Health-related quality of life in dermatology: measurement, interpretation and application

Prinsen, C.A.C.

Citation for published version (APA):
Interpretation of Skindex-29 scores
ABSTRACT
In this Commentary, we compare two categorizations of a dermatological health–related quality-of-life (HRQoL) instrument, the Skindex-29. One was created on the basis of an anchor-based method, the other on a distribution-based method. Differences between the two classifications are discussed, emphasizing the importance of the interpretability of HRQoL measures.
One essential property of a measurement instrument is the interpretability of its data. An understanding of the meaningfulness of a result is necessary for physicians in choosing treatments; for patients in understanding their conditions and their changes over time; and for policy makers in evaluating relationships among benefits, adverse effects, and cost. In the field of health-related quality of life (HRQoL), the problem of establishing meaningfulness is especially challenging. Although we easily understand, for example, the significance of an increase or a decrease of one degree Fahrenheit or Celsius in body temperature, it is not clear how to interpret a five-point change in a HRQoL scale measurement. In this issue, Prinsen et al. report scores for mildly, moderately, and severely impaired HRQoL for the three subscales—emotions, symptoms, and functioning—of the Skindex-29. This study completes a previous study, in which the authors identified cutoff points only for severely impaired HRQoL. To identify the categories, the authors used an anchor-based method that consists of comparing measures of HRQoL with other measures or phenomena that have relevance to patients. In particular, they used a cross-sectional method based on patients’ ratings of their HRQoL.

The categorization of the Skindex-29 was previously proposed by Nijsten et al. (2009), using mixture analysis (a distribution-based method) to obtain cutoff scores. This analysis assesses whether a distribution of a variable consists of different overlapping but independent “subdistributions” and categorizes the observations in different mixture components using posterior probabilities.

The categories obtained with the two approaches differ in several respects. In Table 1 we compare the categorization of data from studies performed by our group (and not previously used in the mixture analysis) using both sets of cutoffs. The first of each pair of columns, labeled “Psoriasis IMPROVE,” refers to data from the IMPROVE study of in-patients with psoriasis; the second of each pair, labeled “Survey 2010,” refers to data from a survey of more than 2,500 outpatients with various dermatological conditions. Mild or very mild HRQoL impairment, although reported in separate groups by Nijsten et al. (2009), are presented together by Prinsen et al. (2011, this issue). The discrepancies in the results are quite evident when dealing with a sample of all dermatological conditions (“Survey 2010”), several of which seldom highly impair patients’ HRQoL. Thus, in the Prinsen distribution the first category includes most cases. In addition, the range of some of the Prinsen classes is very narrow, so only a small percentage of patients fall into such categories. On the other hand, some Nijsten categories include a wide range of scores (the “mild” class of the emotions scale and the “moderate” class of the functioning scale, for example).

Our aim here is not to judge the correctness of the proposed classifications but to show how different methods may lead to different results. Both categorizations are valid and built according to well-established methods, but they may not represent the final say. In fact, to focus on a single example, it is difficult to reconcile results as discrepant as those seen for the symptoms scale in a sample of patients hospitalized for psoriasis. In this example, the upper limit of the “mild” category in the Prinsen classification is well above the lower limit of the “severe” category in the Nijsten classification; therefore, different decisions would be made if these categories were used to decide a patient’s eligibility for systemic treatment.
Table 1. Percentages of a sample of 936 patients hospitalized with psoriasis \(^6\) and a sample of 2,576 consecutive outpatients with various dermatological conditions \(^7\) according to the categorization of Skindex-29 by Prinsen et al. and Nijsten et al.

