Institution: University of Sheffield		
Unit of Assessment: C-17 Business and Management Studies		
Title of case study: Improving measurement of health outcomes for NICE appraisals		
Period when the underpinning research was undertaken: 2016–2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Stephen Pudney	Professor of Econometrics	2016–present
Period when the claimed impact occurred: 2018–2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact (indicative maximum 100 words)

NICE makes important cost-effectiveness recommendations on medical and social care practices. Its work requires that health outcomes are valued consistently across conditions and treatments. Health outcomes are typically reported by patients using a simple questionnaire (EQ-5D) which has recently been updated, and NICE now requires an updated method of valuing outcomes from the new measure. The research examined the only valuation system that has so far been developed for new EQ-5D and demonstrated to NICE that (i) its adoption could lead to major changes in recommendations and (ii) that it is based on flawed data and statistical analysis. As a consequence, NICE guidance has been revised, and new research has been launched to put valuation on a sounder basis. This has changed current NICE policy and will have important long-term implications for NHS patients and spending.

2. Underpinning research (indicative maximum 500 words)

The UK National Institute for Health and Care Excellence (NICE) (and similar bodies internationally) makes recommendations on a wide range of medical technologies and practices that affect the health of millions of people. This work is largely based on clinical trials that compare health outcomes of patient groups receiving alternative treatments. Outcome measures are used to generate conclusions about the relative cost-effectiveness of those treatments. The most widely used outcome measure that can directly be used in cost-effectiveness analysis is the EQ-5D, which was originally developed by The EuroQol Group in 1990 to measure health outcomes in five domains (mobility, self-care, usual activities, pain, anxiety/depression). A complication is that EQ-5D has been subsequently updated, to give better sensitivity by measuring each domain on five levels (EQ-5D-5L) rather than the original three levels (EQ-5D-3L). These two alternative versions are both in current use.

For purposes of cost-effectiveness analysis, outcome measures are converted into utility values representing societal views of the relative seriousness of outcomes. The utility values can be used to compare benefits on a common basis (using quality-adjusted life years or QALYs), across alternative treatments and different conditions. Construction of the required utility values is done by means of hypothetical choice experiments, which elicit relative valuations from representative members of the general population. The first valuation system proposed for use of EQ-5D-5L in England was published by Devlin, et al, in *Health Economics* 2017. NICE were initially minded to adopt this system but, following presentation at the NICE Technical Forum of research underpinning publications [R1-R3], a proposal for quality assurance work was

discussed and prioritised by the Policy Research Unit Oversight Group comprising senior representatives from the Department for Health and Social Care (DHSC), NHS England and NICE.

Those discussions led to commissioning of the research described here, which is part of a sustained programme of research on outcome measurement, conducted by Steve Pudney and two colleagues in Sheffield's School for Health and Related Research (ScHARR) – Mónica Hernández Alava and Allan Wailoo. This research received funding from DHSC (under the auspices of NICE), the ESRC and MRC; and it was carried out within EEPRU (the Policy Research Unit in Economic Methods of Evaluation in Health and Social Care Interventions) established to provide evidence to help the Department of Health and Social Care, and its constituent bodies, make the best use of scarce resources. Pudney's main contribution was to the overall design of review work, empirical analysis of experimental data, technical analysis of modelling methodology and guidance on software review. The research was prompted by the team's appreciation of the long-term importance of the choice of valuation system for EQ-5D-5L and our concerns about some aspects of the proposed valuation system.

The research was aimed at improving methodology in two ways:

1) By re-examining a large number of real-world cost-effectiveness studies to demonstrate that adopting the new valuation system would have potentially important consequences for NICE decisions (R3, R4). This research made clear the need to establish whether or not the system met a reasonable quality threshold.

2) By conducting an in-depth critical evaluation of the data, statistical methodology and estimation software that had been used to construct the proposed valuation system (R5, R6). Specifically, the evaluation comprised:

    - Statistical work providing an empirical assessment of the quality of the experimental data on which the value set was based.
    - Re-estimation and technical evaluation of the econometric specification and methods, and Bayesian software implementation used to construct the value set.

3. References to the research (indicative maximum of six references)

R1. Hernández-Alava, M., & Pudney, S. (2017). Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. *Journal of Health Economics*, *55*, 139–152. https://doi.org/10.1016/j.jhealeco.2017.06.013

R2. Hernandez Alava M., Wailoo A., Pudney S. (2017). *Methods for mapping between the EQ 5D 5L and the 3L*. NICE Decision Support Unit. http://nicedsu.org.uk/wp-content/uploads/2020/06/Mapping-5L-to-3L-DSU-report.pdf.

R3. Hernandez Alava, M., Wailoo, A., Grimm, S., Pudney, S., Gomes, M., Sadique, Z., Meads, D., O'Dwyer, J., Barton, G., & Irvine, L. (2018). EQ-5D-5L versus EQ-5D-3L: The Impact on Cost Effectiveness in the United Kingdom. *Value in Health*, *21*(1), 49–56. https://doi.org/10.1016/j.jval.2017.09.004. (Awarded *Value in Health* best paper of 2019).

