Author/s:
Palmer, AJ;Campbell, JA;de Graaff, B;Devlin, N;Ahmad, H;Clarke, PM;Chen, M;Si, L

Title:
Population norms for quality adjusted life years for the United States of America, China, the United Kingdom and Australia

Date:
2021-08-01

Citation:
Palmer, A. J., Campbell, J. A., de Graaff, B., Devlin, N., Ahmad, H., Clarke, P. M., Chen, M. & Si, L. (2021). Population norms for quality adjusted life years for the United States of America, China, the United Kingdom and Australia. Health Economics United Kingdom, 30 (8), pp.1950-1977. https://doi.org/10.1002/hec.4281.

Persistent Link:
https://hdl.handle.net/11343/302969

Population Norms for Quality Adjusted Life Years for the United States of America, China, the United Kingdom and Australia.

Running title: QALY population norms

Andrew J. Palmer[1,2], Julie A. Campbell[1], Barbara de Graaff[1], Nancy Devlin[2], Hasnat Ahmad[1], Philip M Clarke[2], Mingsheng Chen[3,4], Lei Si[1,5]

[1] Menzies Institute for Medical Research, The University of Tasmania, Hobart, Tasmania 7000, Australia

[2] Centre for Health Policy, School of Population and Global Health, The University of Melbourne, Victoria, 3035, Australia

[3] School of Health Policy & Management, Nanjing Medical University, Nanjing, China

[4] Creative Health Policy Research Group, Nanjing Medical University, Nanjing, China

[5] The George Institute for Global Health, UNSW Sydney, Kensington, Australia

Address for correspondence:

Professor Andrew J Palmer

Menzies Institute for Medical Research, University of Tasmania

17 Liverpool Street, Hobart, Tasmania, Australia 7000.

Andrew.Palmer@utas.edu.au

Funding: This study was funded by the Centre of Excellence in Population Ageing Research, Australian Research Council (CE170100005) and the National Health and Medical Research Council (NHMRC) (Grant number: APP1127827). L.S. is supported by a NHMRC Early Career Fellowship (Grant number: GNT1139826). The funding body was not involved in the design of the study, or data collection, analysis, and interpretation or in writing the manuscript.

Conflict of interest: The authors declare no conflict of interest.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hec.4281.

Population Norms for Quality Adjusted Life Years for the United States of America, China, the United Kingdom and Australia.

Summary

Health economics uses quality adjusted life years (QALYs) to help healthcare decision makers. However, unlike life expectancy for which age- and sex-dependent national life tables are available, no general population norms exist to use as a benchmark against which to compare observed or modelled projections of QALYs in sub-populations or patients. We developed a 2-state Markov model to generate QALY population norms for the USA, UK, China and Australia. Annual age- and sex-specific probabilities of all-cause mortality were taken from life tables combined with general population country-specific age- and sex-specific health state utilities for the EQ-5D-3L (all countries); and SF-6D (Australia) multi-attribute utility instruments (MAUI). To validate our QALY benchmark model we found that the model closely predicted population life expectancies. Using EQ-5D-3L, undiscounted QALYs for males/females aged 18 years ranged 54.62/58.90 (USA), 55.55/60.21 (China), 57.11/60.16 (Australia), and 58.01/61.43 (UK) years. SF-6D benchmark QALYs for Australia were consistently lower than those generated from the EQ-5D-3L. The gap in undiscounted QALYs between the UK (highest) and the USA (lowest) was 2.53 QALYs in women and 3.39 QALYs in men aged 18 years. Our model's QALY population norms can be used for internal validation of future health economic models for the country-specific value sets for the instruments that we adopted, and when quantifying burden of disease in terms of QALYs lost due to illness compared to the general population. We have created a publicly available repository to continuously include QALY benchmarks that use country-specific value sets for other MAUIs and life expectancies.

1 INTRODUCTION

1.1 Health Economics and Resource Allocation

Confronting choices about how to spend society's limited health care resources will never be easy (Neumann and Sanders 2017, Lancsar, Gu et al. 2020). Economic evaluation of health care using methods such as cost-effectiveness analyses, investigates competing alternatives for scarce healthcare resources (Drummond, Sculpher et al. 2015). The unprecedented global COVID-19 pandemic has tested the allocation of scarce healthcare resources to unparalleled limits – we have observed the critical and rapid mobilisation of these resources to deal with multiple, complex and interdependent healthcare challenges ranging from caring for the gravely ill to suppressing the spread of COVID-19 within the broader community (McAnulty and Ward 2020). Nevertheless, in the aftermath of the pandemic, it is likely that the allocation of healthcare resources will be increasingly challenging against the backdrop of a global economic recession (Chilton, Nielsen et al. 2020, OECD 2020). Therefore, robust health economic modelling, including the derivation of quality-adjusted life years (QALYs) in cost-utility analysis (CUA) to help healthcare decision makers, will be even more relevant in the post-pandemic environment.

1.2 Benchmarking for QALYs

QALYs are a metric of health economic evaluation outcome commonly used in CUAs (a form of cost-effectiveness analyses), particularly in ever-increasing health technology assessments worldwide. The reliability of the health economic evaluation is often tested by the validation process. For example, the life expectancies generated from the health economic model are commonly compared to the population-specific life tables (Si, Winzenberg et al. 2015). While QALYs are among the important outcomes in CUA, no population norms for health economics QALYs exist to provide a benchmark against which the QALYs of population sub-groups, or patients with particular diseases, can be compared.

More specifically, QALYs (as a composite measure included within the broader category of health-adjusted life years [HALYs]) are frequently used to summarise the quality and quantity of life associated with a condition/disease or intervention (Drummond, Sculpher et al. 2015). QALYs combine health state utilities (HSUs; that assess the strength of preference for an individual's health state relative to perfect health=1 and death=0) with life expectancy, providing a common measure that can be used to compare health outcomes. HSUs used to calculate QALYs can be derived from a range of 'direct' and 'indirect' patient-reported methods. Direct measurement seeks a patient-reported outcome 'directly' by asking the person to value their own health under conditions of uncertainty by using a valuation technique such as time trade-off or standard gamble methodologies (Torrance 1987). Indirect measurement seeks a patient-reported response to a set of pre-determined questions on a multi-attribute utility instrument's (MAUI) questionnaire. The patient-reported responses to this MAUI-specific questionnaire are then attributed a utility value through a pre-determined set of utilities (country-specific value sets) for the particular MAUI (Ara and Brazier 2014). The value sets that are derived for the patient-reported responses to the MAUI questionnaires reflect the value systems of the country population for which they are derived and this will differ from country to country (Clemens, Begum et al. 2014). Additionally, country (or even region) specific population norms (including age- and sex-specific) also reflect the value systems of the population for which they are derived (Kularatna, Byrnes et al. 2017).

Importantly, QALYs are commonly used in decisions about health care technologies – for example, the UK's National Institute for Health and Care Excellence and Australia's Pharmaceutical and Medical Benefits Advisory Committees when considering reimbursement of new healthcare interventions (NICE 2018). Use of cost effectiveness, based on QALYs, continues to grow as new countries introduce Health Technology Appraisal (e.g. India

(Downey, Mehndiratta et al. 2017, Prinja, Downey et al. 2018) and China (MacQuilkan, Baker et al. 2018)).

Despite QALYs having been used in economic evaluation for many decades, no population benchmark QALY data have been published to date, so comparisons of QALYs or QALY loss between a study cohort and the general population of any given country, or between countries, are not possible. This is an important gap.

1.3 Benchmarking Population Health Summary Measures – State of Play

Additionally, several population health summary measures have been developed to evaluate the health status of populations (Gold, Stevenson et al. 2002, Flanagan, Boswell-Purdy et al. 2005). These summary measures are used in both the health policy context, and as metrics of health outcomes in economic evaluation (Drummond, Sculpher et al. 2015). Reflecting both the context for the development of these measures and the inherent challenges of using a single measure to summarise the health of a population, two broad categories have emerged: life expectancy/mortality and HALYs – a broad class of measures (including QALYs) that combine mortality with a measure of morbidity, disability or health-related quality of life (Gold, Stevenson et al. 2002).

For life expectancy, a range of population summary measures are available for many countries, including life tables providing age- and sex-specific life expectancy, infant mortality, all-cause and disease-specific mortality (Gold, Stevenson et al. 2002). These measures allow comparisons between countries and are also used as a benchmark to evaluate the impact of disease and medical interventions (Gold, Stevenson et al. 2002). A key limitation of these measures, however, is that the impact of problems with function, disability, pain, distress or quality of life on health is not captured (Field and Gold 1998).

