ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/11595008

Do SF-36 Summary Component Scores Accurately Summarize Subscale Scores?

Article *in* Quality of Life Research · February 2001

DOI: 10.1023/A:1012552211996 · Source: PubMed

CITATIONS
215

READS
329

3 authors, including:

Jan Karlsson
Örebro University Hospital

84 PUBLICATIONS **10,163** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Psychosocial functioning related to stigma: construct validity, responsiveness and Swedish reference data norms of the Stigma-related Social problem scale (SRSP) View project

All content following this page was uploaded by Jan Karlsson on 07 June 2016.

The user has requested enhancement of the downloaded file.

Quality of Life Research **10**: 395–404, 2001.
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Do SF-36 summary component scores accurately summarize subscale scores?

Charles Taft, Jan Karlsson & Marianne Sullivan
Health Care Research Unit, Sahlgrenska University Hospital, Göteborg University, Göteborg, Sweden

Accepted in revised form 29 March 2001

Abstract

Standard scoring algorithms were recently made available for aggregating scores from the eight SF-36 subscales in two distinct, higher-order summary scores: Physical Component Summary (PCS) and Mental Component Summary (MCS). Recent studies have suggested, however, that PCS and MCS scores are not independent and may in part be measuring the same constructs. The aims of this paper were to examine and illustrate (1) relationships between SF-36 subscale and PCS/MCS scores, (2) relationships between PCS and MCS scores, and (3) their implications for interpreting research findings. Simulation analyses were conducted to illustrate the contributions of various aspects of the scoring algorithm to potential discrepancies between subscale profile and summary component scores. Using the Swedish SF-36 normative database, correlation and regression analyses were performed to estimate the relationship between the two components, as well as the relative contributions of the subscales to the components. Discrepancies between subscale profile and component scores were identified and explained. Significant correlations ($r = -0.74$, -0.67) were found between PCS and MCS scores at their respective upper scoring intervals, indicating that the components are not independent. Regression analyses revealed that in these ranges PCS primarily measures aspects of mental health (57% of variance) and MCS measures physical health (65% of variance). Implications of the findings were discussed. It was concluded that the current PCS/MCS scoring procedure inaccurately summarizes subscale profile scores and should therefore be revised. Until then, component scores should be interpreted with caution and only in combination with profile scores.

Key words: Health indexes, Quality of life, SF-36, Summary scores

Introduction

The Medical Outcomes Study (MOS) SF-36 is one of the most widely used generic measures of health status today. Originally, the SF-36 was designed to measure functioning and well-being in eight basic health dimensions. Scores derived for these dimensions are commonly presented in the form of health profiles.

Although profiles of health dimension scores provide comprehensive and diversified information, they may delineate measured health status in more detail than is required or economical for the purposes of a particular study or survey. Furthermore, multiple health dimensions complicate statistical analyses and interpretation [1]. Consequently, the developers of the SF-36 recently made available standard scoring algorithms for aggregating scores from the eight SF-36 subscales in two distinct, higher-order summary scores: Physical Component Summary (PCS) and Mental Component Summary (MCS). Derived from principal Components analyses of subscale scores in a general US population sample, these components were found to explain between 80–85% of the reliable variance in subscale scores. Internal consistency and test–retest reliability estimates in general populations surveys have been reported to be high [2]. The validity of these summary measures in discriminating among clinically meaningful groups has, as noted by the developers, also been good [1, 2].

Compared to SF-36 health profiles, the summary component scores have been used as primary or secondary outcome measures in relatively few empirical studies to date. Although the bulk of these studies has confirmed the validity of these measures, two recent studies have indicated that all may not be well with PCS and MCS. One study, which tested the construct validity of these components in a sample of patients with rheumatoid arthritis, found that 10% of the variance in MCS mental health scores was explained by an instrument measuring physical disability [3]. In the second study, presenting cross-sectional and prospective data from a sample of patients initiating depression treatment, discrepancies were found to exist between SF-36 health profile scores and component scores [4]. Impaired physical health indicated in profile scores was not reflected in the PCS score at baseline and at follow-up improvement in physical health subscales was instead shown as a decline in PCS. The results from these two studies raise fundamental questions regarding the validity of the SF-36 component summary scores, i.e. what the components are measuring and the accuracy of those measurements, and warrant further examination.

Potential problems associated with the SF-36 summary components are most simply illustrated by comparing the range of the profile scores with that of the summary scores. SF-36 health profiles represent a state of optimal physical and mental health as one in which all eight subscale values reach the maximum scale value of 100, while poorest possible physical and mental health is represented by minimum values (0) on all subscales. As summary indexes it follows that PCS and MCS values computed from these subscale values should also reflect best vs. worst possible health status in their respective spheres. The range of expected scores, given the range of profile scores, is about 20–58 for PCS and 17–62 for MCS. However, the true range of PCS and MCS scores is in fact nearly double that expected from profile scores (2–76 for PCS and −1–81 for MCS). How is it possible to attain, for example, a PCS score above 58, when profile scores are all already at optimal values? What does a MCS score below 17 imply, when at this point all subscale values already indicate poorest possible health status?

