



Review Article (Meta-analysis)

# Measuring Change in Health Status Over Time (Responsiveness): A Meta-analysis of the SF-36 in Cardiac and Pulmonary Rehabilitation



Flora-Joan van Rotterdam, PhD, MSc <sup>a</sup>,  
Michael Hensley, MB, PhD <sup>b</sup>,  
Michael Hazelton, RN, MA, PhD <sup>c</sup>

<sup>a</sup> School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia

<sup>b</sup> Hunter and New England Hospital, Newcastle, New South Wales, Australia

<sup>c</sup> School of Nursing and Midwifery, University of Newcastle, Newcastle, New South Wales, Australia

## KEYWORDS

Cardiac rehabilitation;  
Outcome assessment,  
health care;  
Quality of life;  
Rehabilitation

**Abstract** *Objective:* The purpose of this meta-analysis is to gather and investigate pooled information on the responsiveness of the main patient outcome measure in cardiac rehabilitation (CR) and pulmonary rehabilitation (PR). The main outcome measure in CR and PR has been found to be the Medical Outcomes Study Short Form-36 health survey (SF-36).

*Data Sources:* A previous systematic effectiveness review of this literature was used as the basis of this statistical analysis, with the bulk of articles being observational studies.

*Study Selection:* This meta-analysis assessed articles on CR that used SF-36 pre and post “within” (per interventional group) mean scores and in the PR literature that used the SF-36 and the Chronic Respiratory Questionnaire (CRQ) “within” change scores.

*Data Extraction:* Each group of patients in the chosen literature were taken to represent a single group, so that studies such as randomized controlled trials were listed twice. We undertook a correlation analysis between SF-36 pre and post “within” mean scores in the CR literature. In the PR literature, we undertook a correlation analysis between SF-36 and the CRQ “within” change scores; this involved Spearman correlation coefficients.

*Data Synthesis:* The SF-36 Mental Composite Score domain is the most responsive of the composite SF-36 domains, with the Physical Composite Score showing less ability to discriminate in the higher SF-36 scores. In the individual domains, Role Emotional scored  $r=0.52$ ,  $P<.001$  with only

*List of abbreviations:* CR, cardiac rehabilitation; CRQ, Chronic Respiratory Questionnaire; MCS, Mental Composite Score; PCS, Physical Composite Score; POM, patient outcome measure; PR, pulmonary rehabilitation; QOL, quality of life; RCT, randomized controlled trial; SF-36, Medical Outcomes Study Short Form-36 health survey.

Disclosures: none.

Cite this article as: Arch Rehabil Res Clin Transl. 2021;3:100127

<https://doi.org/10.1016/j.arrct.2021.100127>

2590-1095/© 2021 The Authors. Published by Elsevier Inc. on behalf of American Congress of Rehabilitation Medicine. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

27% of the variance explained, and Role Physical with  $r=0.49$ ,  $P\leq.005$  had only 24% of the variance explained. In the PR literature Spearman rank correlation coefficient shows that SF-36 Physical Composite Score has a weaker correlation to the CRQ at 0.39 than the SF-36 Mental Composite Score, which was 0.63.

**Conclusions:** This suggests that the SF-36 is not suited as a pre- to postprogram assessment tool for CR and PR. More studies, however, need to be conducted particularly in CR with regard to the responsiveness of the SF-36.

© 2021 The Authors. Published by Elsevier Inc. on behalf of American Congress of Rehabilitation Medicine. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The prevalence of chronic conditions may be high, but their incidence is relatively low. Chronic illness presents many different challenges, and cardiac rehabilitation (CR) and pulmonary rehabilitation (PR) have become a component of standard medical care when dealing with these challenges.<sup>1,2</sup> These therapies form an individualized program that helps control and alleviate symptoms, optimizes functional capacity, improves health-related quality of life (QOL), and may reduce overall health care use.<sup>1,2</sup>

Literature reviews conducted in relation to both CR and PR literature report that both fields of chronic disease rehabilitation are effective in reducing hospital admissions and improving QOL in patients attending these programs.<sup>3-5</sup> Reviewers, however, also reported that the quality of the primary randomized controlled trials (RCTs) was variable, and limitations in their methodological quality led to downgrading of some of the evidence. Several potential sources of bias were noted, including unconcealed randomization, unblinded study personnel, incomplete outcome data, and loss to follow-up; also, the literature revealed a need for the standardization of methods and reporting.

Researchers believe that responsiveness of patient outcome measures (POMs) is an important but often overlooked area. This may influence the variability of evidence currently available because some POMs may not be responsive to the rehabilitation process and may not be suitable outcome measures in this area of research. Generic instruments are useful in measuring health outcomes against normative data; however, they are often not able to pick up specific outcomes for disease-associated programs. Disease-specific instruments, although focused on the problem at hand, are not able to measure outcomes across disease programs, and individualized instruments, although they specifically address the patient's own concerns, these instruments are often difficult to administer. All of these factors contribute toward determining the ideal POM to evaluate chronic disease rehabilitation programs.<sup>3-5</sup>

We used our previous systematic effectiveness review of this literature as the basis of this statistical analysis.<sup>6</sup> The bulk of articles were observational studies, and the main emphasis of the review was to provide an overview of the performance of the POMs. This review also conducted a quality assessment of the literature, which included RCTs, comparative studies, observational studies, and retrospective studies.<sup>7,8</sup> The quality assessment tool used was adapted from the graded Cochrane Collaboration's system,<sup>9</sup> with an emphasis on the assessment of responsiveness of POMs in these articles.