HALY is an umbrella term for a range of measures that combine mortality with measures that reflect the impact of disease and/or disability. As well as the QALY, another commonly used HALY is the disability-adjusted life year (DALY) that was developed for the Global Burden of Disease Study in the early 1990s to quantify disease and disability within a given population (Murray 1994, Murray and Lopez 1994). DALYs aim to quantify the number of years of life lost due to poor health, compared to an ideal health situation free of disease and disability (Culyer 2014, Kassebaum, Arora et al. 2016). Another measure is the disability free life expectancy (DFLE) measure that was developed to assess whether gains in life expectancy achieved in recent decades have resulted in time spent in 'good or bad health' (Jagger, McKee et al. 2013), with a healthy condition represented by the absence of disability or limitations in the functioning of individuals (Eurostat 2018). This measure can apply a dichotomous definition of disability (Sullivan 1971, Imai and Soneji 2007), or alternately, grade disability using categories of long-term limitations on activity. This latter approach provides a measure referred to as Healthy Life Years (HLY) (Wolfson 1996), and is used as a key health indicator across the European Union (Jagger, McKee et al. 2013).

HALYs can be calculated using one of the prevalence or incidence methods. The most commonly used method for the prevalence approach is Sullivan's, which combines mortality data from life tables with a measure of disability, morbidity or health related quality of life (Sullivan 1971). For example, a QALY can be calculated by multiplying the quality of life HSU by life expectancy. Alternatively, incidence-based methods use multi-state life tables and data on transitional probabilities between health states. The life course of cohorts can then be estimated using a microsimulation model, which incorporates age, sex, transitions between risk factor classes and changes between disease states over time. Examples of this approach include the Population Health Model (POHEM) developed by Statistics Canada

(Hennessy, Flanagan et al. 2015) and the Netherlands Chronic Disease Model (Hoogenveen, van Baal et al. 2010).

Importantly, the relevance of the above background for our study is that for the suite of measures described above, with the exception of QALYs, *population benchmarks are available.* For example, the Global Burden of Disease groups have published DALY data that allow for comparisons between and within countries, for populations and specific diseases (Institute for Health Metrics and Evaluation 2013, G. B. D. DALYs and Hale Collaborators 2017). Such population benchmark data can also be used to demonstrate the impact of diseases, such as multiple sclerosis, in comparison to the general population.

1.4 Aims of our study

To support the drive for open source health economics (Dunlop, Mason et al. 2017), we aimed to produce the first validated set of age- and sex-specific QALY population norms/benchmarks for four countries. To achieve our aim we developed and validated a tool to generate multi-country, population-based QALYs using HSU data derived from two MAUIs namely the EQ-5D-3L MAUI (a globally prevalent MAUI used in over 60% of health economic evaluations (Richardson, Iezzi et al. 2016) for the United States of America (USA), the United Kingdom (UK), China and Australia. For sensitivity analyses we also used the SF-6D MAUI for Australia only to reveal differences in benchmark QALYs, with the addition of further country-, age-, sex- and instrument-specific QALYs informed by different country-specific MAUI value sets planned in the future.

As a parallel aim, we will foster a community of interest through the sharing of QALY benchmarks through an integrated online open source repository.

2 METHODS

2.1 Model structure and parameters

2.1.1 Markov model

A probabilistic Markov model was developed with two health states ('alive' and 'dead') incorporated in the benchmark QALY model (Figure 1). Simulated patients transited from 'alive' to 'dead' with a 1-year cycle length based on annual mortality risks by age, sex and population, otherwise they stayed in the 'alive' state with updated age and HSU for the next simulation cycle. When a patient transited in the 'dead' state, a cumulative QALY was calculated.

2.1.2 Mortality risk

Four countries were included in this initial version of the model, namely the USA, China, UK, and Australia. Age- and sex-specific mortality rates for the general population were obtained from the Centers for Disease Control and Prevention (Xu, Murphy et al. 2016), the National Bureau of Statistics of China (China 2010), Office for National Statistics of the UK (Statistics.), and Australian Bureau of Statistics (Statistics 2015), respectively. The annual mortality rates were converted to annual risks using the following formula:

$$p = 1 - \exp(-r)$$

where p is the mortality risk, r is the annual mortality rate by age, sex and population.

2.1.3 EQ-5D-3L and SF-6D MAUIs

The EQ-5D-3L is a globally prevalent MAUI that assesses 243 health states derived from the 5 question and 3 level questionnaire (Dolan 1997). EQ-5D-3L value sets are available for many developed and developing countries and the instrument's algorithm assesses health states considered worse than death (less than zero) for some countries (Clemens, Begum et al. 2014, Devlin and Brooks 2017). We adopted the 3L for the inaugural set of benchmark

QALYs given the prevalence of this 'legacy instrument' and the existence of value sets for all four countries for the 3L.

For sensitivity analyses, we adopted the SF-6D for Australia only to test for differences in benchmark QALYs for this preferentially sensitive instrument for complex and chronic disease and to show the ability of including other MAUI HSUs in the model (Brazier, Roberts et al. 2004, Kularatna, Byrnes et al. 2017, Campbell, Jelinek et al. 2020). The SF-6D's HSUs are derived from patient-reported responses to the SF-36 (11 questions) or SF-12 questionnaire and the SF-6D version 1 algorithm assesses HSUs for 18,000 separate health states (Brazier, Roberts et al. 2002). The SF-6D's algorithmic range for the Australian value set (-0.363 – 1.0) is similar to the EQ-5D-3L (-0.217 – 1.0 model 3) (Viney, Norman et al. 2011, Norman, Viney et al. 2014).

Each type of MAUI may provide different utility values, even when used in the same study population including the general population norms for a country (Campbell, Palmer et al. 2016). The overall differences in the most commonly used MAUI's HSUs for the same country is predominantly attributable to their differing descriptive systems whereby the absolute average difference between the most common instruments is 0.135 utility points (Richardson, Iezzi et al. 2015). Additionally, the same instrument is likely to value different HSUs for different countries (Augustovski, Irazola et al. 2009, Viney, Norman et al. 2011, Clemens, Begum et al. 2014).

Therefore, to generate the first example of benchmark QALYs, the EQ-5D-3L MAUI's HSUs for the general population in the USA, China, UK and Australia, and the SF-6D MAUI's HSUs for Australia only were obtained from published literature (Dolan 1997, Kind, Hardman et al. 1999, Shaw, Johnson et al. 2005, Sullivan and Ghushchyan 2006, Viney, Norman et al. 2011, Clemens, Begum et al. 2014, Norman, Viney et al. 2014, Si, Shi et al.

2017). HSU data is generally not distributed normally, therefore, beta distributions were used to reflect the left skewed nature of HSUs in the general population (Gray, Clarke et al. 2010).

2.2 Validation of the benchmark QALY model

Face validation and verification of life expectancy generated from the model versus published life tables were conducted to ensure model validity. Face validity of the model refers to whether model structure, data sources, problem formulation and results are meaningful (Eddy, Hollingworth et al.). In addition, we have generated age- and sex-specific life expectancies from the model by disabling the HSUs (i.e. only accounted for years of living) component and comparing them to published life expectancies that were generated from other statistical methods. Goodness-of-fit analysis was conducted, life expectancies calculated by the model were regressed against published data and the slope of the regression line was provided to gauge the accuracy of model predictions compared to published data. In addition, R-squared of the regression line and the root mean square error (MSE) were reported to indicate how well the regression model performed to measure in the difference between the model predictions and that from the literature.

2.3 Prediction of QALYs

TreeAge Pro for Health Care 2017 version (Williamstown, Massachusetts, USA) was employed to develop the benchmark QALY model and to run simulations. Mean, 2.5^{th} and 97.5^{th} percentiles of QALYs by country were generated by second order Monte-Carlo simulation resampled 100,000 times. Abridged residual lifetime QALYs by country were generated from Monte-Carlo simulation with 5-year intervals and were then fitted into polynomial models to generate age-specific QALYs. Goodness-of-fit analysis was conducted with R-squared for each polynomial model to ensure its accuracy of predicting age-specific QALYs. The polynomial model is expressed as:

$$y=\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3$$

As discounting of QALYs is often required in cost-utility analyses in many jurisdictions (Severens and Milne 2004), we have also provided benchmark QALYs with annual discount rates of 3% and 5% in addition to non-discounted values.

3 RESULTS

3.1 Model validation and Benchmark QALYs

QALY calculation involves age- and sex- specific mortality and HSUs in that given time. To ensure that our model can accurately predict QALYs, we needed to ensure that our model can produce reliable life expectancies. Therefore, we conducted the life expectancy verification by comparing the life expectancy that were generated from our model against that from published life tables in each country - a total of 70 sets of life expectancy comparisons were conducted to ensure that our model could accurately predict QALYs. There were a total of 104 pairs of life expectancies generated from the model and that from the literature compared as the life expectancy verification process. Ten pairs for the Australian population; 34 pairs for the Chinese population; 32 pairs for the UK population and 28 pairs from the US population (Appendix 1 – tables 1, 2, 3 and 4 respectively). All comparisons are graphed using a linear approach (Figure 2). The slope of the regression line for life expectancies from published literature and modelled predictions was 0.998. In addition, the R-squared and the root MSE of the linear regression model was 0.99 and 0.32 respectively, indicating a good reproducibility of life expectancy from the benchmark QALY model (Figure 2).

Age- and sex-specific QALYs by country generated from the model are shown in Table 1. In addition to undiscounted QALYs, we have also presented QALYs discounted at 3% and 5% per annum in Tables 2 and 3. Table 4 reveals the differences in QALYs when using either the EQ-5D-3L or the SF-6D for Australia.

