



ORIGINAL ARTICLE

## Reproducibility of an extensive self-report questionnaire used in secondary coronary prevention

KARI PEERSEN<sup>1,2</sup>, JOHN MUNKHAUGEN<sup>3,4</sup>, LARS GULLESTAD<sup>2,5</sup>,  
TORIL DAMMEN<sup>4</sup>, TORBJORN MOUM<sup>4</sup> & JAN ERIK OTTERSTAD<sup>1</sup>

<sup>1</sup>Department of Medicine, Hospital of Vestfold, Norway, <sup>2</sup>Faculty of Medicine, University of Oslo, Norway, <sup>3</sup>Department of Medicine, Drammen Hospital, Norway, <sup>4</sup>Department of Behavioural Sciences in Medicine, Faculty of Medicine, University of Oslo, Norway, and <sup>5</sup>Department of Cardiology, Oslo University Hospital, Norway

### Abstract

**Aims:** Self-reported information from questionnaires is frequently used in clinical epidemiological studies, but few provide information on the reproducibility of instruments applied in secondary coronary prevention studies. This study aims to assess the test–retest reproducibility of the questionnaire applied in the cross-sectional NORwegian CORonary (NOR-COR) Prevention Study. **Methods:** In the NOR-COR study 1127 coronary heart disease (CHD) patients completed a self-report questionnaire consisting of 249 questions, of which there are both validated instruments and de novo questions. Test–retest reliability of the instrument was estimated after four weeks in 99 consecutive coronary patients. Intraclass Correlation Coefficient (ICC) and Kappa ( $\kappa$ ) were calculated. **Results:** The mean interval between test and retest was 33 ( $\pm 6.4$ ) days. Reproducibility values for questions in the first part of the questionnaire did not differ from those in the latter. A good to very good reproducibility was found for lifestyle factors (smoking:  $\kappa = 1.0$ ; exercise: ICC = 0.90), medical factors (drug adherence: ICC = 0.74; sleep apnoea: ICC = 0.87), and psychosocial factors (anxiety and depression: ICC = 0.95; quality of life 12-Item Short-Form Health Survey (SF12): ICC = 0.89), as well as for the majority of de-novo-created variables covering the patient's perceptions, motivation, needs, and preferences. **Conclusions: The present questionnaire demonstrates a highly acceptable reproducibility for all key items and instruments. It thus emerges as a valuable tool for evaluating patient factors associated with coronary risk factor control in CHD patients.**

**Key Words:** Secondary prevention, coronary heart disease, questionnaires, reproducibility, coronary risk factors

### Introduction

As a result of the contemporary management of coronary heart disease (CHD), an increasing proportion of patients survive and require optimal secondary prevention [1]. A high prevalence of unhealthy lifestyle and poor risk factor control in CHD patients was demonstrated in a large European multicentre study [2]. The reasons for these findings are complex and somewhat poorly understood, and the identification of optimal patient management and healthcare factors of importance for an improved coronary risk profile remains a public health priority [3]. The aims

of an ongoing cross-sectional study, the NORwegian CORonary (NOR-COR) Prevention Study [4], are to identify medical and psychosocial factors associated with unfavourable risk factor control after a coronary event. Most of the data to be explored has been collected through a comprehensive self-report questionnaire.

Self-report questionnaires are frequently used in health research because they are easy to utilize, feasible, and cheap to apply. In order to ensure reproducibility and reliability, a test–retest study is of great

Correspondence: Kari Peersen, Department of Medicine, Hospital of Vestfold, 3103 Tønsberg, Norway.  
Email: kari.peersen@siv.no

(Accepted 7 December 2016)



importance. The reliability of such a test is assessed by measuring the responses of the same study sample to an identical questionnaire at two or more points in time. A reproducibility test will assess random measurement errors as well as the stability of the construct measured, but cannot in itself distinguish between the two [5]. Thus, one must take into consideration that any real change in the phenomenon of interest that may have occurred during the intervening period between tests will result in seemingly low levels of reliability.

There are no standards for the ideal time span between the initial test and the retest in reproducibility studies. The interval should be long enough to prevent memory effects and short enough to ensure that no real clinical change has occurred among participants [6]. Intervals of one to two weeks [7] and one month [8] have been suggested.