The specific aims of this paper were to explore (1) the relationship between SF-36 subscale scores and PCS/MCS scores, (2) the relationship between PCS and MCS scores, and the magnitudes of their potential effects on one another and (3) their implications for interpreting research findings.

Methods

PCS/MCS scoring algorithm

The SF-36 summary components are computed following a standardized three-step procedure. First, all eight subscale scores (range = 0–100) are standardized using a linear z-score transformation. Z-scores are calculated by subtracting subscale means for the general US population sample from each individual's subscale scores and dividing the difference by the standard deviation of the US sample. Second, z-scores are multiplied by the subscale factor score coefficients (see Table 1) for PCS and MCS and summed over all eight subscales. Finally, t-scores are calculated by multiplying the obtained PCS and MCS sums by 10 and

Table 1. US subscale means, standard deviations and factor score coefficients for PCS and MCS

Subscale	Mean	STD	PCS coefficients	MCS coefficients
PF	84.52404	22.89490	0.42402	−0.22999
RP	81.19907	33.79729	0.35119	−0.12329
BP	75.49196	23.55879	0.31754	−0.09731
GH	72.21316	20.16964	0.24954	−0.01571
VT	61.05453	20.86942	0.02877	0.23534
SF	83.59753	22.37642	−0.00753	0.26876
RE	81.29467	33.02717	−0.19206	0.43407
MH	74.84212	18.01189	−0.22069	0.48581

Reproduced from Ware et al. [2].

adding 50 to the product, to yield a mean of 50 and a standard deviation of 10 for the US norm population. For scoring international versions of the SF-36, it is recommended that *z*-scores be calculated using national means and standard deviations, but that US factor score coefficients be maintained in order to facilitate international comparisons. The scoring procedure is explained in detail elsewhere [2, 5].

Subject sample

The Swedish SF-36 national normative sample comprises 8930 subjects (15–93 years, mean age 43) from seven general population studies (average response rate = 68%) conducted between 1991–1992 [6]. Data was collected using mail self-administered questionnaires in all seven studies. The distributions of major sample characteristics (sex, age, education, and marital and employment status) corresponded to those of the general Swedish population. Since the PCS/MCS scoring algorithm requires that respondents have scores on all eight subscales for PCS/MCS scores to be computed, the normative sample used here consisted of 8004 individuals with computable component scores [5].

Analyses

The analyses are divided into three sections. The first section consists of theoretical analyses explaining and illustrating how various aspects of the SF-36 scoring algorithm influence the relationship between subscale and component scores. The second and third sections consist of analyses using the Swedish SF-36 normative database to test the empirical implications of the algorithm on the relationships between the two component scores and between subscale and component scores. In the second section, product-moment correlations were calculated between PCS and MCS scores at different intervals of the scoring range. Based on the analyses in the first section, it was hypothesized that correlations between upper range PCS and MCS scores would be significant and negative; correlations between lower range PCS and MCS scores would be lower, but still significant and negative; and correlations over the entire scoring range would be low. In the third section, stepwise regression analyses were performed to examine the proportion of the variance in PCS and MCS scores explained by each of the eight SF-36 subscales at different scoring levels. It was hypothesized that at upper scoring ranges each component would primarily measure subscales assumed to be most strongly associated with the other component, i.e. most of the variance in upper range PCS scores would be explained by mental health subscales and that most of the variance in upper range MCS scores would be explained by physical health subscales.

The first set of analyses used US population means and standard deviations for computing *z*-scores, while analyses using the Swedish SF-36 normative database used Swedish means and standard deviations in accordance with SF-36 guidelines [2]. All data analyses were performed using SAS version 6.12.

Results

How do PCS and MCS scores relate to subscale scores?

According to the scoring algorithm, PCS and MCS scores are a function of not only the factor score coefficients, but also the size and direction of the deviation of individual subscale scores from the subscale population mean, as well as the population subscale standard deviation. Obviously, where individual scores equal population means (although empirically impossible) neither the population standard deviation nor the scoring coefficients contribute to component scores. On the other hand, the greater the deviation from the mean and the smaller the variance associated with a particular subscale, the greater is the impact of the scoring coefficients on component scores. The direction of the impact is determined both by the valence of the scoring coefficient as well as the direction of the deviation of scores in relation to the population mean. It is important to note that all eight subscales are used in the calculation of both PCS and MCS scores. US means, standard deviations and scoring coefficients are shown in Table 1 [2].