Overall, and similar to life expectancy, women were found to have higher QALYs compared to men. Undiscounted QALYs generated by the EQ-5D-3L HSUs for males aged 18 years ranged from 54.62 years (USA), 55.55 (China), 57.11 (Australia), and 58.01 (UK). Additionally, for females aged 18 years, undiscounted QALYs were 58.9 years (USA), 60.16 (Australia), 60.21 (China), and 61.43 (UK). QALYs decreased with age in both sexes (Tables 1,2,3).

Similar trends were reflected in the undiscounted benchmark QALYs generated for the Australian value set using the SF-6D MAUI whereby females consistently recorded higher QALYs than males. Due to the expected differences in the utility valuations of the two MAUIs (Richardson, Iezzi et al. 2015) the undiscounted benchmark QALY's for Australia also recorded a mean negative difference of 3.0 QALYs for males from the EQ-5D-3L estimates to the SF-6D estimates, and for females 4.58 QALYs. These results also translated to a mean 10% difference for males and a mean 14% difference for females. (Figure 3, Table 4).

In regard to the country differences, overall the UK population was estimated to have the highest QALYs compared with the other three countries. The difference in undiscounted QALYs between the UK (highest) and the USA (lowest) was 2.53 QALYs in women and 3.39 QALYs in men aged 18 years old.

Table 5 shows the cubic polynomial model that was chosen as the QALY prediction model for continuous age. Moreover, the QALY prediction model was fitted against values that were generated from the QALY benchmark model, all prediction models had an explanatory power (R-squared) greater than 0.99 (Appendix 2). A complete benchmark continuous QALY table is provided in Appendix 3 and is uploaded online in our open source repository of all benchmark QALYs: http://menzies.utas.edu.au/qaly_benchmarks.

DISCUSSION

This study provides the first QALY benchmarks for the USA, China, UK and Australia. Using a probabilistic Markov model with annual transitions, we have generated benchmark QALYs using both the EQ-5D-3L and SF-6D MAUIs for country-specific value sets for males and females for age 18 years and from age 20 years to age 85 years with a 5-year interval. In addition, we have generated continuous annual QALYs based on the abridged QALYs using polynomial equations. To reflect requirements of decision makers, QALYs have also been reported with 3% and 5% discount rates applied. We have demonstrated the importance of generating QALYs for each instrument, and that QALYs generated from one instrument or another country-specific value set are not comparable with those generated from another.

Importantly, our QALY benchmarks have addressed a pressing information gap in the health economics and policy-making literature by providing much needed population benchmarks that will enable health economists to further internally validate their cost-utility models used for the economic evaluation of health interventions. Ultimately, robust decisions regarding the allocation of scarce healthcare resources from one alternative to another can be made with additional confidence if the health economics cost-utility model's output metrics (including QALYs) used to inform decision makers can be further demonstrated to be valid with our QALY benchmarks. Another example of the usefulness of our QALY benchmarks relates to a recent paper that explored the societal concerns about the equitable distribution of resources within a model that adjusts for severity probabilities (Versteegh, Ramos et al. 2019). The pretext of this paper was that an often overlooked issue was the uncertainty and heterogeneity in the estimate of severity. We suggest that the findings of that paper are synergistic with our work whereby the outcomes of severity probabilities could also be validated within our QALY benchmarks frameworks to make the necessary comparisons.

These data can also be used to demonstrate the impact of different health states on QALYs in comparison to the general population. An example of this is provided by a recent Australian study (43) on the impact of multiple sclerosis (MS). This Australian study reported that life expectancy for 35-year old females with MS was 43.3 years (42.2-44.2); 6.9 years less than that for a 35-year old Australian woman from the general population. QALYs for 35-year old females were estimated to be 28.7 years (26.7-30.4), 12.6 years less than the age and sex-specific population norm (Ahmad, Taylor et al. 2017). Another example is the lifetime projection of life expectancy and QALYs for type 2 diabetes patients in the UK (Clarke, Gray et al. 2004), in which the UKPDS group reported life expectancies for males with type 2 diabetes aged 45-50 of 25.54 years versus 30.05 years in the general population, a difference of approximately 5 years. The UKPDS were unable to compare their QALY estimates to any UK population norms. However, we can now compare their reported projections of 18.82 undiscounted QALYs with our estimate of 28.15 QALYs for the general UK population, a difference of 9.33 QALYs. Use of QALY benchmarks as comparators in these studies allowed for a more nuanced description of the impact of MS and diabetes on individuals living with these diseases.

At present, the QALY benchmarks have been calculated using utility data generated from the EQ-5D-3L for the USA, China and the UK. Importantly, to further demonstrate the capacity of our model of incorporating HSUs of different MAUIs, we also calculated the QALY benchmarks for Australia using both the SF-6D and EQ-5D-3L. We expected that the QALY benchmarks for the SF-6D to be reduced compared to the EQ-5D-3L. One of the reasons for this expectation is due to the differences in the two instruments' classification systems leading to an increased 'ceiling effect' for the EQ-5D-3L and a reduced sensitivity of the EQ-5D-3L in the upper bounds of good health where participants respond as approaching full health including perfect health (utility 1.0). It is noted that the EQ-5D-5L was developed to

address the limited sensitivity (lack of descriptive richness and serious ceiling effects (Herdman, Gudex et al. 2011)) of the EQ-5D-3L. Nevertheless, the EQ-5D-5L still suffers from ceiling effects as respondents approach full health (Campbell, Palmer et al. 2016).

Another important reason is that differences between country-specific value sets inherently include differences in protocols, the model selected to estimate the value set, country-specific cultural differences, and the study date for the generation of the value set. To illustrate, a recent study described the comparisons of value sets between different countries for the EQ-5D-3L and noted that the health-state description of 11223 (no problems with mobility or self-care, some problems performing usual activities, moderate pain or discomfort, and extremely anxious or depressed), the HSUs from the UK and USA value sets are 0.255 and 0.506, respectively (Clemens, Begum et al. 2014). This is substantially different and leads to differences in QALY estimates. Therefore country-specific value sets should be used whenever possible. The same study also established that using the Australian value set for the EQ-5D-3L resulted in significantly different mean HSUs compared to the UK. However, the paper also noted that the absolute differences were small and less than the minimal clinically important difference (mean HSUs: 0.87, 0.86, 0.89 using the Australian, UK and USA value sets, respectively) (Clemens, Begum et al. 2014).

The current model was employed for the USA, China, UK and Australia using the EQ-5D-3L HSUs to generate benchmark QALYs. We also used the SF-6D age- and sex-specific HSUs for Australia to provide further evidence of model validation and benchmark QALYs. Therefore, the differences described above have translated to the expected divergence in our QALY benchmarks across our four selected countries for the EQ-5D-3L; and within the same country (Australia) for the SF-6D that we expected to be reduced from the EQ-5D-5L because of the inherent ceiling effects (and therefore decreased discriminatory sensitivity) of the EQ-5D-3L compared to the SF-6D. As a consequence, these results highlight the need for

country- and MAUI-specific QALY benchmarks for health economists to use to internally validate their health economics models. Importantly, this model specification is flexible, and therefore can be used in other study and country populations once relevant population-specific mortality and utility valuation data has been included.

Our QALY benchmarks are a useful resource for health economic modellers, policy-decision makers and clinicians alike. We will continue to update the model with population-level utility data elicited from other commonly used MAUIs, including the EQ-5D-5L, SF-6D (for additional countries), 15 Dimension (15D), Health Utilities Index (HUI 2 and 3), the AQoL suite of instruments (notably the AQoL-8D), and others (e.g. PROMIS) as value sets become available for individual countries. Other researchers interested in generating similar benchmarks for their particular country or using alternative HSUs generated by other MAUIs are encouraged to contact the corresponding author for instructions on data requirements and format.

Importantly, we have provided an open source repository of the data generated from this study and we will include additional countries and instruments as they become available at the following website: http://menzies.utas.edu.au/qaly_benchmarks. Our team has the expertise and track-record in developing a global community of interest (Association 2007, Palmer and Group 2013, Palmer, Si et al. 2018). Over time, we plan for to develop a comprehensive open source repository on QALY benchmarking and to develop a community of interest around this.

Strengths and Limitations

There are several strengths of this study. First, the model underwent extensive internal validation testing (and was comprehensively and robustly proved to be valid), the process and outcomes of this rigour are described in this paper. Second, the model can easily be adapted

to include new countries, additional country-specific HSUs from these instruments used in this study and other MAUIs, and updated life tables as new data become available, to provide a comprehensive resource and community of interest: all QALY benchmarks will continue to be provided on our publicly accessible website.

There are some limitations to our study. The key limitation of our study is that the polynomial equations that were used to generate continuous QALYs do not have a perfect goodness-of-fit with the results generated from the Markov model. For example, the undiscounted mean QALY for the Australian woman at age 18 years was 60.16 years from the Markov model (Table 1) and it was 60.59 years from the polynomial equation prediction. Nevertheless, the cubic equations were generated with an overall reasonable goodness-of-fit. Therefore, health economists are encouraged to validate their QALYs against our results in Table 1-3 for abridged ages. For other ages, they can be compared against results generated from the polynomial equations that are presented in our online tool.