Self-reported information from questionnaires is frequently used in clinical epidemiological studies, but few provide information on the reproducibility of instruments applied in secondary coronary prevention studies. Those available are limited by only addressing single questionnaires with a moderate range of items. So far, few studies have explored whether reproducibility remains satisfactory in a comprehensive questionnaire applied in clinical patient studies. The purpose of this study was to evaluate the test-retest reliability of an extensive self-report questionnaire assembled and created to be used in the NOR-COR study [4]. Given acceptable reproducibility results, such a questionnaire could be valuable in future studies on risk factor control and lifestyle measures in long-term secondary coronary prevention.

## Materials and methods

A complete description of the design and methodology applied in the NOR-COR study is published elsewhere [4]. In the present study, the self-report questionnaire used in NOR-COR was completed twice by 99 stable patients with an interval of four weeks.

### Design of the NOR-COR questionnaire

The NOR-COR questionnaire contains 249 questions derived from a number of medical and psychosocial instruments that have previously, to some extent, been demonstrated to be associated with coronary risk factors, adherence to medication, and prognosis in cardiac patients [9–15]. As there were no validated instruments for revealing the patient's needs and preferences, a number of questions/items were created de novo following an extensive process

[16], described in detail previously [4]. The NOR-COR questionnaire was pilot-tested in two CHD patients in order to incorporate the patients' perspective, and subsequently tested in 20 randomly selected eligible CHD patients in order to establish relevance, acceptance, and feasibility.

The following descriptive variables have been obtained from the questionnaire:

- *Socio-demographic factors:*
  - Marital status;
  - Level of education.
- *Behaviour/lifestyle risk factors:*
  - Smoking status (never, previous, or current smoking);
  - Physical activity (frequency, duration, intensity, and a sum-score) [17];
  - Diet (the frequency of intake of fish, vegetables, and fruits);
  - Alcohol consumption (the past four weeks).
- *Medical factors:*
  - Drug adherence (the 8-item Morisky Medication Adherence Scale) [18];
  - Obstructive sleep apnoea (Berlin Questionnaire) [19].
- *Psychosocial factors*
  - Quality of life (12-Item Short-Form Health Survey (SF12)): a 12-item measure of generic quality of life with a physical health sub-scale Physical Component Summary (PCS12) and mental health sub-scale Mental Component Summary (MCS12) [20];
  - Anxiety and depression (Hospital Anxiety and Depression Scale, HADS) [21];
  - Rumination (Ruminative Response Scale, RRS): a 22-item self-report inventory designed to assess the tendency to ruminate in response to a depressed mood [15];
  - Worry (Penn State Worry Questionnaire, PSWQ): a 16-item measure of pathological worry [22];
  - Type D personality (distressed personality type, Type D Scale, DS-14): a 14-item instrument with seven items each on the subscales of negative affectivity (NA) and social inhibition (SI) [23];
  - Illness perception (Brief Illness Perception Questionnaire, BIPQ): an 8-item measure of illness identity, personal and treatment ability to control the illness, consequences, understanding and concern about the illness rated on a Likert scale from 0 to 10, and one item about what caused the patient's illness [24];
  - Perceived risk perception (PRP): a 3-item measure on a Likert scale from 0 to 10; probability for a new event within 12 months,

- your own ability to reduce coronary risk, and to what degree the disease will limit your activities [13];
- Insomnia (Bergen Insomnia Scale): a 7-item measure of on an 11-point Likert scale from 0 to 10 [25].
- *Treatment desires, perceived needs, beliefs about causes, motivation (de-novo-created questions)*
  - Beliefs regarding what caused the patient's CHD, ranking known CHD risk factors from 0 to 10 on a Likert scale indicating to what extent the patient believed that each risk factor had caused the disease to develop;
  - Motivation for further lifestyle changes and changes already achieved in these lifestyle factors;
  - Perceived needs of sufficient health information about CHD and the risk factors;
  - Participation in healthcare follow-up (cardiac rehabilitation, follow-up visits in primary healthcare);
  - Perception of the information provided by healthcare workers [16] with four assertions: I am cured, but have to change my lifestyle; I am cured and do not need to change my lifestyle; I still have heart disease and need to change my lifestyle; and I still have heart disease, but do not need to change my lifestyle;
  - Perceived needs for further secondary preventive follow-up today in order to meet the goal of prevention (email/telephone, nurse, cardiac rehabilitation, physiotherapist, nutritionist, psychiatrist/psychologist, Internet, and/or mobile app).