Subscale scoring coefficients. PCS and MCS scoring coefficients are factor scoring coefficients derived from a principal components analysis (orthogonal rotation, varimax method) of the US population sample. Factor scoring coefficients represent the relative contribution of each subscale to each derived component. As shown in Table 1, a total of five subscales are positively weighted for PCS and three are negatively weighted, while for MCS four subscales are positively weighted and four are negatively weighted. Furthermore, the size of the coefficients varies across subscales and the rank order of the PCS coefficients is the inverse of the MCS coefficients. Importantly, all subscale scores contribute concurrently and, with the exception of vitality (VT), inversely to PCS and MCS scores (Table 1). Implicit in this system is a theoretical assumption of the interdependence of physical and mental health, where these two components of health are in effect opposed.

Subscale standard deviations. A second determinant of PCS/MCS scores is the population standard deviation of the subscales. As noted earlier, the smaller the standard deviation, the greater the subscale contributes to PCS/MCS scores and vice versa. Importantly, the population standard deviations vary in size between subscales. This implies that the relative weight of each subscale on PCS/MCS scores varies not only in proportion to the size of its scoring coefficient but also its standard deviation since, by simple algebraic transformation, subscale weights are the scoring coefficient divided by the standard deviation. Variation in standard deviations across subscales means that the relative contribution of subscales ascribed PCS/MCS from principal component analyses differs from true subscale weights (Figure 1). For example, about 20% of PCS scores is determined by the subscale role physical (RP) based on coefficients alone, however, due to its relatively large standard deviation this subscale actually contributes only about 14%, making its contribution actually less than the subscales bodily pain (BP) and general health (GH) with lower scoring coefficients. Noteworthy here is that the negative contribution of the mental health (MH) subscale to PCS scores is actually greater than that of RP, meaning that mental health will entirely offset RP when both deviate in the same direction and amount from their respective population means.

Subscale deviations from population means. The direction of the deviation of subscale scores from population means is important in that an individual subscale score above its population mean contributes to component scores in the manner defined by the valence of the scoring coefficient, while scores below the mean contribute inversely. For example, a score above the population mean of 84.52 on the physical functioning (PF) subscale

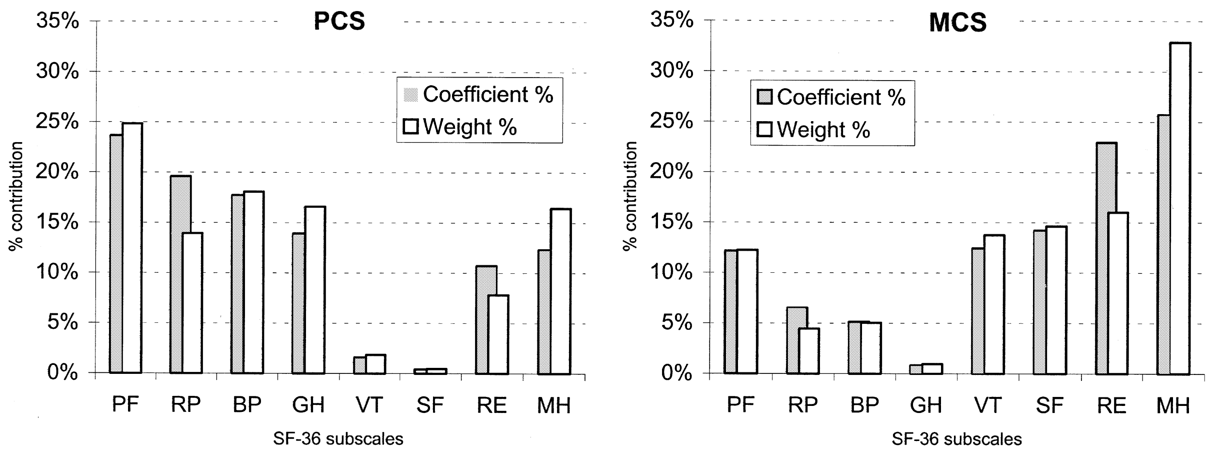

Figure 1. Relative contribution (%) of subscale scoring coefficients vs. weights (coefficients/Std) to PCS and MCS scores, assuming equal deviations of subscale scores from population means. The relatively large standard deviations of the role functioning subscales RP and RE reduce their contribution to component scores. Noteworthy is that the true relative (negative) contribution of MH to PCS scores actually outweighs that of RP. PF = Physical Functioning, RP = Role Physical, BP = Bodily Pain, GH = General Health, VT = Vitality, SF = Social Functioning, RE = Role Emotional, MH = Mental Health, PCS = Physical Component Summary, MCS = Mental Component Summary.

implies that PF contributes positively to PCS and negatively to MCS. Conversely, a score below the mean results in PF contributing negatively to PCS and positively to MCS.