Reviewers in this field of research suggest that there are 2 aspects of responsiveness: internal responsiveness and external responsiveness. Internal responsiveness is the ability of a measure to change over a prespecified time frame, and external responsiveness is the extent to which changes in a measure over a specified time frame relate to corresponding changes in a reference measure of health status.<sup>10</sup> The aim of this present study was to conduct a secondary analysis of this systematic review to statistically examine the main POM, the Medical Outcomes Study Short Form-36 health survey (SF-36), for internal responsiveness.

## Methods

Meta-analysis is a systematic approach to identifying, synthesizing, and, if applicable, combining results of relevant studies; however, this is usually done through assessing RCTs. Observational data in some cases, however, may be needed to assess the effectiveness of an intervention, and the number of meta-analyses using observational studies has increased substantially in recent years.<sup>11</sup>

The statistical package used for this analysis is Stata.<sup>a</sup> The purpose of this meta-analysis was not to assess pooled effect sizes of each study to determine effectiveness of either CR or PR (this has been done many times), but rather to gather pooled information on the responsiveness of the main POMs in this setting. Measurement of responsiveness for pre to post changes in POMs is for change in health status per each interventional group of each study group.

**Hypothesis.** The measurement of QOL by POMs in CR and PR is not responsive to change in health status pre- to post-rehabilitation program.

In articles that use a pre to post study design, it is important to determine if the POM used is responsive to the rehabilitation process, and in most cases, this is done by assessing previous literature to determine if this is the case. Of the 83 articles assessed for the literature review, we found that only 47% checked POMs for responsiveness in previous studies.

The SF-36 was the most commonly used generic measurement tool in both the CR and PR literature. In the PR literature, the disease-specific Chronic Respiratory Questionnaire (CRQ) was the most frequently used instrument.<sup>5</sup> Using the results of the previous literature review, this meta-analysis assessed articles in both the CR and PR literature that used the SF-36 and pooled the statistical results; however, this study was limited by the number of common indices used to measure the magnitude of change over time or responsiveness.

## A statistical assessment of the literature

To find an association between SF-36 pre and post, each individual intervention group (“within”) mean scores in the CR literature were used. Initially, we undertook a correlation analysis to define the degree to which the 2 sets of variables were related. These scores can be taken as similar to a matched study.<sup>12</sup> In the PR literature, we undertook a correlation analysis between SF-36 and CRQ each interventional group's change scores. Correlation is defined as the quantification of the degree to which 2 variables are related, providing that the relationship is linear.<sup>12,13</sup>

We identified 83 studies based on the inclusion and exclusion criteria. Of the articles published after 2000, 25 reported CR studies that used the SF-36; 16 reported PR studies that used the CRQ; and 9 reported PR studies that used the SF-36.

The SF-36 has 2 compound domains, a Physical Composite Score (PCS) and a Mental Composite Score (MCS). These composite scores are also made up of a number of separate domains: physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health.<sup>3,7</sup> Some studies used PCS and MCS composite scores as well as individual domain scores; some studies, such as RCTs, used 2 patient groups, whereas other studies, such as quasi-experimental studies, had up to 4 different patient groups. Each group of patients was taken to represent a single group so that studies such as the RCTs were listed twice.

In the CR literature there were 13 studies that listed PCS and MCS “within” pre and post mean scores, and 22 groups altogether that provided these scores (table 1).

The determination of responsiveness of POMs in CR and PR has not been conducted before in the literature. For the

CR literature, we constructed POMs scatterplots (see figs 1-4 for PCS and MCS domains; see supplemental appendix S1 and supplemental figs S1-S16 for individual domains, available online only at <http://www.archives-pmr.org/>), but a scatterplot can only illustrate the degree of correlation between the 2 variables.<sup>13,14</sup> Each scatterplot shows the variables along with the author of each article and a line of best fit. All scatterplots for the CR literature showed positive slopes, so that as X (the independent variable) increased, Y (the dependent variable) increased.

The overall precision of each domain in terms of responsiveness was determined by how closely the line of best fit and the degree of scatter correlate; the greater the degree of scatter around the line of best fit, the less responsive the domain. For completeness, we also conducted Bland-Altman plots to plot the differences among pre and post measurements against premeasurement of each individual interventional mean score.<sup>15</sup> The Bland-Altman method is designed as an absolute measure of agreement between 2 measuring instruments that are on the same scale of measurement.<sup>16</sup>

A meta-regression was conducted on all 22 groups of data for the PCS and MCS scales of the SF-36 to determine average responsiveness of each composite score (tables 2 and 3). Calculation of standard deviations for pre to post change scores and then standard errors for change scores was based on pre and post SD.<sup>17</sup> Meta-regression analysis was conducted on individual domain scores of the SF-36, but this only included 9 groups of data because some groups did not list standard errors. The meta-regression allowed for calculation of an average change score, which is different from the mean change score; a simple change score assumes the data are from individuals and the only variance is error variance, whereas the meta-analysis assumes there is between study variance as well as error variance.<sup>18</sup>