#### *Study population*

A total of 1127 (83% participation rate) patients aged 31–80 (mean 62) with first or recurrent diagnosis or treatment for CHD (acute myocardial infarction, coronary artery bypass graft operation, and/or elective or emergency Percutaneous Coronary Intervention (PCI)) within the time period from eight weeks to three years previously, participated in the NOR-COR study, and completed the questionnaire. The study was conducted at two Norwegian hospitals, Drammen and Vestfold. Initially, 28 of the participants recruited from Vestfold Hospital completed the NOR-COR questionnaire a second time after four weeks. It was decided to increase the number of participants to approximately 100 in order to obtain sufficient statistical power in this reproducibility study. Accordingly, 71 consecutive patients referred to cardiac rehabilitation in Vestfold Hospital performed an identical retest, with inclusion criteria

identical to those in the NOR-COR study [3]. The participants in this reproducibility study were considered as having been stable with respect to their CHD, and none had been re-hospitalized during the interval between test and retest. The same observer conducted all tests and retests, and was very alert for possible changes in the patients' physical or psychological condition that might affect retest results. In order to evaluate possible group differences, patient characteristics in the reproducibility sample and the entire NOR-COR population were compared.

#### *Statistics*

Descriptive data are presented as means  $\pm$  standard deviations (SDs), while reproducibility results are presented with 95% confidence intervals (CIs). Differences between the reproducibility sample and the NOR-COR population regarding age, sex, education, and type of event were assessed with independent two-sample *t*-tests and chi-square tests. Test–retest reliability was calculated by comparing the data obtained at test sessions 1 and 2 using Intraclass Correlation Coefficients (ICCs) for continuous data and for ordinal variables with at least five response categories [5], and Kappa ( $\kappa$ ) for nominal and ordinal variables [26] for each individual question in the NOR-COR questionnaire, as well as for summarized scores when available, such as for exercise, drug adherence, sleep apnoea, and the psychosocial questionnaires. ICC was calculated based on a two-way mixed-effect analysis of variance with 95% CIs. An acceptable reproducibility was set at the often-recommended level of ICC  $\geq 0.70$  and  $\kappa$  values were defined as acceptable if above 0.5. The guidelines for interpreting  $\kappa$  with strength of agreement based on Landis and Koch [26] suggest that values are fair between 0.21 and 0.4, moderate between 0.41 and 0.6, good between 0.61 and 0.8, and very good above 0.81. These guidelines for  $\kappa$  agreement will also be applied to continuous data using ICC. Internal consistency was calculated with standardized Cronbach's alpha for each set of items or scales. Analyses of covariance were used to examine potential differentials in reproducibility across age, gender, or education. Statistical analyses were conducted with the SPSS version 21 (SPSS Inc., US). The significance level was set at  $p < 0.05$ .

#### *Ethics*

This study was approved by the Regional Committee of Ethics in Medical Research, approval number 2013/1885. Written informed consent was obtained

from all included participants. The study is registered at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) (ID NCT02309255).

## Results

A total of 99 patients completed the retest within an interval of 33 ( $\pm 6.4$ ) days. One patient who broke his leg and a woman who lost her son within the interval between tests were excluded from the reproducibility study. The mean time interval between index hospitalization and first-time completion of the questionnaire was 34 weeks (range 8–83). The amount of missing data was 1.1% in the first test session and 3.0% in the retest, at the same level throughout the questionnaire. Participant feedback revealed that the time used to fill out the questionnaire was 30 to 45 minutes. Reproducibility figures obtained from the first part of the questionnaire did not differ from those of the last part.

There were no significant differences between patient characteristics among the NOR-COR population and the reproducibility study sample (Table I). The reproducibility values were very good for exercise and smoking (Table II), good for the use of alcohol, and moderate for diet. The reproducibility coefficients for drug adherence were acceptable, and very good for obstructive sleep apnoea.

The test–retest reliability calculations of the psychosocial factors presented in Table III show good reproducibility for quality of life (PCS12) and very good for all other psychosocial instruments. The majority of the questions covering the patient’s perceptions, needs, preferences, and motivation were above the limits for acceptable reproducibility (Table IV). The participants were asked about their preferences for follow-up to meet their present needs of optimal prevention. The reproducibility level for these replies was fair to good.