Importantly then, the level of the population mean vis-a-vis individual subscale scores is decisive in determining how a particular subscale contributes to the two components. Moreover, since the range of scores below the mean for all subscales is considerably larger than the range above the mean, subscales have a greater potential negative contribution to their primary component and greater potential positive contribution to their secondary component than vice versa. In fact, the greatest potential contribution to above mean PCS and MCS scores (>50) comes from negatively weighted subscales. As shown in Figure 2, the combined potential positive effects of negatively weighted mental health subscales (MH, role emotional [RE], social functioning [SF], VT = 0) outweigh the combined potential positive effects of physical health subscales in determining PCS scores. The net effect is thus an inflated PCS score, where poor mental health actually increases physical health beyond optimal levels. Similarly, MCS scores are inflated due to poor physical health.

Furthermore, the potential contribution of each subscale to PCS/MCS scores varies between subscales, irrespective of their weights, due to the variation between subscales in the level of their population means. Given that the range of all subscale scores is 0–100, the range of possible deviations below the mean varies between subscales from a high of 84.52 (PF) to a low of 61.05 (VT) and, above the mean, from 38.95 (VT) to 15.48 (PF). This implies that certain subscales with smaller weights have potentially greater impact on PCS/MCS scores than other subscales with larger weights. For example, both BP and GH have greater potential positive impacts on PCS than PF despite lower weights and VT has the second largest potential positive impact on MCS despite it having the lowest positive weight (Figure 2).

Range of expected vs. computable component scores. As pointed out earlier, PCS and MCS values exceed the range of possible scores computable from best versus worst possible health status as indicated by SF-36 profile scores. The reason for this may be illustrated in Figure 2. Summing across optimal subscale values (score = 100) in the figure yields a PCS score of 57.87 and an MCS score of 62.14. Similarly, values corresponding to poorest profile scores sum to

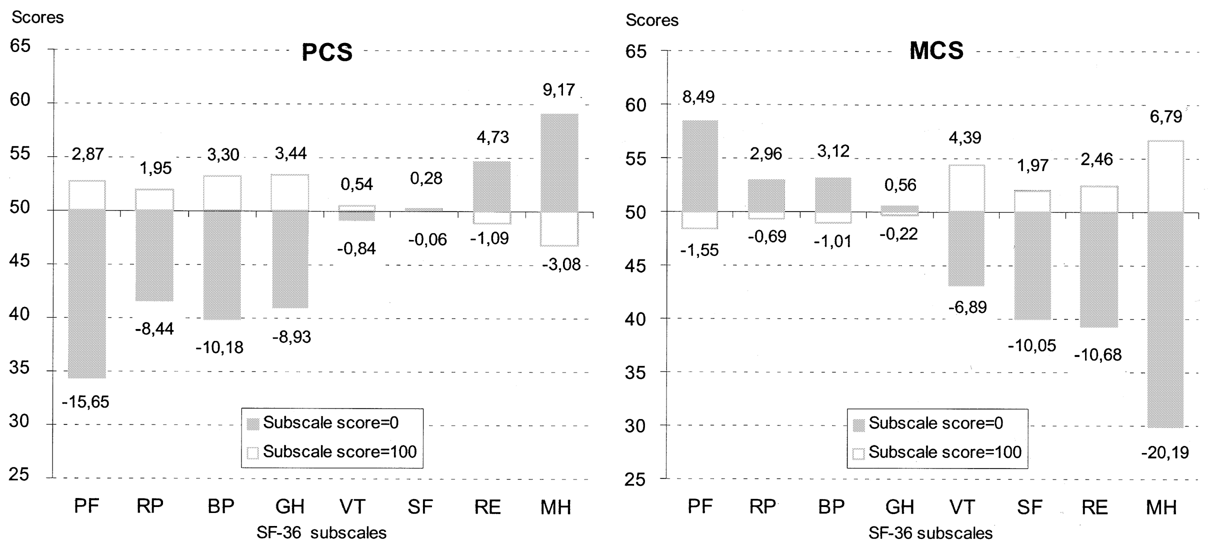

Figure 2. Maximum negative and positive potential contributions of each subscale to PCS and MCS scores. Low subscale scores (= 0) have greater potential impacts on component scores than do high scores (= 100). The combined contribution of low negatively weighted subscale scores on PCS scores is greater than that of high positively weighted subscale scores. Score contribution = [(subscale score − subscale US population mean)/subscale US STD] * subscale scoring coefficient.

