Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease

J Müller-Nordhorn, S Roll, S N Willich

Objective: To investigate whether a shorter health status instrument, the short form (SF)-12, is comparable with its longer version, the SF-36, for measuring health related quality of life of patients with coronary heart disease.

Design: Prospective cohort study with follow up at six and 12 months.

Setting: 18 cardiac rehabilitation centres in Germany.

Patients: Patients were enrolled at admission to the rehabilitation centres after myocardial infarction, coronary artery bypass grafting, and percutaneous transluminal coronary angioplasty.

Analyses: Correlation coefficients were calculated between SF-12 and SF-36 physical component summary (PCS-12/-36) and mental component summary (MCS-12/-36) scores and the respective change scores. Responsiveness to change was determined with the standardised response mean.

Main results: 2441 patients were enrolled (78% men, mean (SD) age 60 (10) years; 22% women, 65 (10) years). Baseline PCS-12 and PCS-36 scores were highly correlated (r = 0.96, p < 0.001), as were baseline MCS-12 and MCS-36 scores (r = 0.96, p < 0.001). Similarly, change scores between baseline and 12 months were highly correlated (PCS-12/-36: r = 0.94, p < 0.001; MCS-12/-36: r = 0.95, p < 0.001). There was no difference in standardised response means between the SF-12 and SF-36 scales.

Conclusions: The SF-12 summary measures replicate well the SF-36 summary measures and show similar responsiveness to change. The SF-12 appears to be an efficient alternative to the SF-36 for the assessment of health related quality of life of patients with coronary heart disease.
SF-36 and SF-12 questionnaires
The SF-36 is a generic health status instrument with 36 items, eight subscales that aggregate 2–10 items each, and two summary measures that aggregate the subscales.1,2 The two summary measures of the SF-36 will be referred to as physical component summary 36 (PCS-36) and mental component summary 36 (MCS-36) scales. Compared with the SF-36, the SF-12 has only one or two items from each of the eight health concepts of the SF-36.13 The SF-12 items allow the calculation of the PCS (PCS-12) and MCS (MCS-12) scales but not of the subscales. The items selected for the SF-12 and the scoring algorithms for the summary measures were cross validated in nine countries.19 The German version of the SF-12 was translated and validated according to the methods developed by the International Quality of Life Assessment Group.20 We calculated the SF-12 scores from items embedded in the SF-36, which has been shown to be equivalent to calculating the SF-12 scores from the stand alone SF-12 items.13–21 As SF-12 items are relatively heterogeneous, internal consistency estimates of reliability underestimate the reliability of SF-12 measures and are not applicable to single item measures.22 Higher SF-12 and SF-36 scores indicate better HRQoL, a positive change in SF-12 and SF-36 scores indicates improvement in HRQoL, and a negative change indicates deterioration.

For the calculation of the SF-36 scales, missing data were imputed according to the recommendations of the SF-12 user’s manual.1 In the recommended algorithm, a person specific estimate for any missing item is substituted when the respondent answered at least 50% of the items of a subscale. The summary scales are then calculated from the subscales and are set as missing if the respondent is missing any one of the eight SF-36 subscales. The SF-12 summary scales, on the other hand, are calculated directly from the 12 items. It is recommended that the SF-12 summary scales be set as missing if the respondent is missing any one of the SF-12 items in the survey.23

Responsiveness to change
Responsiveness to change was compared between the SF-12 and the SF-36. Indices of responsiveness are the responsiveness statistic, effect sizes, relative efficiency, and the standardised response mean (SRM).24 There is no consensus on which is the most appropriate index. We chose the SRM, as it is one of the most commonly used indices of responsiveness and takes the variation of change into account.25–28 The SRM is calculated by dividing the mean change in scores by the standard deviation of the change. It can be easily interpreted by Cohen’s interpretation of effect sizes with SRMs of at least 0.2 being regarded as small, SRMs of at least 0.5 as moderate, and SRMs of 0.8 or greater as large.27

Statistical analysis
Descriptive statistics were calculated for the PCS-12 and MCS-12 scores and for the PCS-36 and MCS-36 scores at baseline and during follow up. The correlation between baseline PCS-12/-36 and MICS-12/-36 scores and the correlation between PCS-12/-36 and MCS-12/-36 change (12 months – baseline) scores were estimated by the Pearson correlation coefficient. Responsiveness to change was analysed as described above. Linear regression analysis was used to determine how much of the variation in PCS-36 and MCS-36 scores is explained by the respective PCS-12 and MCS-12 scores. The models were adjusted for age and sex. To test whether there was any significant effect modification by age or sex, the interaction terms PCS-12/MCS-12-age and PCS-12/MCS-12-sex were used. All tests for significance were two sided; the significance level was α = 0.05. Statistical analyses were done with SPSS version 10.0 for windows (SPSS Inc, Chicago, Illinois, USA).