Another limitation is the use of the EQ-5D-3L where the value sets for each country are different and are now being replaced by the EQ-5D-5L value sets, nevertheless (and as noted in the paper) we have developed the first set of QALY benchmarks with the 'legacy' EQ-5D-3L and this instrument remains prevalent in the literature (including for some new studies). A related limitation is the variation of utility values across MAUIs. An important recommendation regarding the use of QALY benchmarks is the need to compare like with like. Utility values generated from different instruments can vary and this is a limitation. On the other hand, this is a known fact in the international health preferences research and health economics literature (Richardson, Iezzi et al. 2015) and our paper has demonstrated this with a comparison of two instruments for one country. For example, Australian population norm data for young females has been reported to be 0.78 using the SF-6D (ages 18-30 years) (Norman, Church et al. 2013), 0.89 using the AQoL-8D (ages 16-24 years) (Maxwell, Ozmen

et al. 2016) and 0.90 using the EQ-5D-3L (ages 18-24 years) (Clemens, Begum et al. 2014). Therefore, using the QALY benchmarks and study-specific QALYs generated from the same instrument and country-specific value sets if available is recommended.

CONCLUSIONS

The principal aim of this paper was to provide age-and sex- population norms for QALYs in four countries to serve as a benchmark for health economic modellers when they conduct internal validation of their economic model. We have developed an accessible health economics tool to generate country-, sex- and MAUI-specific benchmark QALYs that will be refined, expanded and updated as new data become available, and will be made publicly available on the Internet. This open source repository aims to become the gold-standard resource for QALY benchmarks, and also provide the platform for a community of interest to develop. We have demonstrated the importance of generating QALYs for each instrument, and that QALYs generated from one instrument, or another country-specific value set, are not comparable with those generated from another.

Using these benchmarks now and into the future will provide the opportunity for researchers and policy-makers to quantify and further understand the impact of diseases on QALYs versus the general population.

Funding

This study was funded by the Centre of Excellence in Population Ageing Research, Australian Research Council (CE170100005) and the National Health and Medical Research Council (NHMRC) (Grant number: APP1127827). L.S. is supported by a NHMRC Early Career Fellowship (Grant number: GNT1139826). The funding body was not involved in the design of the study, or data collection, analysis, and interpretation or in writing the manuscript.

References

Ahmad, H., B. V. Taylor, I. van der Mei, S. Colman, B. A. O'Leary, M. Breslin and A. J. Palmer (2017). "The impact of multiple sclerosis severity on health state utility values: Evidence from Australia." Mult Scler **23**(8): 1157-1166.

Ara, R. and J. Brazier (2014). Cost-effectiveness modelling using health state utility values. Encyclopaedia of Health Economics. A. J. Culyer. Amsterdam, Elsevier.

Association, A. D. (2007). "Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting." Diabetes Care **30**(6): 1638-1646.

Augustovski, F. A., V. E. Irazola, A. P. Velazquez, L. Gibbons and B. M. Craig (2009). "Argentine valuation of the EQ-5D health states." Value Health **12**(4): 587-596.

Brazier, J., J. Roberts and M. Deverill (2002). "The estimation of a preference-based measure of health from the SF-36." J Health Econ **21**(2): 271-292.

Brazier, J., J. Roberts, A. Tsuchiya and J. Busschbach (2004). "A comparison of the EQ-5D and SF-6D across seven patient groups." Health Econ **13**(9): 873-884.

Campbell, J. A., G. A. Jelinek, T. J. Weiland, N. Nag, S. L. Neate, A. J. Palmer, B. Mulhern, A. De Livera and S. Simpson-Yap (2020). "SF-6D health state utilities for lifestyle, sociodemographic and clinical characteristics of a large international cohort of people with multiple sclerosis." Qual Life Res. 2020 Sep;29(9):2509-2527.

Campbell, J. A., A. J. Palmer, A. Venn, M. Sharman, P. Otahal and A. Neil (2016). "A Head-to-Head Comparison of the EQ-5D-5L and AQoL-8D Multi-Attribute Utility Instruments in Patients Who Have Previously Undergone Bariatric Surgery." Patient. 2016 Aug;9(4):311-22.

Chilton, S., J. S. Nielsen and J. Wildman (2020). "Beyond COVID- 19: How the 'dismal science' can prepare us for the future." Health Econ 2020 Aug;29(8):851-853.

National Bureau of Statistics, China. (2010). Sixth National Population Census of the People's Republic of China.

Clarke, P. M., A. M. Gray, A. Briggs, A. J. Farmer, P. Fenn, R. J. Stevens, D. R. Matthews, I. M. Stratton and R. R. Holman (2004). "A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68)." Diabetologia **47**(10): 1747-1759.

Clemens, S., N. Begum, C. Harper, J. A. Whitty and P. A. Scuffham (2014). "A comparison of EQ-5D-3L population norms in Queensland, Australia, estimated using utility value sets from Australia, the UK and USA." Qual Life Res **23**(8): 2375-2381.

Culyer, A. J. (2014). Culyer AJ. Encyclopaedia of Health Economics. 1st ed. Oxford:Elsevier.

Devlin, N. J. and R. Brooks (2017). "EQ-5D and the EuroQol group: past, present and future." Appl Health Econ Health Policy **15**(2): 127-137.

Dolan, P. (1997). "Modeling valuations for EuroQol health states." Med Care. 1997 Nov;35(11):1095-108.

Downey, L. E., A. Mehndiratta, A. Grover, V. Gauba, K. Sheikh, S. Prinja, R. Singh, F. A. Cluzeau, S. Dabak and Y. Teerawattananon (2017). "Institutionalising health technology

assessment: establishing the Medical Technology Assessment Board in India." BMJ Glob Health 2017 Jun 26;2(2): e000259.

Drummond, M. F., M. J. Sculpher, K. Claxton, G. L. Stoddart and G. W. Torrance (2015). Methods for the economic evaluation of health care programmes, Oxford university press.

Dunlop, W. C. N., N. Mason, J. Kenworthy and R. L. Akehurst (2017). "Benefits, Challenges and Potential Strategies of Open Source Health Economic Models." PharmacoEconomics **35**(1): 125-128.

Eddy, D. M., W. Hollingworth, J. J. Caro, J. Tsevat, K. M. McDonald and J. B. Wong "Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7." Value Health **15**(6): 843-850.

Eurostat. (2018). "Concepts and Definitions." from http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GLOSSARY_NOM_DTL_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=16910968&RdoSearch=BEGIN&TxtSearch=healthy&CboTheme=&IntCurrentPage=1.

Field, M. and M. Gold (1998). Summarizing Population Health: Directions for the Development and Application of Population Metrics. Washington DC, National Academy Press.

Flanagan, W., J. Boswell-Purdy, C. Le Petit and J. M. Berthelot (2005). "Estimating summary measures of health: a structured workbook approach." Popul Health Metr **3**(1): 5.

GBD DALYs and Hale Collaborators (2017). "Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016." Lancet 2017 Sep 16;390(10100):1260-1344.

Gold, M. R., D. Stevenson and D. G. Fryback (2002). "HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health." Annu Rev Public Health **23**: 115-134.

Gray, A. M., P. M. Clarke, J. L. Wolstenholme and S. Wordsworth (2010). Applied methods of cost-effectiveness analysis in healthcare, OUP Oxford.

Hennessy, D. A., W. M. Flanagan, P. Tanuseputro, C. Bennett, M. Tuna, J. Kopec, M. C. Wolfson and D. G. Manuel (2015). "The Population Health Model (POHEM): an overview of rationale, methods and applications." Popul Health Metr **13**: 24.

Hoogenveen, R. T., P. H. van Baal and H. C. Boshuizen (2010). "Chronic disease projections in heterogeneous ageing populations: approximating multi-state models of joint distributions by modelling marginal distributions." Math Med Biol **27**(1): 1-19.

Imai, K. and S. Soneji (2007). "On the Estimation of Disability-Free Life Expectancy: Sullivan' Method and Its Extension." J Am Stat Assoc **102**(480): 1199-1211.

Institute for Health Metrics and Evaluation (2013). GBD Results for the United States, IHME.

Jagger, C., M. McKee, K. Christensen, K. Lagiewka, W. Nusselder, H. Van Oyen, E. Cambois, B. Jeune and J. M. Robine (2013). "Mind the gap--reaching the European target of a 2-year increase in healthy life years in the next decade." Eur J Public Health **23**(5): 829-833.

Kassebaum, N. J., M. Arora, R. M. Barber, Z. A. Bhutta, J. Brown, A. Carter, D. C. Casey, F. J. Charlson, M. M. Coates and M. Coggeshall (2016). "Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015." The Lancet **388**(10053): 1603-1658.

Kind, P., G. Hardman and S. Macran (1999). UK population norms for EQ-5D, Centre for Health Economics, University of York.