20.14 for PCS and 17.32 for MCS. These scores represent the expected upper and lower boundaries for physical and mental health based on profile scores. However, due to the contributions of negatively weighted subscales to each component, the range of PCS scores is extended by 18.41 units (summing the contributions of mental health, RE and SF) both above and below this range, in effect nearly doubling the range of PCS scores (37.73 vs. 74.55). Similarly, the range of MCS values increases by 18.59 (summing the contributions of PF, RP, BP and GH) both above and below the range expected from profiles. Thus, values in the extended range of PCS and MCS scores are a product of negatively weighted subscales. In other words, PCS scores in the upper range do not reflect better ratings on physical subscales but rather impaired mental health. And vice versa, lower range scores indicate better than average mental health rather than poorer physical health. Correspondingly, upper range MCS scores reflect impaired physical health (not better mental health), while lower range scores reflect better than average physical health. In effect, the PCS and MCS scoring procedure may be likened to a seesaw, where below average physical health subscale scores weigh up mental health, while above average scores weigh down mental health. Likewise, below average mental health subscale scores increase PCS scores, while above average scores decrease PCS scores (Figure 3).

How do PCS and MCS scores relate to each other?

Correlations between PCS and MCS scores. PCS and MCS are orthogonal components and thus are

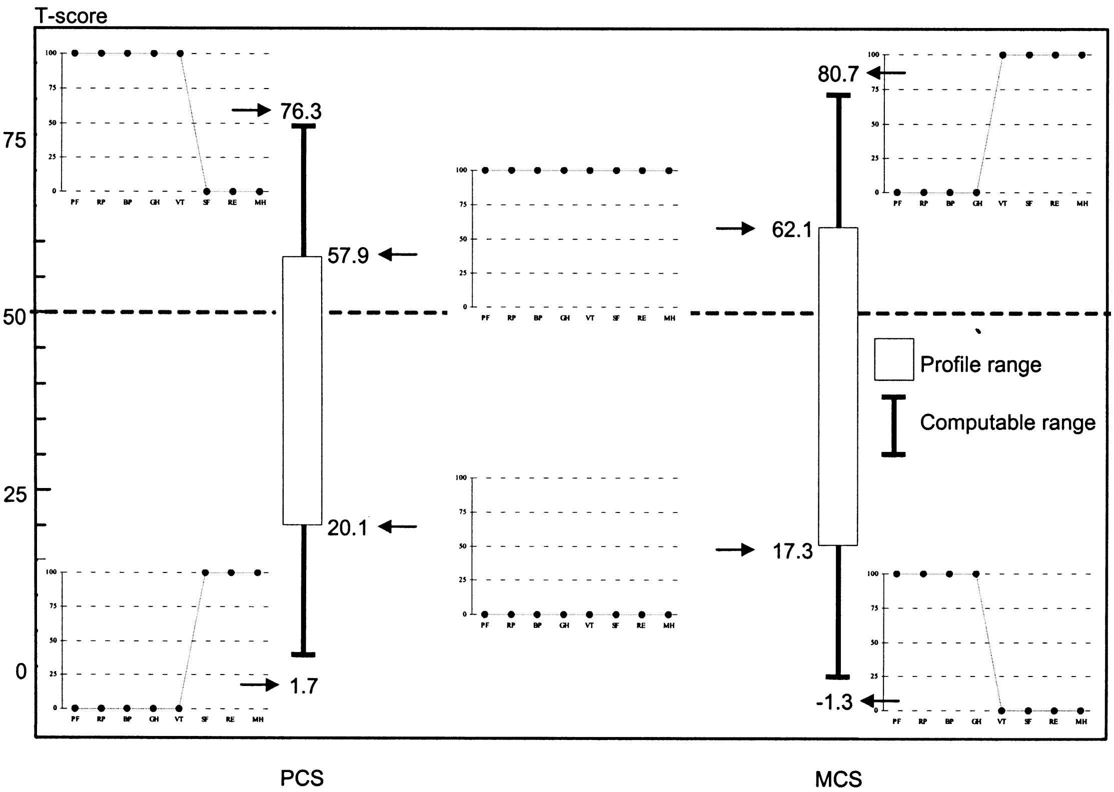

Figure 3. Differences in profile-based and true range of PSC and MCS scores exemplified by summarized profiles. When all subscale scores are at maximum values (100), PCS = 57.9 and MCS = 62.1. However, when mental health subscales are at poorest possible levels (0), the true upper range of PCS scores increases to 76.3. Likewise, the true upper range MCS score (80.7) is attainable when physical health subscales are 0. At the lower end, the range for PCS is extended an additional 18.4 points when mental health subscales are at optimal levels, and 18.6 points for MCS when physical subscales are maximized.

