual and according to the 'checklist misconception-effect' in all persons with missing data in the somatisation subscale of the SCL-90-R (n = 57) (mean, standard deviation and number of persons of the subgroups)

Those n = 27 subjects who were additionally included from model a to model d had a subscale score of 0.65 (SD = 0.37), i. e. markedly lower scores (Fig. 2). Subscale mean and standard deviation of the checklist imputation were comparable to the results of the maximum likelihood estimation that also yielded a lower value.

The correlations of various external criteria and different values of the somatisation subscale derived by different imputation methods do not show a clear pattern (tab. 5). They do not differ much from the manual- based imputation, though.

Of those individuals with a 'checklist misconception-effect' response pattern the correlations of the somatisation subscale (based on the imputed values) and the variables analysed in table 5 were  $r = .523$  ( $p = .002$ ;  $n = 32$ ) with depression (CES-D),  $r = -.476$  ( $p = .004$ ;  $n = 35$ ) with functional capacity (FFbH-R),  $r = .608$  ( $p = .000$ ;  $n = 32$ ) with the rumination (PRSS),  $r = -.194$  ( $p = .265$ ,  $n = 35$ ) with vitality (SF-36), and  $r = .521$  ( $p = .002$ ;  $n = 34$ ) with overall health status.

## Discussion

Missing data in survey research present a major challenge to applied social and epidemiological research, especially

if multivariate analyses are planned for. Since the 1970s a number of elaborate imputation methods for missing data have been developed (Schafer & Graham 2002), e.g. maximum-likelihood estimation, hot-deck-imputation or multiple imputation. Any of these methods are based on general assumptions on the nature of the missingness of the data. In this paper we introduce a specific hypothesis on how a certain pattern of missing item-responses might occur.

The starting point of the present analyses was a close inspection of the raw data. The postulated 'checklist misconception-effect' was prevalent in more than every ninth subject of our sample and presents a relevant problem. If this effect proves to be true, a manual-based imputation of missing data would exclude subjects with low symptom load systematically and result in misclassification of subjects. This will result in a bias towards an overestimation of complaint level in the sample. The degree of this bias will depend on the proportion of persons with a 'checklist misconception-effect'. An exploratory re-analysis of three other epidemiological data-sets yielded a substantial variation in the proportion of subjects with a 'checklist misconception-effect' (Mittag et al. 2005). The present study was not set up to prove the existence of the effect; it illustrates possible consequences for dealing with item-non-response if the checklist misconception-effect is assumed. Therefore, further studies on this phenomenon are needed.

The results of this study show, firstly, that in about every sixth respondent a 'checklist misconception-effect' could be assumed, and almost all missing data could be imputed on a substantive basis, taking the 'checklist misconception-effect' into account. Secondly, a 'transfer' of the effect from one scale to another could depend on the sequence of the questions within the questionnaire. Thirdly, the effect cannot be expected to depend on patient characteristics, and fourthly, the subscale mean will be reduced through imputation of missing values assuming a 'checklist misconception-effect', which was in line with the results of the maximum likelihood estimation procedure. It should be kept in mind, however, that the impact of the imputation based on the 'checklist misconception-effect' should strongly be influenced by the population under study (e.g. population survey with low symptom load vs. clinical samples).

How do these assumptions with regard to a possible checklist misconception-effect relate to the general missing data mechanisms in terms of MCAR, MAR and NMAR? The answer is not straightforward. On the one hand, the perspectives are different. General missing data mechanisms are concerned with the degree of bias in parameter estimation due to missing data and therefore with distributional characteristics based on the whole sample. The specific hypothesis introduced in this paper relies on a psychological explanation of individual re-

spondents' behaviour. It could possibly be applied directly to a diagnostic situation. In addition, the present focus is on the missingness of individual items of a scale that constitute the scale score, while from the perspective of the general missing data mechanisms the appropriate imputation of the scale score would be prioritized. With regard to missing data of the individual items, in the checklist misconception-effect the probability of the response to be missing depends on the true value of the response, i. e. the 'not at all'-response, implying NMAR. From the perspective of the individual person and their respective scale scores, we do not have a clear indication whether certain individuals are more prone to the checklist misconception-effect than others. There are indications of a systematic transition of the effect from one part of the questionnaire to succeeding ones. In addition, the introduction of a cut-off point of a maximum of four missing item responses allowed for computation of the scale score introduces a systematically different treatment of responses below and above this cut-off value, which could partly be dependent on the variable of interest (i. e. the somatisation scale score).

