Round, J. (2018). Once Bitten Twice Shy: Thinking Carefully Before Adopting the EQ-5D-5L. *PharmacoEconomics*, *36*(6), 641-643. https://doi.org/10.1007/s40273-018-0636-3

Once bitten twice shy: thinking carefully before adopting the EQ-5D-5L

Jeff Round[1]

1. University of Bristol, Department of Population Health Sciences, Bristol Medical School

Corresponding author
Jeff Round, PhD
Senior Lecturer in Health Economic Modelling
University of Bristol
Canynge Hall
39 Whatley Road
BS8 2PS
Jeff.Round@bristol.ac.uk
Tel: (+44) 0117 9280 7252

ORCID ID: 0000-0002-3103-6984

Funding: No specific funding was received for the preparation of this work.

Conflicts of Interest: The author has no conflicts of interest to declare.

Abstract

The EQ-5D-3L version has long been the preferred measure of health-related utility estimates for the National Institute for Health and Care Excellence (NICE). Over time, concerns about the EQ-5D-3L have been raised, focusing on its sensitivity to changes in health state, ceiling effects, and the uneven distribution of values across the states. In response, the EuroQol group developed an alternative five-level version, the EQ-5D-5L, and in 2013 NICE recommended the 5L version for use in reference case analyses. Subsequently, the new version of the instrument has been widely adopted in clinical research and for economic evaluation, typically at the expense of the EQ-5D-3L version. Yet this begs the question and assumes the 5L version is simply a minor change to the 3L version that can be adopted without consideration of its own merits. It is right to ask, however, whether it is more than just a minor variant of the 3L. The 3L version has been tested and validated in a wide range of patient groups – the same cannot yet be said for the 5L version. Additionally, the 3L version was adopted at a time when reliable alternative measures of outcome were scarce – the same cannot be said now. Here I argue that analysts and reimbursement authorities such as NICE should exercise caution and consider the whole range of alternative tools before adopting the 5L version as the reference case.

1. The EQ-5D-5L

The Euro-Qol group introduced the EQ-5D-5L descriptive system in 2009 [1], with the 1st publication describing the tool appearing in 2011 [2]. This revised version of the EQ-5D instrument introduced five response-levels, alongside changes to the wording used to describe responses in the mobility domain. The five-level version of the EQ-5D descriptive system was developed in response to perceived failings of the three response-level EQ-5D-3L, notably in its sensitivity to changes in health [3], but also ceiling effects [2] and an 'uneven' distribution of responses as measured by the valuation tariff [4].

The first UK valuation study of the 5L descriptive system was formally available in 2017 [4]. As Brazier, Bryan and Briggs recently summarize [5], the 5L version of the instrument has been found to reduce ceiling effects and exhibits a more even distribution of responses. Hernandez et al [6] have shown that as a consequence of the changes to the tool and the new valuation system, utility scores derived using the 5L version are compressed towards the full health end of the measurement scale, suggesting that while the 5L version describes more distinct health states, the values placed on them are potentially less sensitive for detecting clinically meaningful changes in health utility. Given the existing 3L and 5L valuation tariffs, cost-effectiveness analyses using different versions of the tool are likely to lead to different results and possibly different decisions on the cost-effectiveness of interventions.

2. Minor changes, major impact?

The response of the National Institute for Health and Care Excellence (NICE), the UK body responsible for health technology appraisal and reimbursement decisions, to the challenges above has been to commission further research into the impacts of changing from the 3L version of the instrument to the 5L version. This is sensible for two reasons that are widely acknowledged, and a third that is less so. First, a change from 3L to the 5L, and the consequent change in tariffs, will lead to questions being raised over the historical decisions taken by NICE – would an intervention that was cost-effective as assessed under the 3L be cost-effective if the 5L had been used? NICE are undertaking work to consider the implications of this [7].

The second reason relates to the choice of comparators in evaluations. If the 3L and 5L tariffs give different answers to empirical questions, is it appropriate to compare across them when judging whether interventions are cost-effective? Technology appraisals frequently rely on decision models that draw on a wide range of sources. Are comparisons within a technology appraisal valid if the benefits of an existing intervention have been valued from the 3L version, while benefits of a new intervention are measured using the 5L? One approach that would allow this would be to map between tools (assuming the availability of patient level data), and this is being considered as part of the review of the evidence commissioned by NICE [7]. But in many cases, it will not be possible to map, and guidance will need to be developed to inform analysts when faced with this challenge.

The third reason to be cautious about adoption of the 5L – and one that is not knowingly being explored by NICE, is that the two instruments, are, even if only subtly, different. And just as we would not expect the EQ-5D-3L and the SF-6D to lead to the same utility scores and decisions, what justification would there be to believe, *a priori*, that the 3L and 5L versions of the EQ-5D would give rise to the same weights and same decisions?