by definition uncorrelated. Corroborating this, cross-validations of the two summary components have yielded correlations ranging from −0.01 to 0.07 [2]. The corresponding correlation in the Swedish population survey database was 0.05. However, given the reciprocal contributions of all subscales (except VT) to the two components, it may be expected that PCS and MCS should in fact have a significant negative correlation. An explanation for the low correlations between PCS and MCS may then be that because they were computed over the entire range of scores on the two components, the reciprocal effects of subscale scores may be cancelled out. Using the Swedish SF-36 norm database (n = 8004), product-moment correlations were therefore calculated for different intervals of PCS and MCS scores: for scores above and below the means of 50, above and below profile-based score ranges and within profile-based score ranges. Table 2 shows the correlations between MCS and PCS for each of these intervals.

As shown in Table 2, correlations between PCS and MCS were highest and significant in scores above the range of expected values. Such high negative correlations are expected in this interval given the inverse effects of subscales on their secondary components. Also as expected, correlations of scores below expected values were lower than those above since the range of subscale scores above population means is smaller and thus limits the deflationary effects on secondary components. However, as is to be expected in general population surveys, sample sizes in this lower range were small and thus the correlations at this interval should be interpreted with caution. Within the expected range of scores, the correlations were significant but low. This would also be expected due to the cancellation of reciprocal effects of subscale scores above and below the mean.

What do PCS and MCS measure?

Regression analyses of PCS/MCS. The results from the correlation analyses indicate that PCS and MCS scores are most strongly interrelated outside the range of scores expected from profiles, particularly in the upper range. As shown earlier, scores in the upper range are attainable only when all positively weighted subscales reach or approach maximum values, or minimum values for scores in the lower range. With all positively weighted scales at maximum or minimum values, nearly all variance in scores outside the expected range must therefore be accounted for by negatively

Table 2. Product-moment correlations between PCS and MCS scores in the Swedish normative database (n = 8004) at different intervals of the score range

Score range	PCS score	$r_{PCS/MCS}$	MCS score	$r_{PCS/MCS}$
Full scoring range[a]	−1.8–76.2 (n = 8004)	0.05*	−6.7–79.4 (n = 8004)	0.05*
Within profile-based range[b]	17.1–57.4 (n = 6433)	0.22*	14.1–58.6 (n = 7032)	0.20*
Outside profile-based range[c]				
Above	>57.4 (n = 1540)	−0.74*	> 58.6 (n = 917)	−0.67*
Below	<17.1 (n = 31)	−0.29 ns	< 14.1 (n = 55)	−0.31*
Mean				
Above	>50.0 (n = 5363)	−0.35*	> 50.0 (n = 5291)	−0.29*
Below	<50.0 (n = 2641)	0.11*	< 50.0 (n = 2713)	0.05*

[a] Swedish scoring ranges differ from US ranges due to differences in population subscale means and standard deviations.
[b] Range of PCS and MCS scores computable from SF-36 scale profiles when all subscale scores are at maximum (100) or minimum (0) values.
[c] Range of PCS and MCS scores above and below profile-based range.
*$p < 0.05$; ns = not significant.

Table 3. Summary of stepwise regression procedure for PCS and MCS in upper scoring intervals. Data is from the Swedish SF-36 normative database (n = 8004)

Variable entered	Partial R^2	Model R^2	F	Prob > F
PCS > 57.4 (n = 1540)				
MH	0.45	0.45	1236.53	0.0001
RE	0.12	0.57	425.55	0.0001
GH	0.14	0.71	746.93	0.0001
BP	0.15	0.86	1691.03	0.0001
RP	0.07	0.93	1724.00	0.0001
PF	0.06	0.99	17703.60	0.0001
VT	0.01	0.99	16624.49	0.0001
SF				ns
MCS > 58.6 (n = 917)				
PF	0.42	0.42	670.51	0.0001
MH	0.10	0.52	180.75	0.0001
RP	0.13	0.64	320.01	0.0001
VT	0.09	0.73	291.94	0.0001
BP	0.10	0.83	495.13	0.0001
SF	0.09	0.92	1020.76	0.0001
RE	0.08	0.99	34286.83	0.0001
GH				ns

Negatively weighted subscales are in bold.

weighted subscales, i.e. MH, RE and SF for PCS and PF, RP, BP and GH for MCS. To corroborate this, stepwise regression analyses were performed for scoring intervals above the expected scoring range with the eight SF-36 subscales as independent variables and PCS and MCS as dependents (Table 3).

As shown in Table 3, the largest proportion of the variance in PCS and MCS scores at the upper intervals was explained by negatively weighted subscales. A full 57% of the variance in PCS was accounted for by negatively weighted mental health subscales, while 65% of MCS was explained by physical health subscales. Thus, in the upper range of PCS scores, this component is primarily measuring mental health. Similarly, MCS in this range of scores is mostly measuring physical functioning.