On a scale score level and taking into account other variables observed in the data set, it should be considered whether the missingness of the scale score was conditional on the other variables observed, implying Missing-At-Random (MAR). The assumption of a MAR might not hold true because there was no indication of systematic relationships of the 'checklist misconception-effect' with subjects' demographic or clinical characteristics. Therefore, the application of more elaborate imputation methods, e. g. maximum-likelihood estimation or multiple imputation (Schafer & Graham 2002; Wirtz 2004) is not necessarily first choice, although these techniques have proved to be quite robust against violations of MAR-assumptions (Collins et al. 2001). In the present case, we know that the level of somatisation is related to other health status indices (general health, depression, vitality etc.), therefore the unknown level of somatisation might be adequately inferred even under MAR-assumptions. The similar results of the EM-algorithm and the checklist misconception-effect imputation with regard to the level of somatisation in the sample might be taken as a possible indication for both the adequacy of the MAR-assumption and the checklist misconception-effect. In addition, in those persons with a 'checklist misconception-effect' response pattern the correlations between the somatisation subscale and other external criteria (depression, functional capacity etc.) were in the expected direction and size, indicating validity of the effect.

However, the results we present here do not prove the existence of the 'checklist misconception-effect'; instead it can only rely on plausible assumptions. More research has to be done before we should recommend to identify possible

'checklist misconception-effect'-missings in data sets and to re-analyse data in terms of a sensitivity analysis.

Hypotheses on the emergence of the 'checklist misconception-effect' might be deduced from the analysis on the stability of the effect in different parts of the questionnaire. A possible explanation should take 'transfer-effects' into account. The 'checklist misconception-effect' was possibly transferred from the items of the somatisation subscale only to the directly succeeding pain items. The CES-D was hardly at all affected by the 'checklist misconception-effect', possibly because of different preceding response formats. Another explanation might be found in the contents of the item themselves. Questions on different bodily symptoms might be more prone to checklist-responses compared to the assessment of carefully worded psychological states. The deviating response format of the SCL-90-R items mentioned above ('not at all' was coded by a '1' instead of a '0') might serve as another possible explanation. However, a first exploratory comparison with data sets that used the response format specified in the original SCL-90-R questionnaire did not support this assertion. It should be noted that viewing the somatisation subscale of the SCL-90-R as a checklist might have been encouraged by presenting the subscale in one block, which is known to change psychometric characteristics of the scale (Franke 1995; Krampen et al. 1992; cf. Mittag et al. 2005).

The existence of a 'checklist misconception-effect' is still hypothetical, although not unlikely. A valid investigation of the effect needs to be targeted by methodological studies that analyse response behaviour in subjects in an experimental setting with different response options (e. g. 'true' checklist question vs. standard format). Another approach would be the application of cognitive survey techniques to identify the thoughts and action-related cognitions of subjects while filling out the questionnaire (Mittag et al. 2003).

#### *Acknowledgements*

This paper is part of the project 'Rehabilitation needs assessment in a sample of statutory pension insurees: validation of the Luebeck Algorithm' ('Die Abschätzung von Rehabilitationsbedarf bei aktiven Mitgliedern der Gesetzlichen Rentenversicherung: der Lübecker Algorithmus und seine Validierung', A1-project, principal investigator: Prof. Dr. Dr. H. Raspe) within the framework of the Northern German Network for Rehabilitation Research (NVRF) in the Joint Research Programme 'Rehabilitation Science' by BMBF and the German Pension Insurance (VDR) (FKZ: 02 1 06).

We are indebted to Prof. Heiner Raspe for his advice and discussions preparing this paper, Birgit Sprengel and Dr. Natalie Glaser-Möller for help in translation.

## Zusammenfassung

### Fehlende Werte durch den 'Listenfragen-Fehldeutungseffekt'

**Fragestellung:** Item non-response kann die Gültigkeit von Studienergebnissen erheblich beeinträchtigen. Am Beispiel der Subskala Somatisierung der SCL-90-R postulieren wir ein Antwortmuster ('Listenfragen-Fehldeutungseffekt'), bei dem die Probanden bei Beschwerdefreiheit statt der vorgesehenen 'überhaupt nicht'-Kategorie keine Antwort markieren. Die vorliegende Studie analysiert Ausmaß und Bedeutung dieses 'Listenfragen-Fehldeutungseffekts'.

**Methoden:** Als Analysegrundlage dient eine postalische Befragung von  $n = 228$  Versicherten einer Arbeiterrentenversicherung mit Antrag auf medizinische Rehabilitation. Der 'Listenfragen-Fehldeutungseffekt' wurde operationalisiert durch Antwortmuster (1) mit mindestens einem fehlenden Wert und (2) mindestens einem gültigen Wert und (3) Fehlen von 'überhaupt nicht'-Antworten.

**Ergebnisse:** 75 % der Befragten hatten komplette Daten, 16,2 % einen postulierten 'Listenfragen-Fehldeutungseffekt'. Substanzielle Kovariationen mit soziodemografischen Merkmalen und Gesundheitsindikatoren wurden nicht gefunden. Durch zusätzliche Imputation der fehlenden Werte unter der Annahme des 'Listenfragen-Fehldeutungseffekts' reduzierte sich der Anteil fehlender Skalenwerte im Vergleich zur einfachen Manual-gestützten Vorgehensweise von 12,3 % auf 0,4 %. Korrelationen mit verschiedenen Außenkriterien (Gesundheitszustand, Funktionsfähigkeit, Depressivität) blieben konstant.