Fair internal consistency was found for sleep apnoea Berlin Category 1 sum (Cronbach’s alpha = 0.45 in test 1 and 0.35 in retest); however, the values improved to good (Cronbach’s alpha = 0.68 in test 1 and 0.66 in retest) if item 4 (“does your snoring bother others”) was deleted from computation. Moderate internal consistency was found for the 8-item Morisky Medication Adherence Scale (Cronbach’s alpha = 0.54 in both tests) and SF12 (Cronbach’s alpha = 0.65 in test 1, 0.61 in retest). All other scales showed good to very good internal consistency and Cronbach’s alpha ranged from 0.69 to 0.95, with slightly higher values in the second test.

Significant differences in the level of reproducibility across gender, age, or education level were found in a small proportion of the variables; however, there was no consistency regarding which subgroup showed the highest level of reproducibility.

Table I. Demographic and medical characteristics of the NOR-COR sample and the reproducibility sample.

	NOR-COR <i>n</i> = 1127	Reproducibility <i>n</i> = 99	<i>p</i> -value
<b>Age, mean (SD)</b>	61.6 (9.6)	63.2 (8.8)	ns
<b>Female gender, %</b>	21	17	ns
<b>Living alone, %</b>	26	24	ns
<b>Low education,<sup>a</sup> %</b>	62	55	ns
<b>Coronary diagnosis</b>			ns
Non-ST elevation MI, <i>n</i> %	561 (50)	44 (44)	ns
ST elevation MI, <i>n</i> %	335 (30)	38 (38)	ns
Stable/unstable CHD, <i>n</i> %	231 (21)	17 (17)	ns

NOR-COR: NORwegian CORonary Prevention Study; *n*: sample size; SD: standard deviation; ns: non-significant; MI: myocardial infarction; CHD: coronary heart disease; ST: ST-segment.

<sup>a</sup>Low education was defined as completion of primary or secondary school only.

## Discussion

The present study analysed the reproducibility of the questionnaire applied in the NOR-COR study. Our findings demonstrated acceptable to excellent values for almost all of the variables explored. This level of reproducibility in data from the NOR-COR questionnaire will be valuable in performing further analyses, findings, and, indeed, conclusions of the project.

There were few missing data in both tests, and the reproducibility remained high throughout the rather extensive questionnaire. The test–retest sample had similar patient characteristics to those of the total NOR-COR population. Information obtained by self-report questionnaires may be distorted by systematic errors such as the patient giving socially desirable answers, using scales and response options in idiosyncratic ways, as well as recall bias. Systematic errors and biases are hard to assess and control, and would in fact tend to boost test–retest correlations. On the other hand, poor or oscillating understanding of the underlying meaning of a question, being distracted or confused, or responding based on current mood will introduce random error or noise in measurements, thereby reducing statistical associations of substantive interest, as well as test–retest correlations. Test–retest correlations allow for estimates of random measurement errors to be established, given that the underlying construct is stable [5]. Thus, acceptable intra-individual reproducibility is reassuring in the sense that one has apparently minimized the risk of committing type II errors because of random error or noise. Conversely, reliability estimates typically based on internal consistency within a set of items tend to be boosted by systematic errors such as response scale effects and thus may yield misleadingly favourable results [27].

Table II. Test–retest reliability of lifestyle risk factors and medical factors.

	Test 1 Mean (SD)	Test 2 Mean (SD)	ICC, 95% CI	κ, 95% CI
<b>Exercise</b>				
Frequency, times per week	3.0 (1.5)	2.9 (1.4)	0.85 (0.78–0.90)	
Exercise sum score <sup>a</sup>	9.2 (1.2)	9.0 (1.1)	0.90 (0.85–0.94)	
<b>Smoking</b>				
Current smoking, <i>n</i> (%)	15 (15)	15 (15)		1.0
Previous smoking, <i>n</i> (%)	66 (67)	69 (70)		0.94 (0.87–1.02)
Never smoked, <i>n</i> (%)	28 (28)	25 (25)		0.87 (0.76–0.98)
<b>Diet</b>				
Fish >3 times/week, <i>n</i> (%)	51 (53)	53 (55)		0.49 (0.32–0.66)
Fruit/veg ≥ 2 times/day, <sup>b</sup> <i>n</i> (%)	41 (43)	39 (40)		0.44 (0.26–0.62)
<b>Alcohol last 4 weeks, <i>n</i> (%)</b>	81 (84)	84 (86)		0.75 (0.56–0.94)
<b>Drug adherence</b>				
Morisky scale sum score	7.4 (0.9)	7.3 (1.0)	0.74 (0.61–0.83)	
<b>Obstructive sleep apnoea</b>				
Berlin category 1 sum <sup>c</sup>	1.59 (1.2)	1.55 (1.2)	0.87 (0.80–0.91)	
Berlin category 2 sum	0.45 (0.8)	0.41 (0.8)	0.89 (0.83–0.93)	