Kularatna, S., J. Byrnes, Y. K. Chan, C. F. Ski, M. Carrington, D. Thompson, S. Stewart and P. A. Scuffham (2017). "Comparison of the EQ-5D-3L and the SF-6D (SF-12) contemporaneous utility scores in patients with cardiovascular disease." Qual Life Res **26**(12): 3399-3408.

Lancsar, E., Y. Gu, D. Gyrd-Hansen, J. Butler, J. Ratcliffe, L. Bulfone and C. Donaldson (2020). "The relative value of different QALY types. J Health Econ 2020 Mar;70:102303.

doi: 10.1016/j.jhealeco.2020.102303. Epub 2020 Feb 13.

MacQuilkan, K., P. Baker, L. Downey, F. Ruiz, K. Chalkidou, S. Prinja, K. Zhao, T. Wilkinson, A. Glassman and K. Hofman (2018). "Strengthening health technology assessment systems in the global south: a comparative analysis of the HTA journeys of China, India and South Africa." Glob Health Action 2018;11(1):1527556.

Maxwell, A., M. Ozmen, A. Iezzi and J. Richardson (2016). Norms for the AQoL-6D and AQoL-8D multi attribute utility instruments. Research paper 94. Centre for Health Economics, Monash University.

McAnulty, J. M. and K. Ward (2020). "Suppressing the Epidemic in New South Wales." N Engl J Med. 2020 May 21;382(21):e74. doi: 10.1056/NEJMc2011592. Epub 2020 May 8.

Murray, C. (1994). "Quantifying the burden of disease: the technical basis for disability-adjusted life years." Bulletin of the World Health Organization **72**: 429-445.

Murray, C. and A. Lopez (1994). "Quantifying disability: data, methods and results." Bulletin of the World Health Organization **72**: 481-494.

Neumann, P. J. and G. D. Sanders (2017). "Cost-effectiveness analysis 2.0." N Engl J Med **376**(3): 203-205.

NICE National Institute for Health and Care Excellence (2018). "Developing NICE guidelines: the manual. Chapter 7: Incorporating Economic Evaluation."

Norman, R., J. Church, B. van den Berg and S. Goodall (2013). "Australian health-related quality of life population norms derived from the SF-6D." Aust N Z J Public Health **37**(1): 17-23.

Norman R, Viney R, Brazier J, Burgess L, Cronin P, King M, Ratcliffe J, Street D. Valuing SF-6D health states using a discrete choice experiment. Med Desc Making. 2014 Aug;34(6):773-86.

OECD (2020). "COVID-19 Focus on the Global Economy."

Palmer, A. J. and M. H. M. Group (2013). "Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting." Value Health **16**(4): 670-685.

Palmer, A. J., L. Si, M. Tew, X. Hua, M. S. Willis, C. Asseburg, P. McEwan, J. Leal, A. Gray and V. Foos (2018). "Computer modeling of diabetes and its transparency: a report on the eighth mount hood challenge." Value Health **21**(6): 724-731.

Prinja, S., L. E. Downey, V. K. Gauba and S. Swaminathan (2018). Health technology assessment for policy making in India: current scenario and way forward, Springer.

Richardson, J., A. Iezzi and M. A. Khan (2015). "Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects." Qual Life Res **24**(8): 2045-2053.

Richardson, J., A. Iezzi, M. A. Khan, G. Chen and A. Maxwell (2016). "Measuring the sensitivity and construct validity of 6 utility instruments in 7 disease areas." Med Decis Making **36**(2): 147-159.

Severens, J. L. and R. J. Milne (2004). "Discounting Health Outcomes in Economic Evaluation: The Ongoing Debate." Value Health **7**(4): 397-401.

Shaw, J. W., J. A. Johnson and S. J. Coons (2005). "US valuation of the EQ-5D health states: development and testing of the D1 valuation model." Med Care **43**(3): 203-220.

Si, L., L. Shi, M. Chen and A. J. Palmer (2017). "Establishing benchmark EQ-5D-3L population health state utilities and identifying their correlates in Gansu Province, China." Qual Life Res 2017 Nov;26(11):3049-3058.

Si, L., T. M. Winzenberg, Q. Jiang and A. J. Palmer (2015). "Screening for and treatment of osteoporosis: construction and validation of a state-transition microsimulation cost-effectiveness model." Osteoporos Int **26**(5): 1477-1489.

Austraian Bureau of Statistics (ABS). (2015). "Life Tables, States, Territories and Australia." Life tables | Australian Bureau of Statistics (abs.gov.au)

Office of National Statistics. "Office of National Statistics. Statistics of National Life Tables. United Kingdom." National life tables – life expectancy in the UK Statistical bulletins - Office for National Statistics (ons.gov.uk)

Sullivan, D. F. (1971). "A single index of mortality and morbidity." HSMHA Health Rep **86**(4): 347-354.

Sullivan, P. W. and V. Ghushchyan (2006). "Preference-Based EQ-5D index scores for chronic conditions in the United States." Med Decis Making **26**(4): 410-420.

Torrance, G. W. (1987). "Utility approach to measuring health-related quality of life." J Chronic Dis **40**(6): 593-603.

Versteegh, M. M., I. C. Ramos, N. C. Buyukkaramikli, A. Ansaripour, V. T. Reckers-Droog and W. B. Brouwer (2019). "Severity-Adjusted Probability of Being Cost Effective." PharmacoEconomics **37**(9): 1155-1163.

Viney, R., R. Norman, M. T. King, P. Cronin, D. J. Street, S. Knox and J. Ratcliffe (2011). "Time trade-off derived EQ-5D weights for Australia." Value Health **14**(6): 928-936.

Wolfson, M. (1996). "Health-adjusted life expectancy." Health Rep **Summer; 8**(1): 41-46.

Xu, J., S. L. Murphy, K. D. Kochanek and B. A. Bastian (2016). "Deaths: Final Data for 2013." Natl Vital Stat Rep **64**(2): 1-119.

Tables

Table 1: Undiscounted age- and sex-specific quality adjusted life years (QALYs) by country generated by the benchmark QALY model *

Age	USA**		China**		UK**		Australia**		Australia***	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
18	54.62 (54.53, 54.72)	58.90 (58.80, 59.00)	55.55 (55.38, 55.72)	60.21 (60.09, 60.33)	58.01 (57.38, 58.59)	61.43 (60.90, 61.93)	57.11 (56.22, 57.95)	60.16 (59.21, 61.05)	50.9 (50.51, 51.29)	51.79 (51.39, 52.17)
20	52.85 (52.76, 52.94)	57.09 (56.09, 57.19)	53.70 (55.53, 53.86)	58.32 (58.20, 58.44)	56.16 (55.56, 56.72)	59.57 (59.05, 60.05)	55.35 (54.48, 56.16)	59.39 (57.48, 59.26)	49.33 (48.95, 49.71)	50.26 (49.87, 50.63)
25	48.54 (48.46, 48.63)	52.60 (52.51, 52.69)	49.00 (48.85, 49.15)	53.45 (53.34, 53.56)	51.04 (50.60, 51.46)	54.35 (53.88, 54.79)	50.97 (50.52, 51.40)	54.56 (54.09, 55.02)	45.42 (45.08, 45.77)	46.65 (46.23, 47.07)
30	43.24 (43.17, 43.32)	47.04 (46.96, 47.12)	44.30 (44.19, 44.41)	48.41 (48.21, 48.61)	46.54 (46.14, 46.93)	49.77 (49.35, 50.18)	46.58 (46.18, 46.98)	50.09 (49.65, 50.52)	41.52 (41.20, 41.83)	42.83 (42.44, 43.21)
35	39.06 (38.99, 39.12)	42.70 (42.63, 42.78)	39.57 (39.44, 39.70)	43.60 (43.51, 43.69)	41.16 (40.80, 41.51)	44.25 (43.86, 44.63)	41.29 (40.83, 41.74)	44.64 (44.14, 45.12)	37.1 (36.79, 37.40)	38.2 (37.87, 38.53)
40	33.72 (33.65, 33.79)	37.13 (37.06, 37.21)	34.88 (34.75, 35.01)	38.70 (38.54, 38.86)	36.83 (36.51, 37.15)	39.85 (39.50, 40.18)	37.07 (36.65, 37.47)	40.32 (39.86, 40.76)	33.31 (33.03, 33.58)	34.51 (34.2, 34.8)
45	29.76 (29.70, 29.82)	33.05 (32.98, 33.12)	30.81 (30.72, 30.90)	33.93 (33.78, 34.07)	30.07 (29.43, 30.69)	33.16 (32.69, 33.62)	31.42 (31.12, 31.70)	34.42 (34.10, 34.74)	28.78 (28.53, 29.02)	29.87 (29.54, 30.2)
50	25.08 (25.00, 25.15)	28.11 (28.02, 28.19)	26.16 (26.05, 26.27)	29.53 (29.40, 29.65)	26.22 (25.66, 26.76)	29.15 (28.75, 29.55)	27.50 (27.24, 27.75)	30.40 (30.12, 30.68)	25.19 (24.98, 25.40)	26.38 (26.09, 26.67)
55	21.58 (21.51, 21.64)	24.40 (24.32, 24.47)	21.99 (21.88, 22.10)	24.54 (24.38, 24.70)	20.86 (20.36, 21.35)	24.05 (23.62, 24.46)	23.69 (23.47, 23.91)	26.45 (26.21, 26.70)	21.27 (21.05, 21.49)	22.66 (22.37, 22.95)
60	17.87 (17.81, 17.93)	20.33 (20.26, 20.40)	17.62 (17.44, 17.79)	20.07 (19.89, 20.25)	17.52 (17.10, 17.94)	20.42 (20.06, 20.78)	20.02 (19.83, 20.20)	22.59 (22.38, 22.80)	17.97 (17.78, 18.15)	19.35 (19.1, 19.6)
65	14.82 (14.77, 14.87)	16.94 (16.88, 16.99)	14.13 (13.99, 14.27)	16.15 (16.00, 16.29)	14.38 (14.05, 14.70)	16.32 (16.02, 16.61)	15.91 (15.76, 16.06)	18.18 (18.00, 18.35)	14.5 (14.33,14.68)	15.34 (15.1, 15.58)
70	11.45 (11.40, 11.49)	13.18 (13.12, 13.23)	10.66 (10.54, 10.77)	12.14 (11.96, 12.31)	11.40 (11.14, 11.66)	13.10 (12.86, 13.33)	12.72 (12.6, 12.84)	14.70 (14.56, 14.84)	11.59 (11.46, 11.73)	12.41 (12.21, 12.6)
75	8.91 (8.88, 8.95)	10.33 (10.29, 10.37)	8.12 (8.01, 8.23)	9.12 (8.91, 9.31)	8.37 (8.11, 8.63)	9.21 (9.00, 9.42)	9.53 (9.39, 9.67)	11.01 (10.9, 11.12)	8.54 (8.41, 8.67)	9.65 (9.5, 9.8)
80	6.23 (6.19, 6.27)	7.26 (7.21, 7.31)	5.70 (5.49, 5.91)	6.22 (5.97, 6.46)	6.09 (5.90, 6.28)	6.77 (6.61, 6.92)	7.00 (6.89, 7.10)	8.14 (8.06, 8.22)	6.27 (6.18, 6.37)	7.13 (7.02, 7.24)
85	4.57 (4.54, 4.60)	5.28 (5.25, 5.32)	3.48 (3.10, 3.83)	4.35 (3.92, 4.74)	4.24 (4.11, 4.37)	4.74 (4.63, 4.85)	4.91 (4.84, 4.99)	5.69 (5.64, 5.75)	4.4 (4.34, 4.47)	4.99 (4.91, 5.06)

