THE UNIVERSITY OF MELBOURNE

SHINE protocol for specifying model inputs, including rapid review and cross-walking (Version 1.0)

Samantha Howe & Tony Blakely

Scalable Health Intervention Evaluation (SHINE)
Population Interventions Unit, Centre for Epidemiology and Biostatistics
Melbourne School of Population and Global Health
University of Melbourne

SHINE Technical Report Series, No: 1

08 March 2023

Recommended citation:

Howe S, Blakely T (2023) SHINE protocol for specifying model inputs, including rapid review and cross-walking (Version 1.0). SHINE Technical Report Series, No 1. Available at *https://mspgh.unimelb.edu.au/centres-institutes/centre-for-epidemiology-and-biostatistics/research-group/shine#publications.* [Online publication date 08 March 2023]

TABLE OF CONTENTS

PART 1: OVERVIEW 3
WHAT IS SHINE? 3
SHINE MODEL INPUTS 3
PART 2: READILY AVAILABLE INPUT PARAMETERS AND THEIR SOURCES 5
AUSTRALIAN BUREAU OF STATISTICS & AUSTRALIAN INSTITUTE OF HEALTH AND WELLFARE 8
GLOBAL BURDEN DISEASE STUDY 9
AOTEAROA/ NEW ZEALAND COST DATA 9
PART 3: INPUT PARAMETERS THAT MIGHT NEED A RAPID REVIEW 9
PART 4: RAPID REVIEW STRATEGY FOR MODEL INPUTS 10
RAPID REVIEW STEPS 11
DOCUMENTATION 16
CROSS WALKING 16
EXPERT KNOWLEDGE ELICITATION 17
REFERENCES 18
APPENDICES 22

TABLE OF FIGURES

Figure 1. Key inputs for simulation modelling Chart 1: Pie chart sample 4
Figure 2. High vs. low intensity rapid review flowchart Chart 1: Pie chart sample 11
Figure 3. SHINE rapid review literature search process Chart 1: Pie chart sampl 12
Figure 4: Traditional hierarchy of evidence 14

PART 1: OVERVIEW

WHAT IS SHINE?

Scalable Health Intervention Evaluation (SHINE) utilises simulation modelling to determine the health and cost impacts of population interventions. SHINE leverages a vast array of existing epidemiological and economic data, applying these as inputs to simulation models to predict of the future health outcomes and costs resulting from health interventions, as well as impact of such interventions on health inequality in a population.

A variety of both macro and microsimulation models are utilised, depending on the health issue and research question at hand. Examples of SHINE's recent modelling work are:

- Exploring optimal COVID-19 policy options using an agent-based model (ABM) in the setting of Victoria, Australia. (1)
- Tobacco control modelling with a Markov simulation model and Proportional Multistate Life Table (PMSLT), in the setting of Aotearoa/New Zealand. (2)
- Estimating the impact of cold housing on cardiovascular disease using a PMSLT in the setting of Australia. (3)

This document does not go into detail on the various simulation models that SHINE applies. The purpose of this review is to act as a guide for appropriately sourcing the multiple inputs that are required for any project involving or contributing to simulation modelling of health interventions. The next sections will outline where input parameters are sourced for SHINE simulation modelling, when rapid literature reviews are required, and the strategy for these reviews.

SHINE MODEL INPUTS

Key inputs are required to populate any given simulation model, as depicted in Figure 1. This is a two-step process. Firstly, business as usual (BAU) is defined – BAU describes the predicted future under 'current force of change'. Conceptually specifying a BAU scenario is deceptively challenging. For example, what is the future smoking prevalence under BAU? Is it that if we stopped all tobacco control and did nothing more? Or is it if we applied the same intensity of policies in the future as in the past (albeit evolving from plain packaging [already done] to the next most likely intervention)? At times, SHINE will explicitly specify alternative BAU scenarios. The default position, though, is to define BAU as a future that has the same trends seen in the recent past (e.g. last 10 to 30 years) in risk factor, disease rates and all-cause mortality and morbidity rates applied to the future (approximately 20 years, then zero further change). SHINE models do not (yet) allow for migration, thus future BAU demographic data (age, sex and ethnicity) is that determined within the model and future birth predictions. There are also the relative risk (RR) associations between risk factors and the disease (e.g. between smoking and lung cancer). Whilst RRs can change over time (e.g. they have steadily increased over time for the association of smoking with diseases, due to the never smoking population getting healthier and healthier), the default is to posit no change in RRs into the future.

There are also potential BAU trends in health expenditure by disease phase, and income by disease phase. However, the SHINE default is no future trends in expenditure and income – other than that due to inflation, which is handled by all monetary values being real values for the base-year of the model.