ICC: intraclass correlation coefficient; κ: Kappa agreement; CI: confidence interval; SD: standard deviation.

<sup>a</sup>Exercise sum score, sum of frequency, duration, and intensity.

<sup>b</sup>Fruit and/or vegetables at least twice a day.

<sup>c</sup>Berlin category 1 sum, snoring, and sleep apnoea; Berlin category 2 sum, tired or exhausted.

Table III. Test–retest reliability of psychosocial factors.

	Test session 1 Mean (SD)	Test session 2 Mean (SD)	ICC (95% CI)
<b>Quality of life, SF12</b>			
Physical health sub-scale sum score	41.89 (5.5)	41.44 (5.4)	0.77 (0.65–0.85)
Mental health sub-scale sum score	50.83 (8.5)	50.83 (9.6)	0.89 (0.84–0.93)
<b>Hospital Anxiety and Depression Scale (HADS)</b>			
HADS-A sum; anxiety	3.63 (3.7)	3.30 (3.9)	0.92 (0.88–0.95)
HADS-D sum; depression	2.96 (2.9)	3.01 (3.5)	0.94 (0.91–0.96)
HADS-T sum; total score	6.59 (6.1)	6.32 (6.8)	0.95 (0.93–0.97)
<b>Ruminative Response Scale (RRS)</b>			
RRS sum score	29.47 (9.3)	29.29 (11.2)	0.88 (0.81–0.92)
<b>Penn State Worry Questionnaire (PSWQ)</b>			
PSWQ sum score	35.0 (12.2)	34.6 (12.8)	0.91 (0.86–0.94)
<b>Type D personality (DS-14)</b>			
Social inhibition sum score	7.55 (5.5)	7.32 (5.5)	0.90 (0.85–0.94)
Negative affectivity sum score	5.76 (5.5)	5.40 (5.7)	0.91 (0.86–0.94)
<b>Brief Illness Perception Questionnaire (BIPQ)</b>			
BIPQ sum score	25.9 (12.3)	25.8 (13.1)	0.91 (0.86–0.94)
<b>Perceived risk perception (PRP), Likert scale 0–10</b>			
PRP 1	2.1 (2.1)	2.7 (2.4)	0.59 (0.39–0.73)
PRP 2	6.9 (2.5)	6.9 (2.6)	0.67 (0.50–0.78)
PRP 3	2.3 (2.3)	2.6 (2.5)	0.74 (0.60–0.82)
<b>Bergen insomnia scale</b>			
Insomnia sum score	11.2 (10.3)	10.7 (10.5)	0.92 (0.88–0.95)

SD: standard deviation; CI: confidence interval; ICC: intraclass correlation coefficient; CI: confidence interval.

The mean time interval between index hospitalization and first-time completion of the questionnaire was eight months. After this relatively long period the majority had completed the rehabilitation process in our hospital. Possible early problems with medication habituation, anxiety, and depression were considered sufficiently diminished, and the patients’ physical activity level was restored. It is, however, not possible

to guarantee total stability (i.e. lack of “true” change) over four weeks, but the abovementioned should have reduced the risk of clinically important improvements or deteriorations that might have influenced reproducibility in the data presented. Test–retest correlations tend to be higher when the time interval between the two points of measurement is short, because few changes have occurred, but there is also a risk of

Table IV. Test-retest reliability of beliefs about disease causes, motivation, perceived needs, and treatment desires.