While no one suggests that the results are expected to be *identical*, there appears to have been an assumption, as evidenced by the NICE position statement, that the EQ-5D-5L was the natural successor to the 3L [7]. In fact, the 2013 NICE guidance on technology appraisal methods permits the use of the 5L version in reference case analyses [8] in the absence of a valuation set or a body of evidence on the validity of the tool. Given this, and the timing of the position statement from NICE, it seems reasonable

to believe that the EQ-5D-5L was the anointed reference-case successor, and that is it perhaps an unexpected set of results arising from the valuation study that have led to delay in coronation.

3. Validity

Is the 5L version of the EQ-5D simply a minor variation from the 3L version, or is it more than that? This question is especially pertinent when we consider the validity of the tools we use to measure outcomes in a study – the validity of the measurement tool is crucial in assessing the validity of the conclusions drawn based on it. The 3L version has been tested for suitability across a wide range of populations, including diabetes [9], dementia [10], deafness [11] and depression [12] (and that's just a sample from the Ds). The 5L version has not yet been subject to this extensive level of assessment. And while some studies have compared the measurement properties of the 5L version to the 3L version [13], if the 3L version is already not considered adequate, its use as a comparator in validity assessment seems oblique.

The 3L version has also been found wanting in populations, including for many mental health conditions [14, 15], for children [16], and for people at the end of life [17]. In response, alternative tools such as the Recovering Quality of Life instrument [18], CHU-9D [19] and the ICECAP-SCM [20] have been developed for use in these populations. If the EQ-5D-3L was considered not valid in these groups, what can we say about the EQ-5D-5L? Additional testing will be required, and it cannot necessarily be assumed that the 5L will either overcome or have the same measurement problems (though our prior beliefs may vary according to why the EQ-5D-3L was not considered suitable in any given group).

We can only answer the above questions by subjecting the 5L version to a period of assessment and validation work. Many researchers in recent years have included the 5L tool as the outcome measure for economic evaluation without this validation work having occurred – the author included [21] – on the assumption that the two instruments were effectively interchangeable. With hindsight, this assumption looks increasingly inappropriate. Ongoing studies should examine how the EQ-5D-5L performs within the study populations and report this, but whether the economic evaluation results are reliable will need to be assessed on a case-by-case basis.

It seems likely that adopting the 5L version in the absence of the required validation work could lead to the same outcome as with the 3L – an instrument adopted as a de facto gold standard in the absence of a sufficiently robust evidence base. In 2004 NICE determined that 'the most appropriate choice in the UK appears to be the EQ-5D' [22], with a firmer recommendation in 2008 [23]. But 14 years on from that first guidance there are, in 2018, a range of generic preference-based instruments measuring and health and beyond from which to choose – should the EQ-5D-5L be privileged over this set of existing instruments owing to its parentage, or should it be treated as merely the latest addition to this panoply?

4. Proceed with caution

Recognising that the 3L and 5L versions are different, even if they are closely related, leads us to an unavoidable question – is it reasonable to simply adopt the 5L version of the tool in place of the 3L version? NICE are considering the technical implications of doing so, in terms of the outcomes of decision making, and this is important, but it does assume that the replacement for the 3L should be the 5L. Yet it is reasonable to ask – if the decision is made to change the reference case quality of life instrument from the EQ-5D-3L, is the 5L version the best choice? There are a wide range of preference-based measures of quality of life tool available to the health economist, and an expanding range of alternative outcome measures based on concepts such as capabilities [24] or subjective well-being [25]. Many of these alternatives, notably the SF-6D, have a robust body of research spanning many

years and supporting their use in a wide variety of patient groups [26-31]. This is something the EQ-5D-5L currently lacks and will only gain through time. Consideration should at least be given to alternatives before adopting the EQ-5D-5L.

Perhaps it is useful to finish with a thought experiment – if the EQ-5D-3L had never existed, would anyone develop the 5L version? Would it solve a problem not solved by a competing tool? In a competitive marketplace, would adoption of the 5L tool be justified, given its already contested descriptive system and valuation study [5]? Or would a different instrument be preferred? The health economics community has been bitten before by adopting a tool it did not fully understand as the reference case for evaluation – perhaps we should be a little bit shy before adopting its successor.