**Table 1** SF-36 composite domains for CR pre and post “within” change scores

Study	Year	PCS Pre	PCS Post	MCS Pre	MCS Post	PCS Change	MCS Change	SD Change	SE Change
Stauber 1	2013	39.5	40.4	52.1	54.1	0.9	2	10.1237	1.2188
Stauber 2	2013	44.0	48.9	47.4	50.8	4.9	3.4	9.2065	0.4037
Izawa 1	2010	46.3	48.8	48.1	50.1	2.5	2	5.663	0.364
Izawa 2	2010	44.0	45.2	51.1	53.4	1.2	2.3	6.2	0.4384
McGrady	2009	37.9	47.3	47.3	50.1	9.4	2.8	10.4843	0.8017
Fast 1	2009	32.9	42.2	42.5	48.4	9.3	5.9	7.7078	1.1893
Fast 2	2009	33.6	43.3	45.1	51.2	9.7	6.1	7.7485	1.1956
Riaz	2009	42.3	49.9	54.8	54.9	7.6	0.1		2.1791
Sanderson 1	2007	37.0	44.0	49.0	53.0	7	4	10.5357	0.4918
Sanderson 2	2007	33.0	40.0	49.0	54.0	7	5	10.5357	0.8408
Gunstad 1	2007	39.8	44.2	42.8	45.9	4.4	3.1	6.8566	0.5213
Gunstad 2	2007	38.3	43.1	41.9	45.5	4.8	3.6	7.0712	0.8392
Leal	2005	39.1	42.6	40.4	42.0	3.5	1.6	6.9486	0.9827
Auon	2004	39.3	44.8	40.2	47.3	5.5	7.1	23.0671	2.3301
Sin	2004	47.4	62.9	57.0	72.4	15.5	15.4		2.8061
Focht 1	2004	41.9	42.7	51.5	52.8	0.8	1.3	10.4183	1.8136
Focht 2	2004	43.4	45.3	56.0	57.1	1.9	1.1	9.3475	1.478
Focht 3	2004	36.6	40.5	54.9	55.7	3.9	0.8	9.1285	1.5007
Focht 4	2004	43.1	44.9	52.9	56.1	1.8	3.2	8.9835	1.4769
Marrin	2000	37.9	46.1	47.9	49.9	8.2	2	11.2054	0.9983
O'Farrell 1	2000	34.1	39.2	43.7	48.7	5.2	4.9	11.425	1.3655
O'Farrell 2	2000	39.3	45.5	47.9	51.3	6.1	3.7	8.8504	0.4971

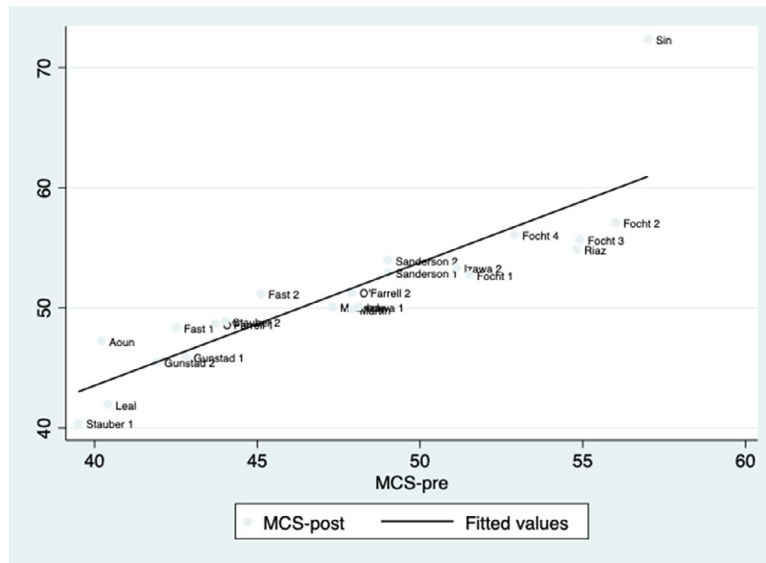


Fig 1 SF-36 MCS pre and post.

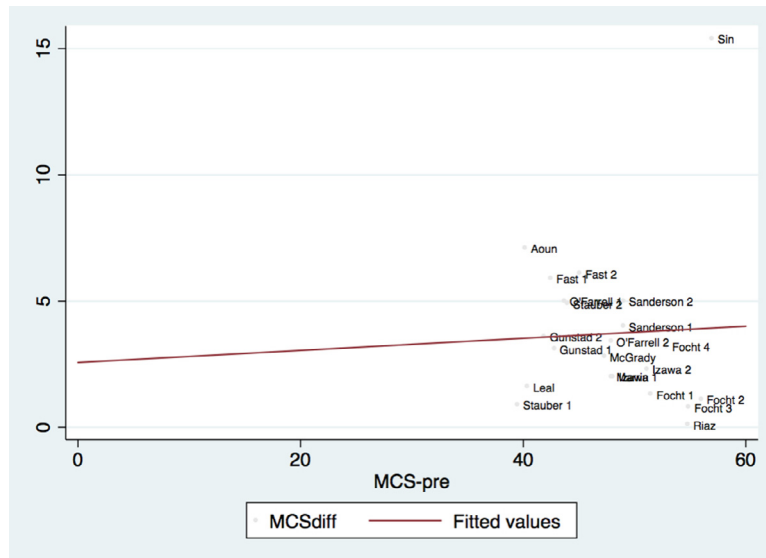


Fig 2 SF-36 MCS Bland-Altman plot with line of best fit.

There were 12 studies that listed individual domain interventional group pre and post mean scores and 19 groups of data (table 4). To calculate average change scores for the individual SF-36 domains, only the 9 articles used for the meta-regression were able to be used (table 5; fig 5), displaying the mean of the average change scores for the individual domain scores of the SF-36 for CR. This allows an estimation of how responsive each individual domain score is in relation to each other.