RESULTS
Study population
A total of 2441 patients were enrolled between January and July 1997. Of all study patients, 78% (n = 1907) were men, with a mean (SD) age of 60 (10) years, and 22% women (n = 534), with a mean (SD) age of 65 (10) years. Primary indications for admission were myocardial infarction (56%, n = 1379), CABG (38%, n = 916), and PTCA (6%, n = 141). The response rates of patients to the questionnaires were 92% (n = 2233) after six months and 85% (n = 2069) after 12 months.

Baseline SF-12 and SF-36 summary scores
Table 1 shows descriptive statistics for the PCS and MCS scales of the SF-12 and SF-36. Normative data of a disease specific US population are given for comparison. Means and standard deviations of the SF-12 and SF-36 summary scores at baseline were similar. Strong correlations existed between baseline PCS-12 and PCS-36 (r = 0.96; p < 0.001) and between baseline MCS-12 and MCS-36 (r = 0.96; p < 0.001) (fig 1A).

Variation in SF-36 scores
About 92% of the variation in PCS-36 scores at baseline was explained by the model (table 2). There was no significant interaction between the PCS-12 scores and either age or sex. About 93% of the variation in MCS-36 scores at baseline was explained by the model (table 2). There was no significant interaction between the MCS-12 scores and either age or sex.

Change and responsiveness to change
PCS-12 and PCS-36 change scores (12 months – baseline) were strongly correlated (r = 0.94; p < 0.001) (fig 1B). Similarly, MCS-12 and MCS-36 change scores were strongly correlated (r = 0.95; p < 0.001). Table 3 shows mean SF-12 and SF-36 summary scores at baseline, six months, and 12 months for patients with complete follow up data, as well as the respective change scores (12 months – baseline).

Responsiveness to change as measured by the SRM was greatest in the PCS-12 and PCS-36 scales of patients after CABG and PTCA followed by the respective MCS-12 and MCS-36 scales (table 3). SRMs were smallest for change after myocardial infarction. Overall, SRMs were similar for PCS-12 and PCS-36 scales, as well as for MCS-12 and MCS-36 scales.

Missing values
The percentage of missing values for the calculation of PCS and MCS scores was 25% (n = 599) for the SF-36 summary scales and 27% (n = 667) for the SF-12 summary scales (table 1). There were no differences between patients with missing SF-12 and those with missing SF-36 values regarding socioeconomic variables such as age (62 (10) vs 63 (10) years), sex (both 72% men) and education (both 10% > 10 years), nor in the result of the exercise ECG (87 (32) vs 86 (32) W). Table 4 shows the percentage of missing values for the 12 items used to score the SF-12 summary scales. Missing values were highest for the items belonging to the role-physical and role-emotional health concepts.

DISCUSSION
The SF-12 summary scores were highly correlated with the SF-36 summary scores for patients with coronary heart disease in our study. High correlations between the SF-12 and the SF-36 have been described for both general populations and for patients with certain diseases such as rheumatoid arthritis and stroke.13,19,24,27 For patients with
coronary heart disease, high correlations between the SF-12 and SF-36 summary measures ($r > 0.9$) were described in a small subsample of the medical outcomes study of patients after recent myocardial infarction.\(^{13}\) Our analyses confirm the findings of the medical outcomes study in a larger sample of patients with coronary heart disease.

In our study, responsiveness to change of the SF-12 and SF-36 summary measures was similar in patients with coronary heart disease. Other studies of patients with different conditions such as congestive heart failure, sleep apnoea, inguinal hernia, and low back pain also reported a similar responsiveness to change of the SF-12 and SF-36 summary scales.\(^{26-30}\) A slightly lower responsiveness to change of the SF-12 compared with the SF-36 summary measures was reported for patients with rheumatoid arthritis and epilepsy.\(^{24-25}\) For patients with coronary heart disease, responsiveness to change of the SF-12 and SF-36 summary measures has not been compared. One small longitudinal study investigated responsiveness to change of the SF-12 summary measures for patients after myocardial infarction.\(^{23}\) The authors reported large to moderate SRMs for the physical SF-12 summary measure during the first six weeks and small SRMs for the mental SF-12 summary measure during the total 24 week follow up period.