R4. Pennington, B., Hernandez-Alava, M., Pudney, S., & Wailoo, A. (2018). The Impact of Moving from EQ-5D-3L to -5L in NICE Technology Appraisals. *PharmacoEconomics*, *37*(1), 75–84. https://doi.org/10.1007/s40273-018-0701-y

R5. Hernández-Aláva, M., Pudney, S.E. and Wailoo, A. (2018). *Quality Review of a Proposed EQ-5D-5L Value Set for England*. EEPRU. http://www.eepru.org.uk/wp-content/uploads/2017/11/eepru-report-eq-5d-5l-27-11-18-final.pdf

R6. Hernandez Alava, M., Pudney, S., & Wailoo, A. (2020). The EQ-5D-5L Value Set for England: Findings of a Quality Assurance Program. *Value in Health*, *23*(5), 642–648. https://doi.org/10.1016/j.jval.2019.10.017

4. Details of the impact (indicative maximum 750 words)

The impact was produced by two complementary strands of research:

1) Examination of the impact of the proposed value set on existing cost-effectiveness studies (R3, R4) were used to inform NICE of the possible difficulties that might be encountered in using the proposed EQ-5D-5L value set uncritically. This led to the commissioning of an in-depth review. Novel methods that enable this work to be conducted are reported in R2 and have subsequently been adopted by NICE in their new updated Methods Guide (S1).

2) Review of the proposed EQ-5D-5L value set. The review reached three main conclusions:

    i. There were serious concerns about the quality of data yielded by the choice experiments underpinning the construction of the value set. In particular: the experimental design covered a low proportion of the health states distinguishable by the EQ-5D-5L instrument; many participants gave responses suggesting a lack of engagement with or understanding of the experiments.

    ii. There were flaws in the specification of the econometric choice model and the method used to draw inferences about relative values of health states.

    iii. The Bayesian estimation procedures were not adequately implemented and did not achieve the required stability properties.

Given these serious concerns about data and methods, our report to DHSC and NICE recommended that “A new programme of further development, including a new data collection initiative, should be considered to put EQ-5D-5L on a sufficiently firm evidential basis” (R5, R6).

The full report (R5) was put out for independent review by NICE to four international reviewers, all of whom accepted the recommendations and one of whom (Prof. Charles F. Manski, Northwestern University and Fellow of the Econometric Society and American Statistical Association) stated: “*The* [...] *recommendations* [...] *to NICE and DHSC are exemplary*”.

NICE updated its position statement on EQ-5D-5L in November 2018 and October 2019 to reflect the team’s recommendation that the proposed valuation system was not of acceptable quality (S2, S3).

A call for expressions of interest was issued for a new data collection and valuation exercise, to be carried out in conjunction with EuroQol (S4). To avoid any perception of conflict of interest,

Pudney, Hernández Alava and Wailoo did not bid for that work. The contract was awarded in June 2020 and the new valuation exercise called for in the report is currently underway.

This decision by NICE, based entirely on the work of Pudney, Hernández Alava and Wailoo, directly affects many of the cost-effectiveness decisions currently made by NICE and will for the foreseeable future (S5, S6). It is impossible to predict with any confidence the number of cost-effectiveness decisions that will be affected by our work in the future. However, between Oct 2019 (when NICE's updated EQ-5D guidance was published) and July 2020, 29 pieces of Technology Appraisal (TA) Guidance were issued to the NHS, and 27 used EQ-5D. Many of these decisions covered new cancer drugs. For example, Palbociclib (TA619) is a treatment option for around 3,300 people per year with advanced breast cancer in England and costs £2,950 for 21 days treatment. Other decisions concerned treatment for very common conditions. Around 90,100 people with type 1 diabetes are eligible for treatment with dapagliflozin (TA597), which costs around £500 per year per patient, and NICE estimates a budget impact of £3.5m per year with 10% uptake. It is still more difficult to predict the long-term impact on NHS patients or the NHS drugs bill (respectively 16.2m referrals to treatment and £19.8bn at list prices in 2018/19 for England). But this work has already made a significant change to NICE methods guidance (S2, S3) and will in due course influence NICE evaluations further by ensuring that health outcomes valuations based on sound statistical evidence and methods will be adopted. Over the long term, this will have large cumulative impacts on choice of drugs and other medical technologies and consequently on human experience of treatment and value-for-money in NHS spending.

NICE methods in relation to the use of EQ-5D go beyond the highest profile guidance NICE issues on new pharmaceuticals. They also cover guidance on diagnostics and on clinical and public health topics. They are seen as the gold standard for non-NICE assessments such as those undertaken by the Department of Health and Social Care; its arms' length bodies such as those issuing guidance for vaccinations and screening; and the portfolio of research, including clinical trials, funded via the National Institute for Health Research (NIHR). NICE's methods, including the use of the UK EQ-5D value set, are also often replicated by countries around the world, meaning that the current change in NICE policy brought about by Pudney's work will also have implications for health care decision making around the world.

5. Sources to corroborate the impact (indicative maximum of 10 references)

S1. 2020 NICE Methods Consultation: The NICE methods of health technology evaluation: the case for change (www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/nice-guidance/chte-methods-consultation/NICE-methods-of-health-technology-evaluation-case-for-change.docx) Section 28, p40.

S2. Nov 2018 update of NICE position statement on the use of EQ-5D-5L, reflecting concerns raised by our research (paragraphs 4 and 6 refer to the review R5).

S3. Further update in Oct 2019 to reflect our recommendation that the proposed 5L value set not be used (https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/EQ-5D-5l), paragraph 3 refers to the review R5.

S4. Call for expression of interest: valuation of EQ-5D-5L in the UK (12 Feb 2020) (https://euroqol.org/blog/call-for-expression-of-interest/).

S5. Testimonial from Senior Scientific Advisor at NICE confirming the importance and influence of the research in NICE's decision to commission a new valuation exercise rather than accepting the value set that had been proposed.

S6. Testimonial from Deputy Director, Medicines and Pharmacy Analysis, Department of Health and Social Care.