	Test session 1 Mean (SD) Likert scale 0–10	Test session 2 Mean (SD) Likert scale 0–10	ICC (95% CI)
<b>Beliefs regarding what caused CHD</b>			
Smoking	3.84 (4.0)	3.80 (4.0)	0.94 (0.91–0.96)
Lack of exercise	4.47 (3.0)	4.91 (3.0)	0.88 (0.83–0.92)
<b>Motivation for lifestyle changes and already applied changes</b>			
Motivation to quit smoking	7.83 (3.6)	8.48 (2.5)	0.87 (0.69–0.95)
Motivation to improve diet	5.26 (2.9)	5.39 (2.9)	0.85 (0.78–0.90)
Motivation to increase exercise	5.06 (3.0)	5.52 (2.8)	0.75 (0.62–0.83)
Have changed diet	5.62 (2.6)	5.43 (2.8)	0.84 (0.76–0.89)
Have increased exercise	4.78 (3.1)	4.93 (2.8)	0.89 (0.84–0.93)
<b>Perceived needs of additional information</b>			
Sufficient information about disease	8.19 (2.2)	8.23 (2.0)	0.73 (0.60–0.82)
Sufficient information about risk factors	8.58 (1.8)	8.20 (1.9)	0.80 (0.70–0.86)
	Test session 1 n (%)	Test session 2 n (%)	$\kappa$ (95% CI)
<b>Healthcare follow-up</b>			
Participated in cardiac rehabilitation	92 (93)	92 (93)	0.69 (0.41–0.98)
Current follow-up general practitioner $\geq$ 3 times/year	48 (49)	44 (44)	0.72 (0.58–0.86)
<b>Perception of the information provided</b>			
Cured from CHD	68 (69)	67 (68)	0.55 (0.38–0.74)
No need to change lifestyle	28 (28)	27 (27)	0.63 (0.46–0.81)
<b>Treatment desires</b>			
Email, SMS, telephone	50 (52)	45 (47)	0.56 (0.39–0.73)
Cardiac nurse	66 (68)	61 (63)	0.34 (0.15–0.53)
Multidisciplinary cardiac rehabilitation	32 (33)	26 (28)	0.54 (0.36–0.72)
Physiotherapist	31 (32)	31 (32)	0.71 (0.56–0.86)
Dietician	32 (33)	31 (32)	0.69 (0.54–0.84)
Psychiatrist/psychologist	14 (14)	7 (7)	0.57 (0.44–0.70)
Internet	27 (28)	23 (25)	0.47 (0.28–0.67)
Mobile app	17 (18)	15 (16)	0.70 (0.50–0.90)
No need for follow-up	32 (36)	31 (33)	0.63 (0.46–0.81)

ICC, intraclass correlation coefficient;  $\kappa$ , Kappa agreement; CI, confidence interval; CHD, coronary heart disease.

memory effects; that is, respondents recalling their response on the first occasion and choosing the same option on the second occasion in order to appear “consistent”. In order to avoid influence of memory effects and to reduce the possibility of significant events and real changes between the two tests, four to eight weeks has been suggested as the ideal time between the two measurements [28,29].

In post-Myocardial infarction (MI) patients, the assessment of type D personality has been shown to be very stable over 18 months [30] and comparable to the good reproducibility of frequency of exercise per week in stable coronary patients when measured with one week interval between the two tests, whereas reproducibility for exercise diminishes with a longer interval between tests [31–33]. In the present context, the majority of our study participants clearly belonged to the category of stable CHD.

We had expected a tendency towards lower reproducibility in questions from the last part of the questionnaire due to tiredness or fatigue. This did

not turn out to be the case, as was also observed in a diet study where the length of the questionnaire had only a minor impact on the response rate and data quality [34].

In the INTERHEART study [35] structured questionnaires were administered to obtain information about socio-demographic factors and cardiovascular risk factors. Repeat measures of risk factors were made in 279 controls at a median interval of 409 days. Except from a nearly identical, and very good agreement rate for smoking in INTERHEART and the present study ( $\kappa = 0.94$  vs. 1.0, respectively), the respective reproducibility values in INTERHEART and our study differed for depression ( $\kappa = 0.44$  vs. ICC = 0.94), regular physical activity ( $\kappa = 0.56$  vs. ICC = 0.85 for frequency), and alcohol ( $\kappa = 0.52$  vs. ICC = 0.75). Different questionnaires and time interval that had elapsed between test and retest may explain these divergences.

The reproducibility of drug adherence, sleep apnoea, and psychosocial factors based upon widely

used questionnaires in the present study was high and in line with most other studies [18–25,36].