Discussion

The reciprocal effects of subscale scores on PCS and MCS component scores are strongest and most apparent in cases of extremely unbalanced profiles, either where physical health subscale scores are at optimal levels and mental health subscales are at pessimal levels or the reverse. Although extreme, such profiles are not unusual, even in general population surveys. For example, data supplied in the US manual indicated that about 14% of all PCS and 3% MCS scores were above expected ranges (PCS > 57.9, MCS > 62.1) in the US normative population [2]. Corresponding figures in the Swedish normative sample were 19% for PCS (> 57.4) and 11% for MCS (> 58.6). As is to be expected in general population surveys, few individuals scored below profile-based ranges.

Even greater proportions are likely to appear in many diseased or injured populations, which may in part account for some puzzling and sometimes counterintuitive findings reported in several clinical studies regarding spinal cord injured patients [7] and disease severity in chronic cholestatic liver disease patients [8]. A striking example is found in the US PCS/MCS manual [2], where known group discriminant validity tests were conducted between MOS patient groups with psychiatric vs. minor medical conditions. Despite better scores on all physical subscales, the group with minor medical conditions had a significantly lower ($p < 0.05$) PCS score than the group with psychiatric conditions. This was due both to the effects of above average mental health subscale scores weighing down PCS in the medical conditions group, and below average mental health subscale scores weighing up PCS in the psychiatric conditions group. In this case, conclusions based solely on PCS values would be misleading.

It is important to keep in mind, however, that all PCS and MCS scores are subject to these effects, not just those outside expected ranges. Therefore, conclusions based solely on PCS and MCS scores without comparison with profile scores may be misleading or inaccurate. For example, based on cross-sectional data from a population survey we reported that in an elderly cohort lower physical health (PCS) is offset by better mental health (MCS) compared to a younger cohort [9]. However, comparison with profile scores shows this conclusion to be unjustified. As shown in Figure 4, the group under 35 years had profile values above the population means on all eight subscales. Although their PCS score correctly reflects above average physical health, their MCS score falsely indicates poorer than average mental health. In contrast, the 65+ group had profile scores below norm values on all subscales, yet their MCS score indicates significantly better mental health than both the norm and the younger group.

Extending this example to the case of a prospective study (pre-treatment = 65+, post-treatment < 35), it is evident that a treatment primarily improving physical health may appear to negatively impact on mental health. Furthermore, physical health improvements may be overestimated due to the inverse effects of mental health subscale scores on physical health.

Summary and conclusions

Physical and mental health are described as distinct higher-order components by the developers, yet the PCS/MCS scoring algorithm involving negatively weighted subscales assumes the opposite, i.e. physical and mental health are in fact dependent. Moreover, the relationship is reciprocal and stipulates that good physical health presupposes poor mental health and vice versa. In its extreme, the scoring system dictates that achievement of maximum physical health status is contingent on worst possible mental health and vice versa. At such extremes, PCS is primarily measuring impaired mental health and MCS impaired physical health.

