

Advantages of Norm-Based Scoring

The interpretation of SF-36v2™ Health Survey results has been greatly simplified with the norm-based scoring of its health domain scales and component summary measures. It is recommended that users base their interpretations on norm-based scores (Mean = 50, *SD* = 10) rather than 0–100 scores. The advantage of norm-based scoring can be illustrated by comparing the SF-36v2™ Health Survey profile scored using the original 0–100 algorithms with the profile based on the norm-based scoring algorithms for the same sample. For purposes of this comparison, the survey was scored both ways for a sample of asthmatic patients who participated in a clinical trial (Okamoto, Noonan, Boisblanc, & Kellerman, 1996).

The original 0–100 scoring produced the profile shown in Figure 7.2. The shape of this profile—the peaks and valleys due to higher and lower scores, respectively, across scales—reflect both the impact of asthma on health domains, as well as arbitrary differences in the ceilings and floors of the scales. Three scales, namely GH, VT, and MH, measure relatively wide score ranges and set the ceiling relatively high by measuring very favorable levels of those health domains (Ware et al., 1993). Other scales, such as PF and RP, assess a narrower range based on a lower ceiling. For these scales the most favorable levels (scored 100 using the original SF-36® Health Survey algorithms) represent the absence of limitations and do not extend the range into well-being. Thus, when using the original 0–100 scoring, the average score for each scale differs substantially across the profile for reasons that have nothing to do with asthma (see *Norm* in Figure 7.2). Ignoring these norms, a reasonable inference from the profile in Figure 7.2 is that asthma has a greater impact on the Vitality (VT) scale than on the Physical Functioning (PF) scale; however, this inference is incorrect.

General population norms provide a basis for meaningful comparisons across scales (see Figure 7.2). For example, the PF scale general population norm is between 80 and 90 while the VT norm is around 60 on the 100-point scale. In relation to these norms, the impact of asthma is actually much larger on the PF scale than on the VT scale, although both are statistically significant. Using the original 0–100 scoring, these differences in norms must be kept in mind when interpreting a profile. Differences in

standard deviations (which are also substantial across some scales) must also be considered for purposes of comparing results across scales.

In NBS, each scale is scored to have the same average (50) and the same standard deviation (10), meaning each point equals one-tenth of a standard deviation. Without referring to tables of norms, this method makes it is clear that whenever an individual respondent's scale score is below 45, or a group mean scale score is below 47, health status is below the average range. As shown in Figure 7.3, with norm-based scoring, differences in scale scores much more clearly reflect the impact of the disease—in this example, the impact of asthma. Using NBS, clinicians can more quickly and appropriately interpret the effect of asthma on an SF-36v2™ Health Survey profile.

Other advantages of norm-based scoring are shown in Figures 7.3 and 7.4. First, results for the PCS and MCS measures, which have always been transformed to norm-based scores, can be compared directly with results for the eight health domain scales when all are standardized on a common metric in relation to population norms. Because the PCS and MCS measures take into account the correlations among the eight health domain scales, it is clear from the example in Figure 7.3 that asthma has a very broad impact on the physical component of health.

Second, the application of norm-based scoring to a clinical trial of treatment effects is also illustrated in Figure 7.4. Patients treated using an inhaler showed statistically significant improvements (represented by the shaded portions of the bars in Figure 7.4) on the PCS measure and on the PF, RP and GH scales (i.e., three of the four scales most closely associated with physical functioning) relative to baseline after 16 weeks of treatment.

To summarize, the main advantage of NBS is simplified interpretation. When interpreting norm-based scores, one no longer has to remember the norms for eight health domain scales; the general population norm is built into the scoring algorithm. For all scales and summary measures, individual respondent scores below 45 and group mean scores below 47 can be interpreted as being below the average range for the general population. And because the standard deviations for each scale are equalized at 10, it is easier to see exactly how far

below or above the general population mean a score is in standard deviation units, and comparisons of health domain scale and component summary measure scores across the SF-36v2™ Health Survey can be made directly.

Another very important advantage of NBS is that it provides a basis for direct comparisons between scores from the SF-36® Health Survey and SF-36v2™ Health Survey. To facilitate such comparisons, the QualityMetric Health Outcomes™ Scoring Software 2.0 (Saris-Baglana et al., 2007) is available for scoring health domain scales and component summary measures using the 1998 norms (see Chapter 5). In addition to providing much more up-to-date norms for use in interpreting SF-36® Health Survey scales and measures, this software assures the comparability of results from both versions of the SF-36® Health Survey.

Finally, for those conducting research, it is important to *not* “mix” or combine NBS and 0–100 scores for the purpose of analyzing or reporting data. Mixed scores have been reported in the published literature and have resulted in erroneous conclusions about the hypotheses being tested. If a data set includes both NBS and 0–100-based scores, one can use the algorithms presented in Chapter 5 or the QualityMetric Health Outcomes™ Scoring Software 2.0 to convert all scores to a single metric (in most cases, NBS is the recommended metric). It is also important to clearly document the norms and scoring algorithms used in reports of “Study Methods” accompanying results based on the SF-36v2™ Health Survey 1998 U.S. general population norms. Further, because tables and figures are sometimes distributed separately, it is also important to include explicit references to *SF-36v2™ Health Survey 1998 U.S. general population norms* and to *norm-based scoring (NBS)* in tables and figures presenting results based on the more current 1998 U.S. general population norms.