References

1. EuroQol Research Foundation. *EQ-5D-5L | About*. 13th February, 2018]; Available from: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/.
2. Herdman, M., et al., *Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L).* Qual Life Res, 2011. **20**(10): p. 1727-36.
3. Lloyd, A., *EQ-5D: Moving from Three Levels to Five.* Value Health, 2018. **21**(1): p. 57-58.
4. Devlin, N.J., et al., *Valuing health-related quality of life: An EQ-5D-5L value set for England.* Health Econ, 2018. **27**(1): p. 7-22.
5. Brazier, J., A. Briggs, and S. Bryan, *EQ-5D-5L: Smaller steps but a major step change?* Health Econ, 2018. **27**(1): p. 4-6.
6. Hernandez Alava, M., et al., *EQ-5D-5L versus EQ-5D-3L: The Impact on Cost Effectiveness in the United Kingdom.* Value Health, 2018. **21**(1): p. 49-56.
7. National Institute for Health and Care Excellence, *Position statement on use of the EQ-5D-5L valuation set*. 2017, NICE: London.
8. National Institute for Health and Care Excellence, *Guide to the methods of technology appraisal*. 2013.
9. Konerding, U., et al., *The validity of the EQ-5D-3L items: an investigation with type 2 diabetes patients from six European countries.* Health Qual Life Outcomes, 2014. **12**: p. 181.
10. Jensen-Dahm, C., et al., *Discrepancy between self- and proxy-rated pain in Alzheimer's disease: results from the Danish Alzheimer Intervention Study.* J. Am. Geriatr. Soc, 2012. **60**(7): p. 1274-1278.
11. Ramakers, G.G., et al., *Agreement between health utility instruments in cochlear implantation.* Clin Otolaryngol, 2016. **41**(6): p. 737-743.
12. Dickerson, J.F., et al., *Evidence on the longitudinal construct validity of major generic and utility measures of health-related quality of life in teens with depression.* Qual Life Res, 2017.
13. Janssen, M.F., et al., *Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study.* Qual. Life Res, 2013. **22**(7): p. 1717-1727.
14. Brazier, J., et al., *A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures.* Health Technol Assess, 2014. **18**(34): p. vii-viii, xiii-xxv, 1-188.
15. Mulhern, B., et al., *Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D.* Br J Psychiatry, 2014. **205**(3): p. 236-43.
16. Noyes, J. and R.T. Edwards, *EQ-5D for the assessment of health-related quality of life and resource allocation in children: a systematic methodological review.* Value Health, 2011. **14**(8): p. 1117-29.
17. Bailey, C., et al., *'The ICECAP-SCM tells you more about what I'm going through': A think-aloud study measuring quality of life among patients receiving supportive and palliative care.* Palliat Med, 2016. **30**(7): p. 642-52.
18. Brazier, J.E., et al., *First validation of the short recovering quality of life (ReQoL) measure.* Quality of Life Research, 2016. **25**: p. 96-96.
19. Stevens, K., *Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation.* Appl. Health Econ. Health Policy, 2011. **9**(3): p. 157-169.
20. Huynh, E., et al., *Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life.* Soc Sci Med, 2017. **189**: p. 114-128.
21. Jurkovic, D. *Organisation of Early Pregnancy Units and its effects on quality of care*. 2015 13th February, 2018; Available from: https://ukctg.nihr.ac.uk/trials/trial-details/trial-details?trialNumber=ISRCTN10728897.
22. National Institute for Clinical Excellence, *Guide to the Methods of Technology Appraisal*. 2004, National Institute for Clinical Excellence.

23. National Institute for Health and Clinical Excellence, *Guide to the methods of technology appraisal*. 2008.
24. Al-Janabi, H., T.N. Flynn, and J. Coast, *Development of a self-report measure of capability wellbeing for adults: the ICECAP-A.* Qual. Life Res, 2012. **21**(1): p. 167-176.
25. Dolan, P., R. Layard, and R. Metcalfe, *Measuring Subjective Wellbeing for Public Policy: Recommendations on Measures*. 2011.
26. van Stel, H.F. and E. Buskens, *Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease.* Health Qual Life Outcomes, 2006. **4**: p. 20.
27. McCrone, P., et al., *A comparison of SF-6D and EQ-5D utility scores in a study of patients with schizophrenia.* J Ment. Health Policy Econ, 2009. **12**(1): p. 27-31.
28. Sayah, F.A., et al., *Comparative performance of the EQ-5D-5L and SF-6D index scores in adults with type 2 diabetes.* Qual Life Res, 2017. **26**(8): p. 2057-2066.
29. Patel, A.R., et al., *The validity of the SF-12 and SF-6D instruments in people living with HIV/AIDS in Kenya.* Health Qual Life Outcomes, 2017. **15**(1): p. 143.
30. Dritsaki, M., et al., *An empirical evaluation of the SF-12, SF-6D, EQ-5D and Michigan Hand Outcome Questionnaire in patients with rheumatoid arthritis of the hand.* Health Qual Life Outcomes, 2017. **15**(1): p. 20.
31. Yousefi, M., et al., *Comparison of SF-6D and EQ-5D Scores in Patients With Breast Cancer.* Iran Red Crescent Med J, 2016. **18**(5): p. e23556.