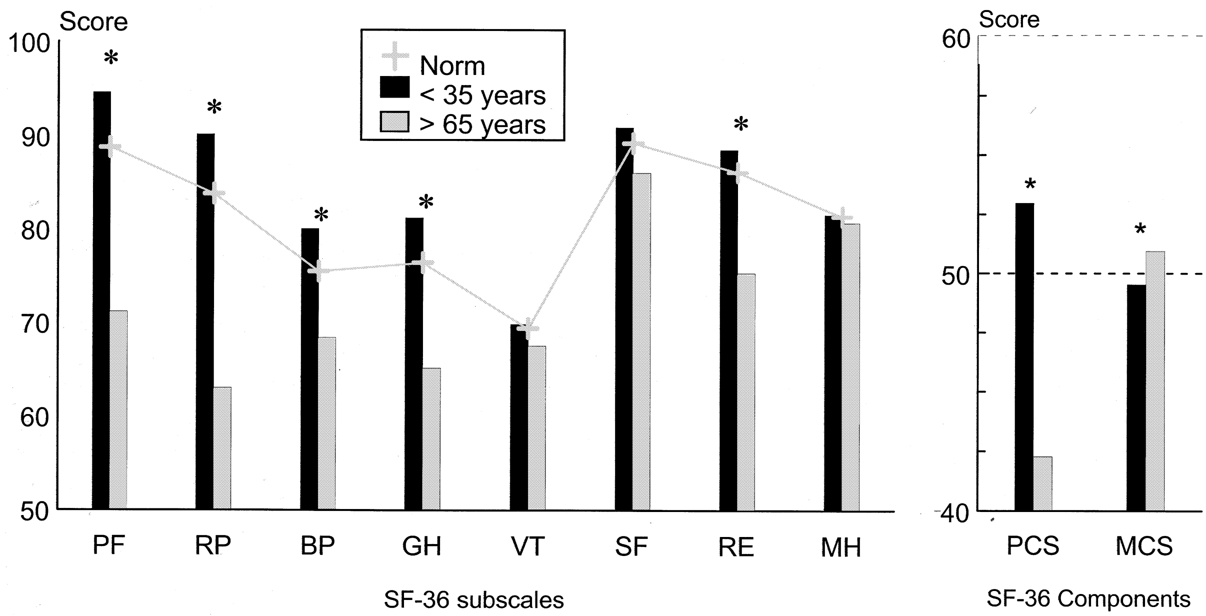

Figure 4. An example of discrepancies between SF-36 profile and component scores. The older group scored worse on all subscales than both the younger group and the normative group, yet scored better on MCS than both groups. Normative mean PCS and MCS = 50. * Significant difference ($p < 0.05$) between younger and older groups.

Not only is the scoring procedure using negative weights inconsistent with the SF-36 conceptual model, it also produces intuitively illogical results. We strongly urge that the current scoring method be revised to bring it more in line with the SF-36 conceptual model of health. Until then, we advise that PCS and MCS scores be scrupulously compared with subscale scores to gauge the potential impacts of scoring artifacts and that profiles be reported concurrently with summary scores.

The SF-36 is a valuable assessment tool with well-established reliability and validity regarding its eight basic health dimensions. Higher-order summary components are potentially a welcome complement to the eight dimensions since they offer the advantage of simplifying data analyses and interpretation. However, aggregating dimension scores to single or few global indices is problematic, and various methods have been proposed and implemented [10]. No method is without its shortcoming and whatever the method used summary scores will, by definition, involve some loss of information. It was for this reason that the SF-36 developers recommended that clinicians interpret PCS and MCS scores in conjunction with profiles when assessing an individual patient's functioning and well-being [2]. A lesser-known danger associated with summary scores is that they may incorrectly summarize the information they are meant to represent. It is important that the SF-36 component scores achieve the same standards as their underlying dimensions in order not to compromise the overall validity of this instrument.

Acknowledgements

This study was supported by grants from Vårdalstiftelsen, The Swedish Foundation for Health Sciences and Allergy Research.

References

1. Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 Health Profile and Summary Measures: Summary of results from the Medical Outcomes Study. Med Care 1995; 33: AS264.
2. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales – A User's Manual. Boston, MA: New England Medical Center, The Health Institute, 1994.
3. Hurst NP, Ruta DA, Kind P. Comparison of the MOS short form-12 (SF12) health status questionnaire with the SF36 in patients with rheumatoid arthritis. Br J Rheumatol 1998; 37: 862–869.
4. Simon GE, Revicki DA, Grothaus L, Von Korff M. SF-36 Summary Score – Are physical and mental health truly distinct? Med Care 1998; 36: 567–572.
5. Sullivan M, Karlsson J, Taft C. The Swedish SF-36 PCS and MCS – Manual and Interpretation Guide. Göteborg: Sahlgrenska University Hospital, 1997.
6. Sullivan M, Karlsson J. SF-36 Hälsoenkät: Svensk manual och tolkningsguide (Swedish Manual and Interpretation Guide). Göteborg, Sweden: Sahlgrenska University Hospital, 1994.
7. Sullivan M, Karlsson J, Taft C, Kreuter M. Interpretation of aggregated health-related quality of life scores: Examples from a clinical research program. Qual Life Res 1999; 8: 612.
8. Younossi ZM, Kiwi ML, Boparai N, Price LL, Guyatt G. Cholestatic liver diseases and health-related quality of life. Am J Gastroenterol 2000; 95: 497–502.
9. Taft C, Karlsson J, Persson L-O, Steen B, Sullivan M. Self-rated health in 70 year old men and women. Clinical relevance of profiles and summary scores of the SF-36. Qual Life Res 1997; 6: 729.
10. Hays RD, Alonso J, Coons SJ. Possibilities for summarizing health-related quality of life when using a profile instrument. In: Staquet MJ, Hays RD, Fayers PM. (eds), Quality of Life Assessment in Clinical Trials. Oxford: Oxford University Press, 1998, 143–153.

Address for correspondence: Dr Charles Taft, Health Care Research Unit, Institute of Internal Medicine, Sahlgrenska University Hospital, SE-413 45 Göteborg, Sweden
Phone: +46-31-342-2340; Fax: +46-31-82-6941
E-mail: charles.taft@medicine.gu.se

View publication stats