EDITORIAL

EQ-5D: Moving from Three Levels to Five

The metric system, developed in France after the revolution, changed the world of measurement. This system arose from the need for standardization and was supported by a standard meter and kilogram that are still held in a safe in Paris. As a tool for measuring health-related quality of life in economic evaluation, the EQ-5D has in some ways almost attained a similar status as a standard. It has not become a standard because of its superior measurement qualities, but because of its easy implementation in studies, widespread use, and the recommendations of bodies such as the National Institute for Health and Care Excellence (NICE). But, although any kilogram can be referenced against the standard in Paris, this is not the case with the EQ-5D because there is no measurement standard for “utility.” As a measurement tool, it may or may not be correct in its measurement accuracy, but if every decision is based on the same standard, then at least there is consistency. Although decision-making bodies such as NICE were largely happy with the tool they were asking others to use, the EuroQol Group itself was not satisfied with it. The EuroQol Group is a research group, not a standards institute.

The problem we face is that ultimate validation is not possible because utility is not a measurable construct: it can only be inferred. The determination of an absolute standard is not possible. That does not mean that one cannot present critical research about an instrument such as the EQ-5D, and indeed there has been criticism over the last 20 years related to concerns about insensitivity, content validity, and the properties of the value sets. The perception of a lack of sensitivity probably arose because the original version had just three levels per dimension. Several concerns related to the EQ-5D-3L UK value set [1] have been noted including the bimodal or even trimodal distribution of data, certain measurement blind spots, and the large proportion of states valued as worse than dead.

The instrument’s fiercest critics are often members of the EuroQol Group itself. The EuroQol Group is a relatively large group of researchers from different academic backgrounds who devote a lot of their research time and energy to examining this instrument. As stated earlier, the EuroQol Group is not a “standards institute” keen to maintain some status quo, but is a research group with an inherent drive to continue to improve the measurement and valuation of health. As such, the group considered that the weight of criticism of EQ-5D-3L should be addressed. A decision was taken to develop what is now the 5L instrument and task forces were established to test new descriptive systems, valuation methods, and analytical methods. The descriptive system emerged first [2] followed by mapping work [3] and most recently de novo scoring weights [4,5]. The developmental work tackled many issues that have been well documented (e.g., Shah et al. [6]).

Recently, the UK valuation work has been published [5,7], and it of course received considerable scrutiny. Stated simply, NICE recognized that different values and quality-adjusted life-year estimates would result from transitioning to the 5L. Keen to know what impact switching from 3L to 5L may have on their decisions, it has commissioned work by the NICE Decision Support Unit.

The first results of this work, by Hernandez et al. [8], suggest that improvements in quality of life are valued less when measured using 5L compared with 3L. This may well be related to a difference in the proportion of negative values: the UK 1997 3L value set attributed negative values to 34.6% of the states, whereas this proportion was only 5.1% in the 2017 5L value set [1,5]. Hernandez et al.’s [8] work provides new insights into the measurement performance of the 3L and 5L instruments. Fundamentally, it is difficult to determine which of the instruments is more accurate. The simple fact that there are differences is not sufficient to conclude which is the best instrument. Hernandez et al. [8] discuss some of these implications in their article here in Value in Health. Their conclusions are dependent on the validity of the assumptions used to estimate 5L and 3L scores, specifically their inventive use of Copula models to map between descriptive systems. What is needed is more data in which both the 3L and 5L are administered simultaneously.

In August 2017, NICE published a short memorandum that stated that it would continue to recommend the 3L as the basis for submissions to the agency and although it supports the use of the EQ-5D-5L questionnaire, those submitting evidence should value the data using the van Hout et al. [9] scoring algorithm, that is, using 3L values [3]. It is awaiting the results of a quality audit (required of all models used to inform government decisions in the United Kingdom) and further work to evaluate the implications of transitioning to the 5L from the 3L for decision making. As more evidence is published, NICE plans to reevaluate its position in the summer of 2018.

Work is underway by a number of groups to try to understand the difference in the UK value sets. It could well be that the “problem” is not the new 5L values, but the old 3L UK values. Moreover, this new work will not only provide insights into the merits of 3L versus 5L but also allow us to better understand the validity of the novel approach of Hernandez et al. [8] as presented here in Value in Health. In any case, we are looking at interesting times to come.

Andrew Lloyd, DPhil
Acaster Lloyd Consulting Ltd., London, UK

Andrew Lloyd works as a consultant to the pharmaceutical industry, is a coeditor of Value in Health, and sits on the Executive Committee of the EuroQol Group. All views and interpretations of events are his own.
REFERENCES