**Schlussfolgerungen:** Die Nicht-Berücksichtigung des 'Listenfragen-Fehldeutungseffekts' führt zu einer Überschätzung der Symptomausprägung. Die Gültigkeit des Effekts muss in weiteren methodischen Studien überprüft werden.

## Résumé

### Non-réponses dues à une mauvaise conception du questionnaire

**Objectifs:** Analyser l'importance d'un "checklist misconception-effect" (type de réponses dans lequel le sujet sans symptômes ne coche pas de réponse au lieu de cocher la réponse «pas du tout»).

**Méthodes:** Les données sont tirées d'une enquête par questionnaire postal auprès d'assurés ( $n = 228$ ) d'une caisse de retraite ayant présenté un dossier pour un traitement de réadaptation. Le type de réponse dont il a été tenu compte pour identifier le "checklist misconception-effect" a été défini comme présentant (1) au moins une réponse manquante, (2) au moins une réponse valable (3), et aucune réponse «pas du tout».

**Résultats:** 75 % des personnes interrogées ont répondu à toutes les questions; chez 16,2 % des personnes on a pu suspecter un "checklist misconception-effect". Nous n'avons pas trouvé de covariations substantielles avec des aspects sociodémographiques ou des indicateurs de santé. Partant de l'hypothèse que ce "checklist misconception-effect" existe, les valeurs manquantes peuvent être remplacées et le nombre de réponses manquantes diminue par rapport au procédé conventionnel de 12,3 % à 0,4 %. Les corrélations avec divers critères (état de santé, vitalité, tendance dépressive) restent alors constantes.

**Conclusions:** En ne tenant pas compte de ce "checklist misconception-effect" on induit une surestimation de l'intensité des symptômes. La validité de cet effet reste à vérifier par d'autres études.

## References

Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (2004). Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP), Langversion. <http://medweb.uni-muenster.de/institute/epi/dae/>.

Bullinger M, Kirchberger I (1998). SF-36 Fragebogen zum Gesundheitszustand – Manual. Göttingen: Hogrefe.

Collins LM, Schafer JL, Kam CK (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychol Methods* 6: 330–51.

Derogatis LR (1997). SCL-90-R, administration, scoring & procedures manual – I for the R(evised) version. John Hopkins University School of Medicine.

Flor H, Turk CD (1988). Chronic back pain and rheumatoid arthritis: predicting pain and disability from cognitive variables. *J Behav Med* 11: 251–65.

Franke GH (1995). SCL-90-R. Göttingen: Beltz Test.

Franke GH (2002). SCL-90-R (2., vollständig überarbeitete und neu normierte Auflage). Göttingen: Beltz Test.

Kohlmann T, Gerbershagen HU (1998). Center of Epidemiological Studies Depression Scale (German version) (CES-D). [http://www.drk-schmerz-zentrum.de/content/07\\_infos/documents/pdf/CES-D.pdf](http://www.drk-schmerz-zentrum.de/content/07_infos/documents/pdf/CES-D.pdf).

Kohlmann T, Raspe H (1996). Der Funktionsfragebogen Hannover zur alltagsnahen Diagnostik der Funktionsbeeinträchtigung durch Rückenschmerzen (FFbH-R). *Rehabilitation* 35: I–VIII.

Krampen G, Hense H, Schneider JF (1992). Reliabilität und Validität von Fragebogenskalen bei Standardreihenfolgen versus inhaltshomogener Blockbildung ihrer Items. *Zeitschrift für Experimentelle und Angewandte Psychologie* 34: 229–48.

Mittag O, Böhmer S, Deck R et al. (2003). Fragen über Fragen: cognitive survey in der Fragebogenentwicklung. *Soz Präventiv Med* 48: 55–64.

Mittag O, Meyer T, Deck R et al. (2005). Fehlende Werte in der Somatisierungsskala des SCL-90-R durch „Listenfrageneffekt“: Häufigkeit in verschiedenen Surveys und Vergleich unterschiedlicher Imputationsmethoden. *Gesundheitswesen* 67: 120–3.

Moher D, Schulz KF, Altman DG, for the CONSORT group (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 357: 1191–4.

Robert Koch Institut (1998). Bundes-Gesundheitssurvey 1998. Fragebogen. Berlin: Robert Koch Institut.

Schafer JL, Graham JW (2002). Missing data: our view of the state of the art. *Psychol Methods* 7: 147–77.

Smith TW (2002). Developing nonresponse standards. In: Groves RM, Dillman DA, Eltinge JL, Little RJA, eds. *Survey nonresponse*. New York: Wiley: 27–40.

Wirtz M (2004). Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. *Rehabilitation* 43: 109–15.

Wyatt J (1997) Quantitative evaluation of clinical software, exemplified by decision support systems. *Int J Med Inform* 47: 165–73.

---

**Address for correspondence**

**Dr. Thorsten Meyer, Psychologist  
Institute for Social Medicine  
University Clinics Schleswig-Holstein  
Campus Luebeck  
Beckergrube 43–47  
D-23552 Lübeck  
Germany  
e-mail: thorsten.meyer@sozmed.uni-luebeck.de**