The CRQ that is used in the PR literature has 20 items that ask about the level of impairment in the domains of dyspnea, fatigue, emotional function, and mastery. Patients express their degree of impairment from 1-7 on a 7-point Likert-type scale.<sup>15,17</sup> There were 8 studies (Puhan 2010, Beauchamp 2010, Maltais 2008, Puhan 2007, Stulberg 2002, White 2002, Singh 2001, Troosters 2000) that provided information for the CRQ scores and 10 groups of data sets; 5

studies (Puhan 2007, Reis 2005, California 2004, Stulberg 2002, White 2002) provided scores for the SF-36 PCS and MCS, with 9 groups of data (table 6). Spearman correlation coefficients were calculated for the SF-36 and CRQ change scores to assess how well the relationship between the 2 variables can be described. The Spearman correlation increases in magnitude as the X and Y variables become closer to monotone functions of each other (table 7).

## Results

### Cardiac rehabilitation; SF-36 composite domain

#### Descriptive analysis of CR SF-36 composite domains

All scatterplots for the CR literature showed positive slopes so that as X (the dependent variable) increased, Y (the

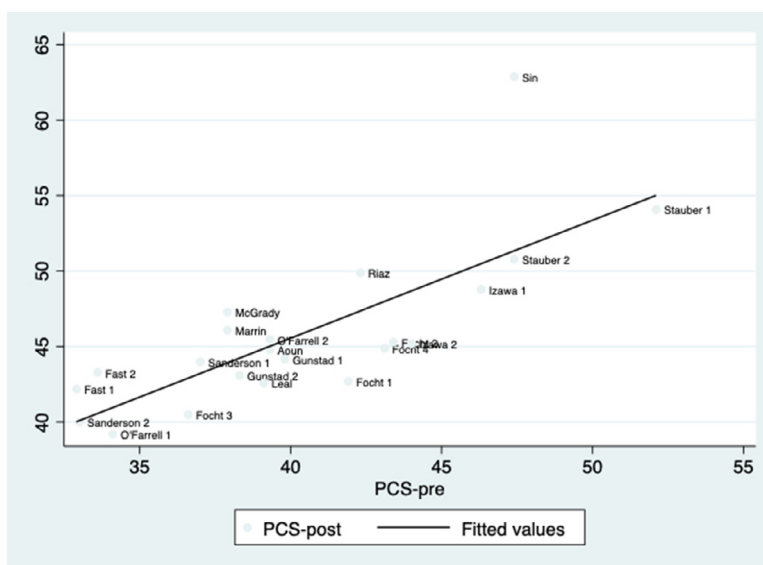


Fig 3 SF-36 PCS pre and post.

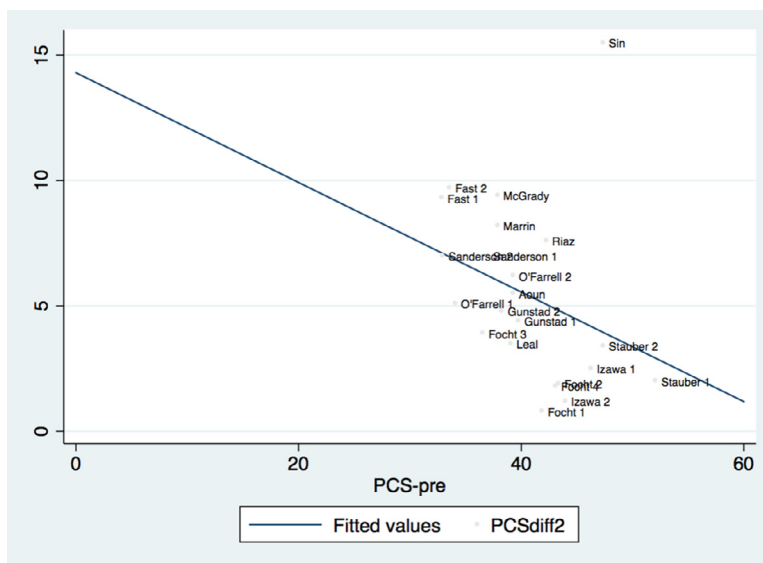


Fig 4 SF-36 PCS Bland-Altman plot with line of best fit.

independent variable) increased with association. The strength of association of the dots refers to the degree of scatter of the plot; if the dots are scattered around a line, the relationship is strong, and if widely spread, the relationship between variables is weak. The mean and SD are listed for each individual domain; they are followed by *r*, the correlation coefficient with its *P* value and the percentage of variance.<sup>13,18</sup>

1. The SF-36 MCS (see [figs 1 and 2](#); [table 3](#)). The average mean change score for the MCS is 2.5. The MCS composite domain showed MCS pre ( $47.61 \pm 5.40$ ) and MCS post ( $51.32 \pm 6.40$ ) with  $r=0.95$ ,  $P \leq .001$ , and 90.63% of the variance explained.

2. The SF-36 PCS (see [figs 3 and 4](#); [table 4](#)). The average mean change score for the PCS is 13.0. There was a positive correlation between the 2 variables PCS pre ( $40.30 \pm 5.11$ ) and PCS post ( $45.79 \pm 5.22$ ) with  $r=0.69$ ,  $P \leq .001$ , and 48.95% of the variance explained.

Both SF-36 composite scores have Sin (2004) as an outlier, which is a retrospective study. Removing this outlier produced little change in the Bland-Altman plots for the composite scores; however, there was some change in the PCS pre–post scatterplot, which showed an increase in space in the lower scores. These changes, however, were not enough to change the overall outcome of these domains.