Our findings of quite acceptable reproducibility data with only few exceptions can be explained by the extensive process used to develop de novo questions and inclusion of questionnaires that have previously been validated and found to have acceptable reproducibility [17–25,36].

These robust reproducibility data will have practical implications for future analysis of the association between potentially modifiable patient factors and unfavourable risk factor control. Since most of the data to be used in this context is derived from the questionnaire, the present findings are reassuring for further NOR-COR projects and for its application in future clinical studies of secondary CHD prevention.

#### Study limitations

The participants of the reproducibility study exclusively represented Vestfold Hospital where nearly 80% attend cardiac rehabilitation. Since only half of the NOR-COR study participants attended such a programme, a selection bias cannot be excluded. However, no socio-demographic differences were observed between the entire NOR-COR population and the reproducibility study sample.

True change in the underlying phenomenon will of course result in low or at least reduced test–retest correlations, thus giving the impression of relatively poor measurement reliability (if no true change is assumed). Weak test–retest correlations, therefore, must be viewed with caution since we may in fact be underestimating reliability. However, in the present study this seems to be a rather unnecessary concern since we have consistently found very high test–retest correlations, also for measures for which one might suspect some true change to have occurred in the time period between test and retest (e.g. the reproducibility for physical activity frequency was found to be ICC 0.85). The risk of overestimating measurement reliability because of artificially boosted test–retest correlations (caused by memory effects, stable biases and/or response styles, mode of administration, etc.) can only be assessed by applying alternative research designs, such as having access to a gold standard, systematically altering instrument style and formatting, switching modes of administration etc., which is clearly outside the scope of the present paper.

#### Conclusion

Reliability studies based on test–retests are essential elements when it comes to establishing the quality of self-report data. A good to very good reproducibility

was found for almost all of the items and scales used in the comprehensive NOR-COR questionnaire. Thus, this instrument emerges as a valuable tool for evaluating risk factor control in CHD patients in general, laying the foundation not only for further analyses, findings, and conclusions in the NOR-COR project, but also for similar comprehensive questionnaires applied in future clinical patient studies.

#### Acknowledgements

The NOR-COR project was carried out at Drammen and Vestfold Hospitals and was developed in collaboration with communities at the medical faculty of University of Oslo.

#### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a research grant from Extrastiftelsen, Norway (grant number 76728).

#### Supplemental material

The online supplementary material is available at <http://sjp.sagepub.com/supplemental>

#### References

1. Townsend N, Nichols M, Scarborough P, et al. Cardiovascular disease in Europe – epidemiological update 2015. *Eur Heart J* 2015;36(40):2696–705.
2. Kotseva K, Wood D, de Bacquer D, et al. EUROASPIRE IV: a European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. *Eur J Prev Cardiol* 2016;23(6):636–48.
3. Perk J, de Backer G, Gohlke H, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012). *Eur Heart J* 2012;33(13):1635–701.
4. Munkhaugen J, Sverre E, Peersen K, et al. The role of medical and psychosocial factors for unfavourable coronary risk factor control. *Scand Cardiovasc J* 2016;50(1):1–8.
5. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? *Qual Life Res* 2004;13(3):571–86.
6. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34–42.
7. Deyo RA, Diehr P and Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991;12(4):142–58.
8. Spertus J, Dorian P, Bubien R, et al. Development and validation of the Atrial Fibrillation Effect on Quality-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation. *Circ Arrhythm Electrophysiol* 2011;4(1):15–25.