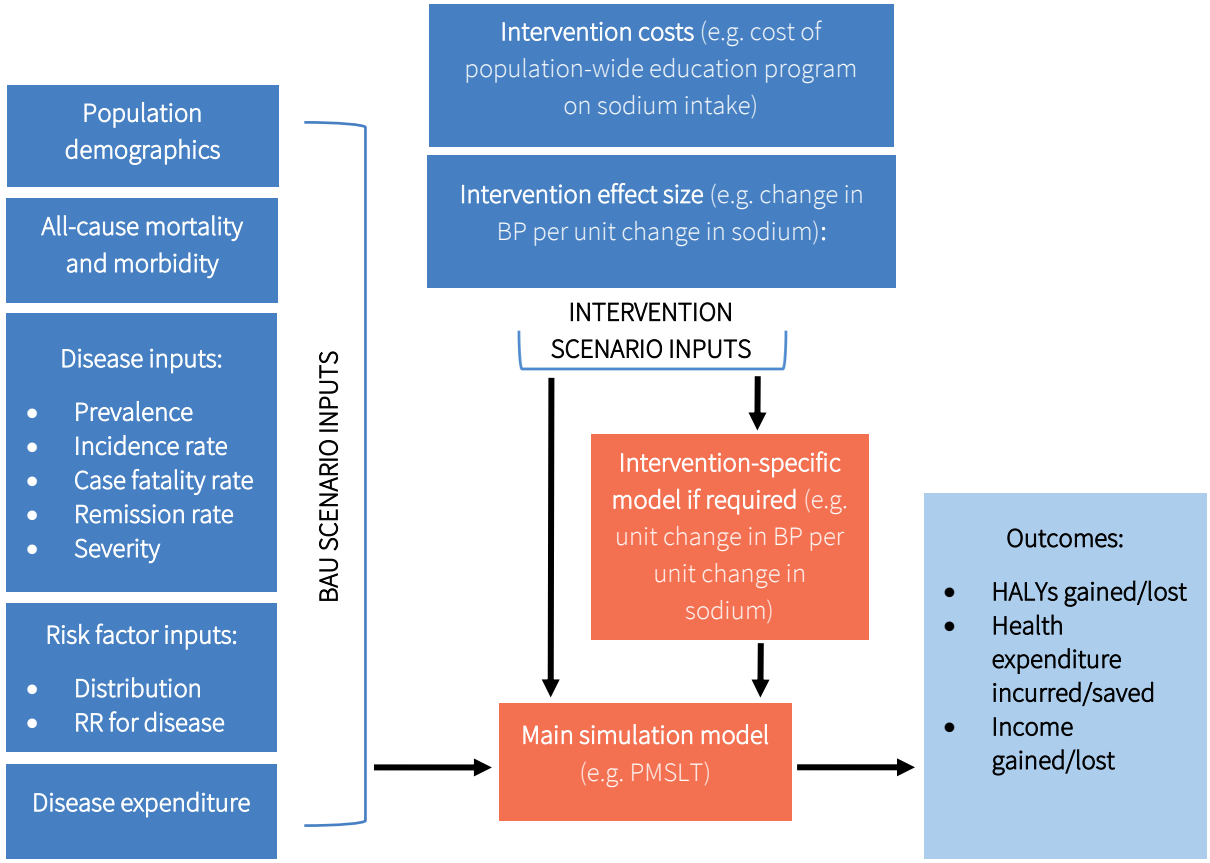

Figure 1. Key inputs for simulation modelling

BP = Blood Pressure HALY = Health-Adjusted Life Year; RR = Relative Risk.

Intervention inputs are then built on top of the BAU scenario(s), to test the impact of interventions on the population and answer questions such as "what are the health gains and costs (or cost savings) of intervention X on outcome Y, compared to BAU, and for whom?" Intervention inputs include the effect size for the relationship between an intervention and a risk factor or disease. These effect sizes may apply to the whole population of interest, or may be different for specific subgroups, depending on the research question and conceptualization of the intervention. Some interventions will also require their own intervention-specific model (e.g. how changes in sodium change systolic blood pressure, varying by sex, age and starting blood pressure, e.g. how a price change in high-salt foods will impact consumption of both high-salt and all other foods using own- and cross-price elasticities).

The full breadth of cost inputs depends on the perspective selected. The SHINE default is a healthcare perspective (capturing healthcare system costs, both Government and patient out-of-pocket funded), plus one additional social cost – namely changes in income earnings. However, a research question may require a broader societal perspective, considering all costs to all relevant parties. For example, this may include additional productivity costs such as that due to presenteeism, absenteeism, or carer costs, etc. (4) A societal perspective can be complex to fully parameterise; therefore, simplified approaches might be used, such as the 'health system + Gross Domestic Product (GDP) perspective' applied for recent COVID-19 modelling in Victoria (1), with the GDP perspective approximating the productivity losses resulting from lockdown throughout the pandemic simulation.

Population heterogeneity is a source of variation in simulation modelling. It is necessary to determine the level of heterogeneity that the population needs to be divided by for a given research question and model. The minimum level of heterogeneity is sex by age strata. However, addressing intervention impacts on health inequality is an important focus of SHINE modelling, such as heterogeneity by ethnicity (e.g., Aboriginal and Torres Strait Islanders and non-Indigenous people in the Australian setting, or Māori and non-Māori in Aotearoa/New Zealand) or socioeconomic status (SES). Modelling this heterogeneity is a demanding task, requiring all modelling and all model inputs to be not only by strata of sex and age, but also by strata of the additional heterogeneity covariate(s).