**Table 2** Meta-regression of SF-36 PCS composite domain pre and post “within” scores

Meta-regression				No. of Observations	22
REML estimate of between-study variance				Tau 2	5.645
% residual variation due to heterogeneity				I-squared res	60.92%
With Knapp-Hartung modification					
PCS change	Coefficient	SE	<i>t</i>	<i>P</i> value>  <i>t</i>	95% confidence interval
cons	5.111457	.6829245	7.48	0.000	3.691238-6.531676
PCS average change score=13.0					

NOTE. REML means, residual (restricted) maximum likelihood and is an improved algorithm for the estimation of the between-study variance  $r$  squared. “res” is the residual variation, “cons” is the constant variable.

**Table 3** Meta-regression of SF-36 MCS composite domain pre and post “within” scores

Meta-regression				No. of observations	22
REML estimate of between-study variance				Tau 2	1.677
% residual variation due to heterogeneity				I-squared res	62.15%
With Knapp-Hartung modification					
MCS change	Coefficient	SE	<i>t</i>	<i>P</i> value>  <i>t</i>	95% confidence interval
cons	2.644835	.4902175	5.40	0.000	1.625372 -3.664298
MCS average change score=2.5					

NOTE. REML means, residual (restricted) maximum likelihood and is an improved algorithm for the estimation of the between-study variance  $r$  squared. “res” is the residual variation, “cons” is the constant variable.

### Analysis of SF-36 individual domains

Individual domains ( $n=19$ ) showed a positive correlation between pre and post variables for all domains. Listed for each individual domain is the mean and SD; this is followed by  $r$ , the correlation coefficient with its  $P$  value and the percentage of variance.<sup>13,18</sup>

1. The SF-36 Physical Function domain (see supplemental appendix S1, supplemental figs S1 and S2, available online only at <http://www.archives-pmr.org/>). The scatterplot for this domain shows most of the studies scoring between 50-70. The Bland-Altman plot shows most scores from 50-70, with Sagar 1 (2012) as an outlier. Physical Function pre ( $48.68 \pm 16.43$ ) and Physical Function post ( $64.84 \pm 13.83$ ) with  $r=0.80$ ,  $P \leq .001$ , and 63.92% of the variance explained.
2. The SF-36 Role Physical domain, (see supplemental appendix S1, supplemental figs S3 and S4, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show a widely scattered field over a large range of SF-36 scores and differences. Role Physical pre ( $32.88 \pm 17.32$ ) and Role Physical post ( $55.76 \pm 12.90$ ) with  $r=0.49$ ,  $P \leq .005$ , and only 24.28% of the variance explained.
3. The SF-36 Bodily Pain domain (see supplemental appendix S1, supplemental figs S5 and S6, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show most scores clustered between 50-70, and the differences are quite widely spread, with Sagar 1 (2012) and Jamieson 2 (2002) as outliers. Bodily Pain pre ( $54.62 \pm 14.25$ ) and Bodily Pain post ( $67.01 \pm 13.42$ ) with  $r=0.88$ ,  $P \leq .001$ , and 78.69% of the variance explained.
4. The SF-36 General Health domain (see supplemental appendix S1, supplemental figs S7 and S8, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show a wide scatter of studies in the lower SF-36 scores and more clustering of studies from the SF-36 scores  $\geq 55-65$ . General Health pre ( $58.34 \pm 9.40$ ) and General Health post ( $62.63 \pm 9.60$ ) with  $r=0.82$ ,  $P \leq .001$ , and 68.06% of the variance explained.
5. The SF-36 Vitality domain (see supplemental appendix S1, supplemental figs S9 and S10, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show studies distributed around the line of best fit. Vitality pre ( $47.88 \pm 9.05$ ) and Vitality post ( $58.21 \pm 7.44$ ) with  $r=0.79$ ,  $P \leq .001$ , and 62.28% of the variance explained.
6. The SF-36 Social Function domain (see supplemental appendix S1, supplemental figs S11 and S12, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show that most of the studies are clustered from 50-70, with Jamieson 2 (2002) as an outlier. Social Function pre ( $56.23 \pm 18.25$ ) and Social Function post ( $71.62 \pm 16.96$ ) with  $r=0.85$ ,  $P \leq .001$ , and 73.17% of the variance explained.
7. The SF-36 Role Emotional domain (see supplemental appendix S1, supplemental figs S13 and S14, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show most studies clustered between 40-80 on the SF-36 range of scores, with Sagar 1 (2012) as an outlier. Role Emotional pre ( $48.65 \pm 17.16$ ) and Role Emotional post ( $68.45 \pm 13.98$ ) with  $r=0.52$ ,  $P \leq .001$ , and only 27.18% of the variance explained.

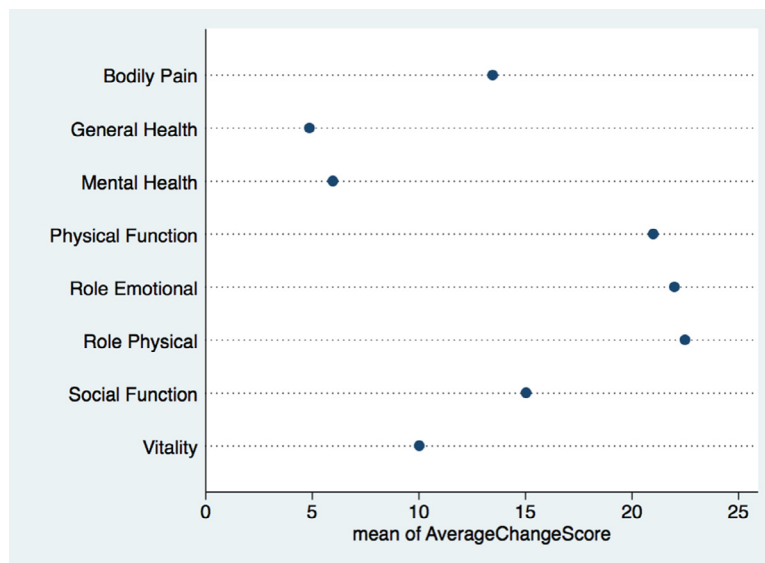
**Table 4** “Within” mean pre and post scores for individual domains