9. Gehi A, Haas D, Pipkin S, et al. Depression and medication adherence in outpatients with coronary heart disease: findings from the Heart and Soul Study. *Arch Intern Med* 2005;165(21):2508–13.
10. Tully PJ, Cosh SM and Baune BT. A review of the affects of worry and generalized anxiety disorder upon cardiovascular health and coronary heart disease. *Psychol Health Med* 2013;18(6):627–44.
11. Svansdottir E, Denollet J, Thorsson B, et al. Association of type D personality with unhealthy lifestyle, and estimated risk of coronary events in the general Icelandic population. *Eur J Prev Cardiol* 2013;20(2):322–30.
12. Grandner MA. Addressing sleep disturbances: an opportunity to prevent cardiometabolic disease? *Int Rev Psychiatry* 2014;26(2):155–76.
13. Broadbent E, Petrie KJ, Ellis CJ, et al. Patients with acute myocardial infarction have an inaccurate understanding of their risk of a future cardiac event. *Intern Med J* 2006;36(10):643–7.
14. Lichtman JH, Froelicher ES, Blumenthal JA, et al. Depression as a risk factor for poor prognosis among patients with acute coronary syndrome: systematic review and recommendations: a scientific statement from the American Heart Association. *Circulation* 2014;129(12):1350–69.
15. Denton EG, Rieckmann N, Davidson KW, et al. Psychosocial vulnerabilities to depression after acute coronary syndrome: the pivotal role of rumination in predicting and maintaining depression. *Front Psychol* 2012;3:288.
16. Perk J, Hambræus K, Burell G, et al. Study of Patient Information after percutaneous Coronary Intervention (SPICI): should prevention programmes become more effective? *EuroIntervention* 2015;10:e1–7.
17. Kurtze N, Rangul V, Hustvedt BE, et al. Reliability and validity of self-reported physical activity in the Nord-Trøndelag Health Study: HUNT 1. *Scand J Public Health* 2008;36(1):52–61.
18. Korb-Savoldelli V, Gillaizeau F, Pouchot J, et al. Validation of a French version of the 8-item Morisky medication adherence scale in hypertensive adults. *J Clin Hypertens (Greenwich)* 2012;14(7):429–34.
19. Popevic MB, Milovanovic A, Nagorni-Obradovic L, et al. Screening commercial drivers for obstructive sleep apnea: translation and validation of Serbian version of Berlin Questionnaire. *Qual Life Res* 2016;25(2):343–9.
20. Ware J Jr, Kosinski M and Keller DS. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34(3):220–33.
21. Wang W, Chair SY, Thompson DR, et al. A psychometric evaluation of the Chinese version of the Hospital Anxiety and Depression Scale in patients with coronary heart disease. *J Clin Nurs* 2009;18(17):2436–43.
22. Pallesen S, Nordhus IH, Carlstedt B, et al. A Norwegian adaptation of the Penn State Worry Questionnaire: factor structure, reliability, validity and norms. *Scand J Psychol* 2006;47(4):281–91.
23. Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. *Psychosom Med* 2005;67(1):89–97.
24. Broadbent E, Petrie KJ, Main J, et al. The brief illness perception questionnaire. *J Psychosom Res* 2006;60(6):631–7.
25. Pallesen S, Bjorvatn B, Nordhus IH, et al. A new scale for measuring insomnia: the Bergen Insomnia Scale. *Percept Mot Skills* 2008;107(3):691–706.
26. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–74.
27. DeVellis R. *Scale development: theory and applications*. 3rd ed. Thousand Oaks, CA: SAGE, 2012, pp.31–58.
28. Arribas F, Ormaetxe JM, Peinado R, et al. Validation of the AF-QoL, a disease specific quality of life questionnaire for patients with atrial fibrillation. *Europace* 2010;12(3):364–70.
29. Svilaas A, Strøm EC, Svilaas T, et al. Reproducibility and validity of a short food questionnaire for the assessment of dietary habits. *Nutr Metab Cardiovasc Dis* 2002;12(2):60–70.
30. Martens EJ, Smith OR, Winter J, et al. Type-D personality is a stable taxonomy in post-MI patients over an 18-month period. *J Psychosom Res* 2007;63(5):545–50.
31. Gremaux V, Iskandar M, Kervio G, et al. The Dijon Physical Activity Score: reproducibility and correlations with physical fitness in patients with coronary heart disease. *Ann Readapt Med Phys* 2008;51(5):366–78.
32. Kurtze N, Rangul V, Hustvedt BE, et al. Reliability and validity of self-reported physical activity in the Nord-Trøndelag Health Study: HUNT1. *Scand J Public Health* 2008;36(1):52–61.
33. Shepard RJ. Limits to the measurements of habitual physical activity by questionnaires. *Br J Sports Med* 2003;37(3):197–206.
34. Subar AF, Ziegler RG, Thompson FE, et al. Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. *Am J Epidemiol* 2001;153(4):404–9.
35. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364(9438):937–52.
36. Roelofs J, Muris P, Huibers M, et al. On the measurement of rumination: a psychometric evaluation of the ruminative response scale and the rumination on sadness scale in undergraduates. *J Behav Ther Exp Psychiatry* 2006;37(4):299–313.