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Psoriasis IMPROVE (%)</th>
<th>Survey 2010 (%)</th>
<th>Symptoms</th>
<th>Psoriasis IMPROVE (%)</th>
<th>Survey 2010 (%)</th>
<th>Functioning</th>
<th>Psoriasis IMPROVE (%)</th>
<th>Survey 2010 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutoff-Prinsen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild 0–23.9</td>
<td>16.4</td>
<td>53.3</td>
<td>Mild 0–38.9</td>
<td>31.5</td>
<td>65.8</td>
<td>Mild 0–20.9</td>
<td>31.3</td>
<td>73.2</td>
</tr>
<tr>
<td>Moderate 24–34.9</td>
<td>15.2</td>
<td>13.6</td>
<td>Moderate 39–41.9</td>
<td>4.2</td>
<td>4.0</td>
<td>Moderate 21–31.9</td>
<td>15.5</td>
<td>9.7</td>
</tr>
<tr>
<td>Severe 35–38.9</td>
<td>7.8</td>
<td>5.8</td>
<td>Severe 42–51.9</td>
<td>17.1</td>
<td>11.4</td>
<td>Severe 32–36.9</td>
<td>4.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Very severe 39+</td>
<td>60.6</td>
<td>27.3</td>
<td>Very severe 52+</td>
<td>47.2</td>
<td>18.8</td>
<td>Very severe 37+</td>
<td>48.5</td>
<td>14.5</td>
</tr>
<tr>
<td>Cutoff-Nijsten</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very little 0–5.9</td>
<td>1.8</td>
<td>19.3</td>
<td>Very little 0–3.9</td>
<td>2.2</td>
<td>25.3</td>
<td>Very little 0–3.9</td>
<td>6.5</td>
<td>37.3</td>
</tr>
<tr>
<td>Mild 6–24.9</td>
<td>14.6</td>
<td>34.0</td>
<td>Mild 4–10.9</td>
<td>4.5</td>
<td>9.7</td>
<td>Mild 4–10.9</td>
<td>11.1</td>
<td>21.5</td>
</tr>
<tr>
<td>Moderate 25–49.9</td>
<td>37.0</td>
<td>30.1</td>
<td>Moderate 11–25.9</td>
<td>10.6</td>
<td>18.4</td>
<td>Moderate 11–32.9</td>
<td>29.2</td>
<td>24.1</td>
</tr>
<tr>
<td>Severe 50+</td>
<td>46.6</td>
<td>16.6</td>
<td>Severe 26–49.9</td>
<td>29.3</td>
<td>23.5</td>
<td>Severe 33+</td>
<td>53.2</td>
<td>17.1</td>
</tr>
<tr>
<td>Extremely severe 50+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If “moderate to severe symptoms” was the criterion for systemic treatment in patients with psoriasis, 31.5% of patients hospitalized and followed in the IMPROVE study would not have been eligible according to the Prinsen categorization, compared with 6.7% according to Nijsten. Looking at a specific disease such as alopecia areata in the “Survey 2010” study, large discrepancies were observed (data not shown). For example, patients were classified with a mild HRQoL impairment in the functioning scale in 75.5% (Prinsen) versus 50.9% (Nijsten) of cases; similarly, patients were classified with a severe impact on emotions in 32.1% (Prinsen) versus 13.2% (Nijsten) of cases.

There is considerable debate over the advantages and disadvantages of anchor-based and distribution-based methods. Distribution-based approaches are based on the statistical characteristics of the sample and thus are sensitive to the homogeneity of the distribution of the sample from which they are derived. Anchor-based methods, which are based on patients’ ratings, are thought to provide the best measure of the significance of change from the individual’s perspective. However, anchor-based methods rely on global ratings, may vary on the basis of whether those anchors are collected prospectively or retrospectively, and may account for some of the variance in HRQoL scores. In addition, these methods rely heavily on the representativeness of the normative sample; this is even more true of anchor-based methods because the anchor points are often also dependent on subjective choices and sociocultural environments. In other words, how useful are such results when applied to sets of patients with a different case mix, in terms of both relative proportions of different skin conditions and clinical severity of disease within each condition?

Norman et al. (2001) conducted a simulation comparing the two methods and reported equivalent information, whereas Koloktin et al. (2002) reported comparable values for anchor-based and distribution-based methods in obesity-specific quality of life at moderate levels of impairment but markedly different values for those with severe and mild impairments. The debate is far from its conclusion. As suggested by Crosby et al. (2003), it would be useful to integrate information from the two approaches as some studies have attempted to do. In dermatology, it would be desirable to conduct analyses using both methods of categorization, provided that representative samples of the population with skin conditions are available. In any case, it is important to consider that a measurement instrument will be most useful when it is possible to interpret its results. And judgments about the usefulness of interpretive tools such as a given set of “categories of quality of life” will ultimately be based on the performance of such tools in the field, in daily clinical routines, and for different patient populations.

**CLINICAL IMPLICATIONS**

- “Categories of quality-of-life impairment” are necessary for applying quality-of-life data in clinical settings.
- Differences in methods of categorization and normative samples may yield different results.
- The usefulness of a given set of categories of quality-of-life impairment must be judged by the performance of such categories in the field.
REFERENCES