Study	Year	Physical Fn Pre	Physical Fn Post	Role PhysicalPre	Role Physical Post	Bodily Pain Pre	Bodily Pain Post	General Health Pre	General Health Post	Vitality Pre	Vitality Post	Social Fn Pre	Social Fn Post	Role Emot Pre	Role Emot Post	Mental Health Pre	Mental Health Post
Sagar 1	2012	27.7	64.8	18.3	68.3	44	69	44	57.7	43.3	59.7	37.5	57.5	21.5	75.2	48.8	66.9
Sagar 2	2012	47.7	65.5	36.7	73.3	62.5	80.8	66.7	75.3	61	70.7	56.7	78.3	46.6	76.1	62.4	75.5
McKee	2009	67.5	74.4	43.7	51.6	71.7	80.2	66.7	72.2	58.8	62.6	77.2	81.6	58.2	65.7	72.4	74.9
Fast 1	2009	31.3	42.5	28.3	38.4	35.6	45.5	47	48.2	38.2	45.8	34.5	44.6	37.4	43.9	42.8	48.9
Fast 2	2009	32.8	44.6	29.1	39.8	36.5	47.3	47.2	49	43.3	49.2	35.8	47.2	35.7	46.8	47.6	51.6
Riaz	2009	48.8	49.3	41.1	47.3	48.9	53.8	47	51.3	47	56.3	47	51.6	47	51.8	47	54.2
Mandel 1	2007	58.3	70.5	37.9	60	54.7	69.7	64.3	70.8	47	57	66.5	81.4	60.6	84.5	72.2	75.2
Mandel 2	2007	48.6	63.6	16.4	44.4	52.3	63.4	58.3	63.4	41.9	55.8	62.3	77.7	50	66.7	68.3	75.5
Aldana 1	2006	67.9	83.3	43.8	72.8	66.5	79.9	44.6	63.9	54.8	57	66.5	81.4	60.6	84.5	72.2	75.2
Aldana 2	2006	70.4	87.7	49.1	79.6	81.3	86.1	59.1	64.4	68.6	55.8	62.3	77.7	50	66.7	68.3	75.5
Auon	2004	62.2	74.5	30.9	52.7	56.6	68.8	49.9	61.2	43.5	67.5	65.5	82.9	54.8	73.6	69.5	78.8
Jamieson 1	2002	50.7	74.3	12.6	52.5	53.4	70.2	64.1	66.5	42	60	58.5	82.1	46.8	72.1	62.2	76.4
Jameison 2	2002	44.5	69.5	13	47.7	72.3	71	62.5	64.1	37	54.5	50.3	80.8	45.8	73.3	69.6	77.8
Marrin	2000	62.7	78.7	25.5	59.8	58.6	71.4	61.8	67.1	47.4	59.7	66.1	81.6	63	72.7	71.1	74
Sledge	2000	57	71.1	22.2	52.2	54.1	63.4	51.9	58.3	48.4	56.6	66.1	80.3	56.3	85.9	70.7	79.6
O'Farrell 1	2000	52.9	63	17.5	43.8	49.6	60.1	55.2	61.4	42.9	48.2	52.6	75.5	50.7	67.6	63.9	71.5
O'Farrell 2	2000	66.4	81.1	30.8	66.4	62.5	75.9	64.7	69.6	50.1	62.8	65	83.9	67.2	77.4	72.5	75.5

Abbreviations: Emot, emotional; Fn, function.

**Table 5** SF-36 individual mean change scores for CR

Domain	Mean Change Score	SD	SE
Physical Function	19.8	66	5.33
Role Physical	23.12	17.3	5.76
Bodily Pain	12.38	7.32	2.44
General Health	4.32	4.47	1.49
Vitality	10.32	5.63	1.87
Social Function	15.38	8.78	2.92
Role Emotional	19.76	15.83	5.27
Mental Health	8.22	5.79	1.93

**Fig 5** Mean of average change score of SF-36 individual domains for CR.

8. SF-36 Mental Health domain (see supplemental appendix S1, supplemental figs S15 and S16, available online only at <http://www.archives-pmr.org/>). The scatterplot and Bland-Altman plot show most studies clustered between the 60–80, with Sagar 1 (2012) as an outlier. Mental Health pre ( $61.00 \pm 11.72$ ) and Mental Health post ( $69.22 \pm 11.21$ ) with  $r=0.89$ ,  $P \leq .001$ , and 79.85% of the variance explained.<sup>12,16</sup>

A number of the scatterplots show that 2 of the studies have consistent outliers; these are Sagar 1 (2012), which had a population of cardiac patients who came to rehabilitation after coronary artery bypass grafting, and Jamieson 2 (2002), which had a population of older patients (aged 65–74y) in CR. Five of the SF-36 domain scatterplots have either 1, or both, of these studies as outliers; we decided to remove these studies and rerun the scatterplots. Three out of the 5 scatterplots showed some difference after removing these outliers: Physical Function, Role Emotional, and Mental Health. However, the change was not significant enough to warrant changing the outcome of these domains.