The SF-12 summary measures replicate the SF-36 summary measures well. However, a criticism of the SF-12 is that it allows only the calculation of the summary scales but not of the subscales. The summary scales may conceal important information contained in the subscales of the SF-36.\(^{23-30}\) Here, investigators have to trade off between the additional information of the subscales and longer questionnaires or interviews. In clinical trials, for example, the effect of a treatment and the difference between the intervention and control groups may be assessed sufficiently by the physical and mental summary measures. It is, therefore, at the investigators’ discretion to decide whether the use of the summary measures is appropriate for their respective study design.

### Table 1  Baseline scores of the SF-12 and SF-36 summary measures

<table>
<thead>
<tr>
<th></th>
<th>PCS-12</th>
<th>PCS-36</th>
<th>MCS-12</th>
<th>MCS-36</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean (SD)</strong></td>
<td>41 (12)</td>
<td>41 (12)</td>
<td>47 (11)</td>
<td>46 (12)</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>12–63</td>
<td>10–66</td>
<td>16–71</td>
<td>10–74</td>
</tr>
<tr>
<td><strong>Missing</strong></td>
<td>27%</td>
<td>25%</td>
<td>27%</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Mean disease specific US norm population</strong>(^*)</td>
<td>42</td>
<td>43</td>
<td>52</td>
<td>52</td>
</tr>
</tbody>
</table>

\(^*\)US norm population for the condition myocardial infarction, recent.\(^2\)^\(^{13}\)  
MCS, mental component summary; PCS, physical component summary.
A cause for concern are the high percentages of missing values in both the SF-12 and the SF-36 summary measures. The Australian study validating the SF-12 in a heart and stroke population reported a similar percentage of missing values in both the SF-12 and the SF-36 summary measures. Despite different scoring algorithms and the imputation of missing data for the SF-36 subscales, missing rates were similar for the SF-12 and SF-36 summary scales in our study. An explanation is that the imputation of missing data is not possible for certain subscales if more than 50% of the items are missing. Missing rates in our study were highest for items of the health concepts role-physical and role-emotional, which has been reported elsewhere as well.31–33 It seems that some items prevent both direct calculation of the SF-12 summary measures—if reported—of other studies have been between 10–20%.15 16 The high percentage of missing values is, however, not a problem specific to the SF-12 summary measures but also to the SF-36 summary measures. However, missing rates in our study should be interpreted with care, since we used SF-12 items embedded in the much longer SF-36 questionnaire. Missing rates in our study were similar for the SF-12 and SF-36 summary scales in our study. An explanation is that the imputation of missing data is not possible for certain subscales if more than 50% of the items are missing. Missing rates in our study were highest for items of the health concepts role-physical and role-emotional, which has been reported elsewhere as well.31–33

### Table 2 Variation in SF-36 summary scales scores at baseline

<table>
<thead>
<tr>
<th>Item</th>
<th>Baseline</th>
<th>6 months</th>
<th>12 months</th>
<th>Change</th>
<th>SRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS-12</td>
<td>46 (10)</td>
<td>43 (10)</td>
<td>44 (10)</td>
<td>-1.9 (11)</td>
<td>0.18</td>
</tr>
<tr>
<td>PCS-36</td>
<td>46 (10)</td>
<td>44 (11)</td>
<td>45 (10)</td>
<td>-1.7 (11)</td>
<td>0.05</td>
</tr>
<tr>
<td>MCS-12</td>
<td>50 (11)</td>
<td>49 (11)</td>
<td>49 (11)</td>
<td>-0.5 (11)</td>
<td>0.05</td>
</tr>
<tr>
<td>MCS-36</td>
<td>49 (11)</td>
<td>48 (11)</td>
<td>48 (11)</td>
<td>-0.5 (11)</td>
<td>0.04</td>
</tr>
<tr>
<td>CAPEG</td>
<td>35 (11)</td>
<td>42 (10)</td>
<td>43 (10)</td>
<td>7.7 (12)</td>
<td>0.63</td>
</tr>
<tr>
<td>PCS-36</td>
<td>46 (12)</td>
<td>51 (10)</td>
<td>51 (10)</td>
<td>4.8 (13)</td>
<td>0.37</td>
</tr>
<tr>
<td>MCS-12</td>
<td>45 (12)</td>
<td>50 (11)</td>
<td>50 (11)</td>
<td>4.3 (13)</td>
<td>0.33</td>
</tr>
<tr>
<td>MCS-36</td>
<td>36 (11)</td>
<td>40 (10)</td>
<td>41 (11)</td>
<td>4.1 (13)</td>
<td>0.48</td>
</tr>
<tr>
<td>PCS-36</td>
<td>44 (11)</td>
<td>47 (11)</td>
<td>48 (10)</td>
<td>3.5 (12)</td>
<td>0.29</td>
</tr>
<tr>
<td>MCS-36</td>
<td>43 (12)</td>
<td>46 (11)</td>
<td>46 (11)</td>
<td>2.9 (12)</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Data are mean (SD).