*Values are expressed as mean (2.5^{th} percentile, 97.5^{th} percentile) in years; ** EQ-5D-3L used for age- and sex-specific health state utilities; *** SF-6D used for age- and sex-specific health state utilities.

Table 2: Age- and sex-specific quality adjusted life years (QALYs) by country generated by the benchmark QALY model, discounted at 3% annually *

Age	USA**		China**		UK**		Australia**		Australia***	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
18	25.07 (25.03, 25.11)	25.96 (25.91, 26.00)	25.96 (25.86, 26.03)	26.95 (26.90, 27.01)	26.18 (25.90, 26.45)	26.82 (26.58, 27.03)	25.48 (25.09, 25.85)	25.89 (25.48, 26.27)	22.71 (22.54, 22.89)	22.28 (22.11, 22.44)
20	24.73 (24.69, 24.77)	25.65 (25.61, 25.70)	25.59 (25.51, 25.66)	26.64 (26.58, 26.69)	25.86 (25.58, 26.12)	26.52 (26.29, 26.74)	25.19 (24.79, 25.56)	25.62 (25.22, 26.00)	22.45 (22.28, 22.62)	22.05 (21.88, 22.21)
25	23.86 (23.82, 23.90)	24.83 (24.79, 24.87)	24.56 (24.48, 24.63)	25.68 (25.62, 25.73)	24.71 (24.50, 24.91)	25.44 (25.22, 25.65)	24.38 (24.17, 24.59)	25.16 (24.94, 25.38)	21.73 (21.56, 21.89)	21.51 (21.32, 21.71)
30	22.34 (22.30, 22.38)	23.35 (23.31, 23.39)	23.37 (23.31, 23.43)	24.49 (24.39, 24.59)	23.71 (23.51, 23.91)	24.52 (24.31, 24.72)	23.45 (23.24, 23.64)	24.31 (24.09, 24.51)	20.9 (20.74, 21.06)	20.78 (20.59, 20.97)
35	21.22 (21.19, 21.26)	22.30 (22.26, 22.34)	21.99 (21.92, 22.07)	23.25 (23.20, 23.30)	22.09 (21.90, 22.28)	22.96 (22.76, 23.16)	21.89 (21.65, 22.13)	22.82 (22.56, 23.06)	19.67 (19.5, 19.83)	19.53 (19.35, 19.69)
40	19.28 (19.24, 19.32)	20.42 (20.38, 20.46)	20.43 (20.36, 20.51)	21.77 (21.67, 21.86)	20.84 (20.66, 21.02)	21.80 (21.61, 21.98)	20.71 (20.48, 20.94)	21.73 (21.48, 21.96)	18.61 (18.45, 18.76)	18.59 (18.43, 18.75)
45	17.91 (17.87, 17.95)	19.14 (19.11, 19.18)	19.03 (18.97, 19.08)	20.14 (20.06, 20.23)	17.94 (17.56, 18.31)	19.14 (18.87, 19.40)	18.51 (18.34, 18.68)	19.57 (19.39, 19.75)	16.96 (16.81, 17.1)	16.99 (16.8, 17.17)
50	15.88 (15.83, 15.93)	17.15 (17.10, 17.21)	17.04 (16.96, 17.11)	18.50 (18.43, 18.58)	16.51 (16.15, 16.85)	17.76 (17.51, 18.01)	17.10 (16.94, 17.26)	18.25 (18.08, 18.42)	15.67 (15.53, 15.8)	15.84 (15.66, 16.01)
55	14.37 (14.33, 14.41)	15.69 (15.64, 15.73)	15.09 (15.01, 15.16)	16.24 (16.13, 16.34)	13.86 (13.52, 14.18)	15.47 (15.20, 15.74)	15.55 (15.40, 15.69)	16.78 (16.63, 16.93)	13.96 (13.81, 14.1)	14.37 (14.19, 14.56)
60	12.51 (12.46, 12.55)	13.77 (13.72, 13.81)	12.72 (12.60, 12.85)	14.01 (13.88, 14.13)	12.27 (11.97, 12.56)	13.87 (13.63, 14.11)	13.86 (13.73, 13.98)	15.15 (15.00, 15.28)	12.44 (12.31, 12.57)	12.97 (12.81, 13.14)
65	10.89 (10.85, 10.92)	12.06 (12.02, 12.10)	10.72 (10.61, 10.83)	11.88 (11.77, 11.98)	10.61 (10.36, 10.84)	11.70 (11.49, 11.90)	11.62 (11.5, 11.73)	12.88 (12.76, 13.00)	10.59 (10.46, 10.72)	10.87 (10.7, 11.04)
70	8.81 (8.77, 8.84)	9.85 (9.81, 9.89)	8.47 (8.38, 8.55)	9.38 (9.25, 9.52)	8.84 (8.64, 9.04)	9.90 (9.72, 10.07)	9.78 (9.68, 9.87)	11.00 (10.89, 11.10)	8.91 (8.81, 9.02)	9.28 (9.14, 9.43)
75	7.15 (7.12, 7.18)	8.07 (8.04, 8.10)	6.73 (6.63, 6.82)	7.38 (7.22, 7.54)	6.81 (6.60, 7.03)	7.32 (7.15, 7.49)	7.70 (7.59, 7.81)	8.69 (8.60, 8.77)	6.9 (6.8, 7.01)	7.61 (7.49, 7.73)
80	5.18 (5.14, 5.21)	5.88 (5.84, 5.92)	4.90 (4.72, 5.08)	5.25 (5.04, 5.45)	5.18 (5.01, 5.34)	5.63 (5.50, 5.76)	5.91 (5.83, 6.00)	6.75 (6.68, 6.82)	5.3 (5.22, 5.38)	5.91 (5.82, 6)
85	3.86 (3.84, 3.89)	4.36 (4.33, 4.39)	3.08 (2.75, 3.29)	3.80 (3.43, 4.15)	3.74 (3.62, 3.85)	4.10 (4.00, 4.19)	4.32 (4.26, 4.39)	4.93 (4.88, 4.98)	3.87 (3.82, 3.93)	4.32 (4.25, 4.39)

*Values are expressed as mean (2.5^{th} percentile, 97.5^{th} percentile) in years; ** EQ-5D-3L used for age- and sex-specific health state utilities; *** SF-6D used for age- and sex-specific health state utilities.