## Pulmonary rehabilitation

### SF-36 Composite “within” change scores and CRQ “within” change scores

1. The SF-36 Physical Composite Score. The Spearman rank correlation score of 0.3932 indicates a weak correlation between the CRQ and the SF-36 PCS.
2. The SF-36 Mental Composite Score. The Spearman rank correlation score of 0.6360 indicates a moderate correlation between the CRQ and the SF-36 MCS.

## Discussion

It is difficult to make a statistical assessment of responsiveness for the SF-36 in the CR literature because it is only possible to compare 1 domain against another. To say that it is not responsive in this setting is more difficult because correlation does not mean agreement.<sup>13</sup>

Combined with the scatterplots and Bland-Altman graphs, the MCS from this assessment is the most responsive



**Table 6** Respiratory studies listing “within” mean change scores for CRQ and SF-36

Author	Year	CRQ “Within” Mean Change Score	Author	Year	SF-36 PCS “Within” Mean Change Score	SF-36 MCS “Within” Mean Change Score
Nilesh	2013	10.2	Jones	2009	0.3	3.3
Dodd	2012	88.7	Lotshaw	2007	3.13	4.83
van Gestel 1	2010	0.64	Verrill	2005	.77	0.25
van Gestel 2	2010	0.48	Reis 1	2005	1.4	1.9
Beauchamp	2010	1.4	Reis 2	2005	1.3	2.2
Jones	2009	0.79	California	2004	2.7	3.9
White 1	2002	8.0	Stuhlbarg 1	2002	1.7	2.2
White 2	2002	4.6	Stuhlbarg 2	2002	1.3	3.2
Singh	2001	14.5	Stuhlbarg 3	2002	4.5	2.8
Troosters	2000	7.0				

**Table 7** SF-36 composite “within” change scores and CRQ “within” change scores

SF-36 Composite Domain	Spearman Correlation Coefficient
PCS	0.3932
MCS	0.6360

of the composite SF-36 domains with  $r=0.95$  and  $P\leq.001$ ; the PCS in comparison shows less ability to discriminate in the higher SF-36 scores.

All the individual domains of the SF-36 in the CR literature were assessed and graded using scatterplots, Bland-Altman plots, and mean change scores from optimum responsiveness to the least responsive. General Health ( $r=0.82$ ,  $P\leq.001$ ), Mental Health ( $r=0.89$ ,  $P\leq.001$ ), and Vitality ( $r=0.78$ ,  $P\leq.001$ ) were the most responsive. They were followed by Bodily Pain ( $r=0.88$ ,  $P\leq.001$ ) and then Social Function ( $r=0.85$ ,  $P\leq.001$ ). The final 3 are Physical Function ( $r=0.79$ ,  $P\leq.001$ ), Role Emotional ( $r=0.52$ ,  $P\leq.001$ ), and Role Physical ( $r=0.49$ ,  $P\leq.005$ ).

To assess responsiveness in the PR literature is even more difficult. A study in 2014 retrospectively reviewed 41 patients with asthma and chronic obstructive pulmonary disease and an average age of 69 years. At the end of the rehabilitation period there was a small change in the SF-36 MCS score but not the PCS score.<sup>19</sup> Yet another retrospective study in 2000 of 22 patients with chronic obstructive pulmonary disease found a significant change in all domains of the SF-36 except for Pain and Role Physical; the scores for MCS and PCS in particular were highly statistically significant.<sup>20</sup> The disparity between these 2 studies with similar patient populations indicates only one of the difficulties involved in this area of research.

In this meta-analysis, Spearman rank correlation coefficient appeared to show that SF-36 PCS (0.39) had a weaker correlation to the CRQ than the SF-36 MCS (0.63). What this shows is that the SF-36 PCS is measuring something different than the CRQ. The CRQ was developed for PR as a disease-specific instrument and is designed to determine how dyspnea affects patients in everyday life.<sup>17</sup>

The SF-36 is an excellent survey instrument with good validity and reliability, and it allows for comparison of normative data across populations. However, it does not perform as well in a clinical setting because it is a lengthy questionnaire and has a low response rate in the >65 age group.<sup>21</sup> There are also

difficulties with any POM when dealing with patients who have poor literacy; these patients may need help filling out lengthy forms such as the SF-36, resulting in possible interference with the patient's choices.<sup>22</sup> The SF-12 was created with this in mind but has not been as well tested as the SF-36. In a clinical setting with measurements taken before and after clinical intervention, the SF-36 does not perform well, evidenced by the results from this meta-analysis and previous literature review.<sup>6</sup>

The development of a new instrument is more complex than was initially believed because there are a number of issues particular to this clinical setting. To be relevant in the rehabilitation setting, a new instrument needs to be relatively simple, record more than a yes/no answer, reflect what actually changes for the patient, reflect patient language for ease of use and understanding, and, most importantly, take “response shift” into account.<sup>23</sup>

“Response shift” can be divided into recalibration, reprioritization, and reconceptualization. It is the last reconceptualization that is most important in this setting.<sup>24</sup> That is, patients' values change from the start of a rehabilitation program to the end of the program.<sup>23</sup> A questionnaire that reflects this problem has yet to be developed and tested in this setting.