*Numbers refer to patients with complete data at baseline and follow up; †change between 12 months and baseline. Values may not add up due to rounding.

CAPEG, coronary artery bypass grafting; PTCA, percutaneous transluminal coronary angioplasty; SRM, standardised response mean.

### Table 4 Missing values for items of the SF-12

<table>
<thead>
<tr>
<th>Item</th>
<th>Abbreviated description of items</th>
<th>Number missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>In general, would you say your health is...</td>
<td>20 (0.8%)</td>
<td></td>
</tr>
<tr>
<td>Moderate activities</td>
<td>181 (7.4%)</td>
<td></td>
</tr>
<tr>
<td>Climbing several flights of stairs</td>
<td>126 (5.2%)</td>
<td></td>
</tr>
<tr>
<td>Accomplished less than you would like</td>
<td>310 (12.7%)</td>
<td></td>
</tr>
<tr>
<td>Limited in the kind of activities (physical)</td>
<td>376 (15.4%)</td>
<td></td>
</tr>
<tr>
<td>Pain interferes with normal work</td>
<td>56 (2.3%)</td>
<td></td>
</tr>
<tr>
<td>Have a lot of energy</td>
<td>253 (10.4%)</td>
<td></td>
</tr>
<tr>
<td>Health interferes with social activities</td>
<td>56 (2.3%)</td>
<td></td>
</tr>
<tr>
<td>Accomplished less than you would like (emotional)</td>
<td>285 (11.7%)</td>
<td></td>
</tr>
<tr>
<td>Didn’t do activities as carefully as usual (emotional)</td>
<td>371 (15.2%)</td>
<td></td>
</tr>
<tr>
<td>Felt calm and peaceful</td>
<td>207 (8.5%)</td>
<td></td>
</tr>
<tr>
<td>Felt downthearted and blue</td>
<td>233 (9.5%)</td>
<td></td>
</tr>
</tbody>
</table>

*Abbreviated description of items.13

SF-12 summary measures—if reported—of other studies have been between 10–20%.15 16 The high percentage of missing values is, however, not a problem specific to the SF-12 summary measures but also to the SF-36 summary measures. Despite different scoring algorithms and the imputation of missing data for the SF-36 subscales, missing rates were similar for the SF-12 and SF-36 summary scales in our study. An explanation is that the imputation of missing data is not possible for certain subscales if more than 50% of the items are missing. Missing rates in our study were highest for items of the health concepts role-physical and role-emotional, which has been reported elsewhere as well.31–33 It seems that some items prevent both direct calculation of the SF-12 summary scales and calculation of the SF-36 subscales and subsequently of the SF-36 summary scales. However, missing rates in our study should be interpreted with care, since we used SF-12 items embedded in the much longer SF-36 questionnaire. Missing rates might have been different if the SF-12 items had been administered alone (unembedded). Also, a “context effect” of the embedded form with the remaining 24 items of the SF-36 causing responses to the SF-12 items to be different cannot be excluded. Ware and colleagues13 compared the mean scores of SF-12 items embedded in the SF-36 in a sample of 525 employees with the mean scores of the SF-12 items unembedded in the same sample a year later. They found a very high (r = 0.999) product–moment correlation between...
embedded and unembedded SF-12 item means. We therefore assume that the results of the present study are not compromised by the use of the embedded SF-12 items.

Conclusion
The SF-12 appears adequate in replacing the SF-36 in large studies assessing HRQoL of patients with coronary heart disease. The use of the SF-12 may reduce respondent burden and save resources.

ACKNOWLEDGEMENTS
The study was supported by a grant from MSD Sharp and Dohme GmbH, Germany. It was conducted in cooperation with the German Society for Prevention and Rehabilitation and the German Society for Cardiology.

References