Table 3: Age- and sex-specific quality adjusted life years (QALYs) by country generated by the benchmark QALY model, discounted at 5% annually [*]

Age	USA**		China**		UK**		Australia**		Australia***	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
18	17.38 (17.35, 17.41)	17.75 (17.72, 17.78)	18.05 (17.99, 18.10)	18.48 (18.44, 18.52)	18.01 (17.82, 18.19)	18.25 (18.10, 18.40)	17.47 (17.20, 17.73)	17.54 (17.27, 17.80)	15.57 (15.45, 15.69)	15.1 (14.98, 15.21)
20	17.24 (17.22, 17.27)	17.65 (17.62, 17.68)	17.91 (17.86, 17.97)	18.37 (18.34, 18.41)	17.90 (17.70, 18.07)	18.16 (18.00, 18.30)	17.37 (17.10, 17.62)	17.46 (17.19, 17.72)	15.48 (15.36, 15.6)	15.03 (14.91, 15.14)
25	16.90 (16.88, 16.93)	17.34 (17.31, 17.37)	17.49 (17.43, 17.54)	18.00 (17.96, 18.03)	17.38 (17.23, 17.52)	17.68 (17.53, 17.83)	17.08 (16.93, 17.22)	17.40 (17.25, 17.55)	15.22 (15.1, 15.34)	14.88 (14.74, 15.01)
30	16.11 (16.08, 16.13)	16.58 (16.55, 16.61)	16.95 (16.91, 16.99)	17.47 (17.40, 17.54)	16.97 (16.83, 17.11)	17.33 (17.18, 17.47)	16.71 (16.56, 16.85)	17.09 (16.93, 17.23)	14.89 (14.77, 15)	14.61 (14.48, 14.74)
35	15.59 (15.56, 15.62)	16.13 (16.10, 16.16)	16.27 (16.22, 16.32)	16.90 (16.87, 16.94)	16.11 (15.97, 16.25)	16.53 (16.38, 16.67)	15.89 (15.71, 16.06)	16.32 (16.14, 16.50)	14.28 (14.16, 14.39)	13.97 (13.85, 14.09)
40	14.45 (14.43, 14.48)	15.06 (15.03, 15.09)	15.44 (15.38, 15.50)	16.16 (16.09, 16.22)	15.51 (15.38, 15.65)	16.00 (15.86, 16.14)	15.34 (15.17, 15.50)	15.85 (15.67, 16.02)	13.78 (13.67, 13.89)	13.56 (13.44, 13.68)
45	13.71 (13.68, 13.74)	14.41 (14.38, 14.44)	14.71 (14.66, 14.75)	15.29 (15.22, 15.35)	13.65 (13.36, 13.94)	14.35 (14.15, 14.55)	14.01 (13.88, 14.14)	14.58 (14.44, 14.71)	12.84 (12.72, 12.94)	12.65 (12.51, 12.79)
50	12.43 (12.39, 12.47)	13.20 (13.16, 13.24)	13.48 (13.43, 13.54)	14.38 (14.32, 14.44)	12.86 (12.58, 13.12)	13.63 (13.44, 13.81)	13.24 (13.12, 13.36)	13.91 (13.78, 14.03)	12.13 (12.03, 12.23)	12.07 (11.93, 12.2)
55	11.51 (11.48, 11.55)	12.36 (12.32, 12.40)	12.24 (12.18, 12.30)	12.94 (12.85, 13.02)	11.06 (10.79, 11.32)	12.16 (11.95, 12.37)	12.34 (12.22, 12.45)	13.10 (12.98, 13.22)	11.07 (10.96, 11.19)	11.22 (11.07, 11.36)
60	10.26 (10.22, 10.29)	11.11 (11.08, 11.15)	10.59 (10.48, 10.69)	11.46 (11.35, 11.56)	10.05 (9.80, 10.28)	11.19 (10.99, 11.38)	11.28 (11.18, 11.38)	12.13 (12.02, 12.24)	10.13 (10.02, 10.23)	10.39 (10.25, 10.52)
65	9.15 (9.12, 9.18)	9.99 (9.95, 10.02)	9.15 (9.06, 9.24)	9.97 (9.88, 10.06)	8.91 (8.71, 9.11)	9.69 (9.51, 9.86)	9.71 (9.61, 9.80)	10.59 (10.49, 10.69)	8.85 (8.74, 8.95)	8.94 (8.8, 9.08)
70	7.58 (7.55, 7.61)	8.36 (8.33, 8.40)	7.41 (7.33, 7.48)	8.09 (7.97, 8.20)	7.63 (7.46, 7.80)	8.42 (8.27, 8.57)	8.39 (8.31, 8.47)	9.30 (9.21, 9.39)	7.65 (7.56, 7.74)	7.85 (7.73, 7.97)
75	6.30 (6.28, 6.33)	7.02 (6.99, 7.05)	6.02 (5.94, 6.10)	6.52 (6.38, 6.66)	6.03 (5.84, 6.22)	6.40 (6.25, 6.55)	6.79 (6.69, 6.89)	7.56 (7.48, 7.63)	6.08 (5.99, 6.18)	6.62 (6.51, 6.72)
80	4.65 (4.62, 4.68)	5.23 (5.19, 5.26)	4.48 (4.31, 4.64)	4.75 (4.56, 4.93)	4.70 (4.55, 4.84)	5.05 (4.94, 5.17)	5.35 (5.27, 5.43)	6.04 (5.98, 6.09)	4.79 (4.72, 4.87)	5.29 (5.2, 5.37)
85	3.51 (3.49, 3.54)	3.92 (3.89, 3.95)	2.86 (2.56, 3.15)	3.51 (3.16, 3.82)	3.46 (3.35, 3.57)	3.76 (3.67, 3.84)	4.00 (3.94, 4.06)	4.53 (4.48, 4.57)	3.58 (3.53, 3.64)	3.96 (3.9, 4.03)

[*] Values are expressed as mean (2.5^{th} percentile, 97.5^{th} percentile) in years; ** EQ-5D-3L used for age- and sex-specific health state utilities; *** SF-6D used for age- and sex-specific health state utilities.

Table 4: Differences in undiscounted QALYs generated by the EQ-5D-3L and SF-6D for Australian males and females.

Australia	**Males**				**Females**			
Age	EQ-5D-3L	SF-6D	Difference in QALYs*	Percentage difference	EQ-5D-3L	SF-6D	Difference in QALYs*	Percentage difference
18	57.11	50.90	6.21	11	60.16	51.79	8.37	14
20	55.35	49.33	6.02	11	59.39	50.26	9.13	15
25	50.97	45.42	5.55	11	54.56	46.65	7.91	15
30	46.58	41.52	5.06	11	50.09	42.83	7.26	14
35	41.29	37.10	4.19	10	44.64	38.20	6.44	14
40	37.07	33.31	3.76	10	40.32	34.51	5.81	14
45	31.42	28.78	2.64	8	34.42	29.87	4.55	13
50	27.50	25.19	2.31	8	30.40	26.38	4.02	13
55	23.69	21.27	2.42	10	26.45	22.66	3.79	14
60	20.02	17.97	2.05	10	22.59	19.35	3.24	14
65	15.91	14.50	1.41	9	18.18	15.34	2.84	16
70	12.72	11.59	1.13	9	14.70	12.41	2.29	16
75	9.53	8.54	0.99	10	11.01	9.65	1.36	12
80	7.00	6.27	0.73	10	8.14	7.13	1.01	12
85	4.91	4.40	0.51	10	5.69	4.99	0.70	12
Mean	29.41	26.41	**3.00**	**10**	32.05	27.47	**4.58**	**14**

Notes: supported by Figure 3.

Table 5: Quality adjusted life years (QALYs) polynomial prediction model by country, discounted at 0%, 3% and 5% annually*