Another problem that has not been well researched is that there has been an assumption that pre- to postrehabilitation follows a simple model of change implied by a simple pre-test to post-test approach. This may not necessarily be true; individual change is a continuous process, whether it is physical, emotional, or intellectual, and is also likely nonlinear.<sup>25</sup>

## Study limitations

The usual tests of heterogeneity analysis, sensitivity analysis, and publication bias were not conducted in this review because the meta-analysis was not done on a per study basis but rather on a per study group basis. To conduct these tests would require further analysis on a per study basis, and there have been many meta-analyses conducted on the efficacy of both PR and CR.

One aspect of this analysis that was mentioned in the systematic review is the inconsistency in the reporting of results; which limited our ability to include studies. The main issue with this analysis is that it was essentially performed on a comparison basis and was limited by the statistical analysis

that can be conducted for each article. Further reviews in this area would help as more studies are published.

## Conclusions

The SF-36 has been used in surveys and studies for many years. It is well validated and shows consistent validity and reliability. However, is it an appropriate instrument for CR and PR, and is it responsive in this setting? Some of the domains of the SF-36 are not responsive to CR, and the PCS of the SF-36 appears not to be as responsive to PR, particularly in the area of more severe disease. There are also other aspects of this questionnaire that do not strictly suit this chronic disease rehabilitation setting. It is a lengthy questionnaire and does not lend itself to repeated use. However, to fully rule it out in this area, further research needs to be conducted.

## Supplier

- a. Stata Statistical Software, release 16; StataCorp LLC.

## Corresponding author

Flora-Joan van Rotterdam, PhD, MSc, University of Newcastle, 82 Fennell Crescent, Blackalls Park, NSW 2283, Australia. *E-mail address:* [joanvr123@bigpond.com](mailto:joanvr123@bigpond.com).

## References

1. Shah P, Mountain D. The medical model is dead-long live the medical model. *Br J Psychiatry* 2007;191:375-7.
2. Wagner E, Davis C, Schaefer J, Van Korff M, Austin B. A survey of leading Chronic Disease Management Programs: are they consistent with the literature. *Manag Care Q* 1999;7:56-66.
3. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic pulmonary disease. *Cochrane Database Syst Rev* 2015;2:1-15.
4. Golmohammadi K, Jacobs P, Sin D. Economic evaluation of a community-based pulmonary rehabilitation program for chronic obstructive pulmonary disease. *Lung* 2004;182:187-96.
5. Bradley E, Bickenbach J, Glozier N, et al. ICF/International classification of functioning disability and health. Geneva, Switzerland: World Health Organization; 2001.
6. van Rotterdam F, Hensley M, Hazelton M. A Comparative effectiveness review: responsiveness of patient outcome measures in cardiac and pulmonary rehabilitation. *J Cardiopulm Rehabil Prev* 2019;39:73-84.
7. Stewart A. Functional status and well-being of patients with chronic conditions: results from the medical outcomes study. *JAMA* 1989;262:1-7.
8. Sanderson S. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666-76.
9. Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH, on behalf of the Cochrane Applicability and Recommendations Methods Group. The assessment of the risk of bias. *Cochrane Database Syst Rev*.
10. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459-68.
11. Stroup D, Thacker S, Olsen C, et al. Characteristics of meta-analysis related to acceptance for publication in a medical journal. *J Clin Epidemiol* 2001;54:655-60.
12. Juul S, Frydenberg M. An introduction to STATA for health researchers. p 199, 4th ed. College Station, Texas: STATA Press; 2014. p. 199-204.
13. Pagano M, Gauvreau K. Principles of biostatistics. Belmont, California: Wadsworth Publishing company; 1993. p. 363-78.
14. Borenstein M, Hedges L, Higgins J, et al. Introduction to meta-analysis. Pondicherry, India: John Wiley and Sons Ltd; 2009.
15. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
16. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. New York: Oxford University Press; 1989.
17. Puhan M, Guyatt G, Goldstein G, et al. Relative responsiveness of the Chronic Respiratory Questionnaire, St Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respir Med* 2007;101:308-16.
18. Fu R, Vandermeer B, Shamlayan TA, et al. Handling continuous outcomes in quantitative synthesis. *Methods guide for effectiveness and comparative effectiveness reviews*. Rockville: Agency for Healthcare Research and Quality; 2013.
19. Limsuwat C, McClellan R, Amiri HM, et al. Pulmonary rehabilitation improves only some domains of health related quality of life measured by the Short Form-36 questionnaire. *Ann Thoracic Med* 2014;9:144-8.
20. Benzo R, Flume PA, Turner D, et al. Effect of pulmonary rehabilitation on quality of life in patients with COPD: the use of SF-36 summary scores as outcome measures. *J Cardiopulm Rehabil* 2000;20:231-4.
21. Andresen EM, Gravitt G, Aydelotte M, et al. Limitations of the SF-36 in a sample of nursing home residents. *Age Aging* 1999;28:562-6.
22. Jahagirdar D, Kroll T, Ritchie K, et al. Using patient reported outcome measures in health services: a qualitative study on including people with low literacy skills and learning disabilities. *BMC Health Serv Res* 2012;12:431.
23. van Rotterdam FJ, Hensley M, Hazelton M. Chronic pulmonary and chronic cardiac rehabilitation: staff perspectives and patient experiences. *J Med Health Sci* 2017;6:1-7.
24. Ubel PA, Peeters Y, Smith D. Abandoning the language of "response shift": a plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. *Qual Life Res* 2010;19:465-71.
25. Fisher AJ, Newman MG, Molenaar PC. A quantitative method for the analysis of nomothetic relationships between ideographic structures: dynamic patterns create attractor states for sustained post-treatment change. *J Consult Clin Psychol* 2011;79:552-63.