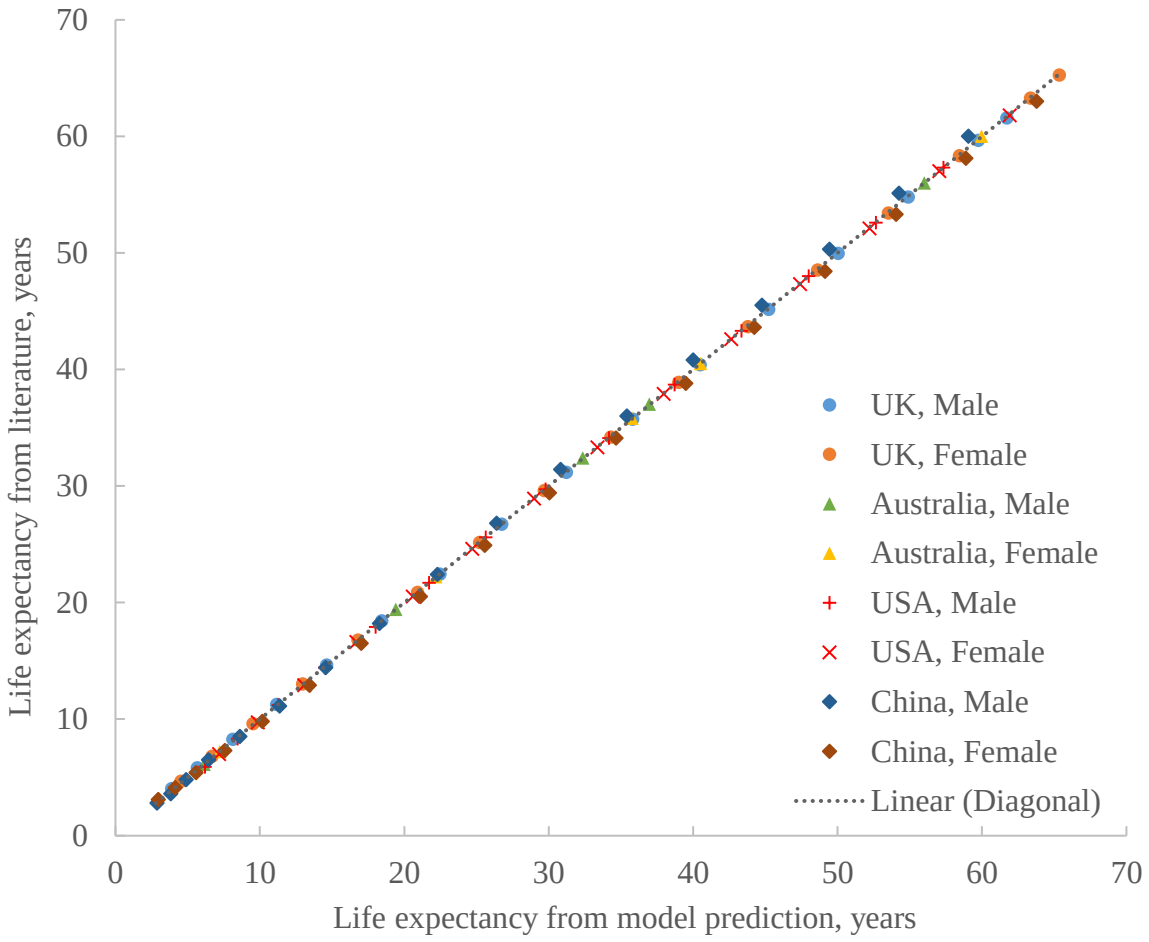
Country		β_0			β_1			β_2			β_3		
		0%	3%	5%	0%	3%	5%	0%	3%	5%	0%	3%	5%
Australia	Male	72.73	25.44	16.41	-0.779	0.133	0.142	-0.006	-0.008	-0.005	$6e^{-5}$	$4e^{-5}$	$2e^{-5}$
	Female	75.83	25.31	16.44	-0.739	0.631	0.135	-0.007	-0.008	-0.004	$6e^{-5}$	$4e^{-5}$	$1e^{-5}$
China	Male	70.48	25.63	16.59	-0.739	0.159	0.174	-0.007	-0.009	-0.006	$7e^{-5}$	$5e^{-5}$	$2e^{-5}$
	Female	74.63	25.82	16.63	-0.674	0.216	0.197	-0.009	-0.010	-0.006	$8e^{-5}$	$5e^{-5}$	$2e^{-5}$
UK	Male	75.38	26.38	16.84	-0.863	0.152	0.18	-0.006	-0.010	-0.007	$8e^{-5}$	$6e^{-5}$	$3e^{-5}$
	Female	77.87	26.33	16.80	-0.796	0.183	0.183	-0.008	-0.010	-0.006	$8e^{-5}$	$5e^{-5}$	$3e^{-5}$
USA	Male	72.17	26.77	17.59	-0.938	0.007	0.053	-0.002	-0.006	-0.004	$5e^{-5}$	$3e^{-5}$	$1e^{-5}$
	Female	76.93	27.63	18.19	-0.962	-0.002	0.024	-0.002	-0.005	-0.003	$4e^{-5}$	$2e^{-5}$	$4e^{-6}$

* QALY prediction model is expressed as a cubic polynomial model: $y=\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3$, where y denotes QALY and x denotes age.

Figures

Figure 1: A two-states Markov model to calculate the quality-adjusted life years

Figure 2: Goodness-of-fit test of life expectancy generated from the model and that from literature outlined in Appendix 1.

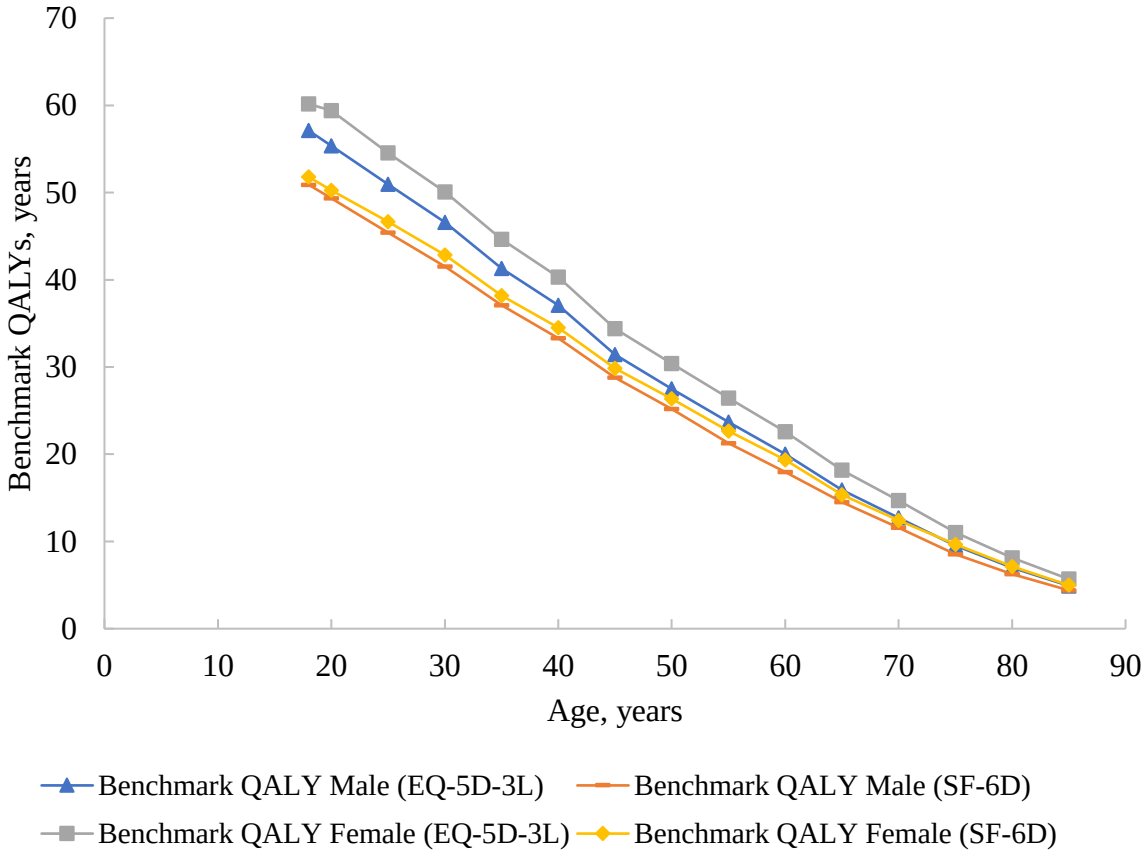


Notes:

A goodness-of-fit test was conducted by regressing life expectancies from model prediction to that from the literature contained in Appendix 1. The slope of the regression line as well as R-squared were provided where the value of 1 indicates perfect goodness-of-fit. In our life expectancy verification analysis, the slope of the regression line was 0.998. In addition, the adjusted R-squared and the root mean square error (MSE) of the regression model was 0.999 and 0.32, indicating life expectancy generated from our model closely matched to that from literature.

Each of the dots represents a pair of comparison, the dotted line represents the diagonal with 45-degree angle, (perfect goodness-of-fit). The slope of the regression line is 0.998. The R-squared and root MSE of the regression model equals to 0.999 and 0.32, indicating the model predictions fit well to that from literature.

Figure 3: Undiscounted age- and sex-specific benchmark QALYs for Australia using the EQ-5D-3L and SF-6D multi-attribute utility instruments



Notes: supported by Table 4.

Figure Legend

Figure 1: A two-states Markov model to calculate the quality-adjusted life years

Figure 2: Goodness-of-fit test of life expectancy generated from the model and that from literature outlined in Appendix 1.

Notes:

A goodness-of-fit test was conducted by regressing life expectancies from model prediction to that from the literature contained in Appendix 1. The slope of the regression line as well as R-squared were provided where the value of 1 indicates perfect goodness-of-fit. In our life expectancy verification analysis, the slope of the regression line was 0.998. In addition, the adjusted R-squared and the root mean square error (MSE) of the regression model was 0.999 and 0.32, indicating life expectancy generated from our model closely matched to that from literature.

Each of the dots represents a pair of comparison, the dotted line represents the diagonal with 45-degree angle, (perfect goodness-of-fit). The slope of the regression line is 0.998. The R-squared and root MSE of the regression model equals to 0.999 and 0.32, indicating the model predictions fit well to that from literature.

Figure 3: Undiscounted age- and sex-specific benchmark QALYs for Australia using the EQ-5D-3L and SF-6D multi-attribute utility instruments

Notes: supported by Table 4.