Accurate sourcing and specification of model inputs is important for the validity of model predictions. Given the time and resources required for model conceptualisation and calibration, and that any one simulation model requires 100s of data inputs (of which at least some will require separate literature reviews), it is critical to obtain model inputs in an efficient manner. The majority of SHINE model inputs are sourced from (derivations of) existing data generated by others. This may be from databases/repositories of population demographics, health expenditure and income by disease data, or groups of large epidemiological papers providing burden of disease data such as those from the Global Burden of Disease study (GBD). These sources are discussed in further detail in part two of this protocol.

However, some inputs, and often critically important inputs, need to be sourced elsewhere – most notably the 'effect size' and 'intervention cost' for the posited intervention. Focusing on the intervention effect size, this may be sourced from traditional primary literature, obtained via rapid review. Following the standard hierarchical order of quality of evidence, these reviews aim to obtain inputs from meta-analyses of randomized trials, single randomized trials, meta-analyses of observational studies (e.g., cohort and case control) or single observational studies. Part three and four of this protocol focuses on when literature searches of existing research are required to specify an intervention effect size.

The perfect data may not exist – in such cases parallel or indirect data may be 'cross-walked' to the setting or population of interest (e.g., if no heterogeneity in price elasticities by ethnicity is available, but findings on heterogeneity by socioeconomic status (SES) exist – then use the SES findings as proxy for ethnic heterogeneity). Where this is not possible, for example in determining the effect size for an intervention that is yet to be evaluated, techniques to obtain reasonable consensus such as expert knowledge elicitation may be necessary. These methods are outlined in part four.

At the outset of a modelling project, all inputs that are expected to be required should be formally listed, specifying the intended source of each (i.e., readily available input source, or a rapid review of the literature; each described further in next sections), how they should be parameterised, and any other important factors requiring consideration. Appendix 1 provides a template for the factors requiring consideration at the outset.

PART 2: READILY AVAILABLE INPUT PARAMETERS AND THEIR SOURCES

In general terms, BAU input parameters are usually readily available from reliable sources and do not require literature reviews. These are inputs that can be sourced from surveys conducted periodically by governing bodies for a given

jurisdiction, such as the Australian Bureau of Statistics, or from the GBD's large repository of disease-specific epidemiological data. These inputs and common sources are summarised in Table 1.

Table 1: readily available input parameters

Parameter	Sources	Considerations/issues
Demographic data	Population distribution by age and sex • ABS, Australia (https://www.abs.gov.au/) • GBD, any country (5) (https://vizhub.healthdata.org/gbd-results/) Life expectancy (life tables) • GBD, global measure (6) (https://ghdx.healthdata.org/record/ihme-data/gbd-2019-life-tables-1950-2019) • ABS, Australia (7) (https://www.abs.gov.au/statistics/people/population/life-tables/latest-release)	Consideration: what level of heterogeneity needs to be included? • What is the research question?
Disease-specific epidemiology	Disease incidence, prevalence, mortality • GBD, any country (5) (https://vizhub.healthdata.org/gbd-results/) • AIHW, Australia (https://www.aihw.gov.au/reports-data/flagships/australias-health) Disability weights for burden of disease calculations • GBD, global measure (8) (https://ghdx.healthdata.org/record/ihme-data/gbd-2019-disability-weights) Disability-Adjusted Life Years (DALYs) by disease/risk factor • Australian Burden of Disease Study, Australia (9) (https://www.aihw.gov.au/reports-data/health-conditions-disability-deaths/burden-of-disease/overview) • GBD, any country (5) https://vizhub.healthdata.org/gbd-results/	GBD study does not report remission rates ABDS and GBD output for Australia differ due to some differing methods. (9) If considering the Australian population, decide whether comparisons to other countries may be needed. If so, best to use GBD output to allow for more accurate cross-country comparisons Consideration: what is the reference year?

Risk factor epidemiology	Risk factor prevalence/distribution • Various national surveys available via ABS/AIHW, Australia e.g. National Drug Strategy Household Survey 2019 (10) Risk factor-disease relationship • GBD, global (11) (https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30752-2/fulltext) (relative risks by age and sex available in Supplementary Appendix 1, page 334)	No single source. Depends on the population of interest Consideration: how recent is the survey?
Disease-specific costs	Overall healthcare system expenditure by disease in Australia • AIHW, Australia (https://www.aihw.gov.au/reports/health-welfare-expenditure/disease-expenditure-australia/contents/about) Hospital expenditure by disease • IHACPA National Hospital Cost Data Collection (NHCDC), Australia (12) (https://www.ihacpa.gov.au/resources/national-hospital-cost-data-collection-nhcdc-public-hospitals-report-round-26-financial-year-2021-22) Income and productivity loss by disease • Blakely et al., New Zealand (13)	Lack of Australian linked income data for diseases • Solution: rich New Zealand population data available, compiled from linked databases by Blakely et al. (13)
Other costs	Household expenditure • ABS, Australia (https://www.abs.gov.au/statistics) Health system costs • AIHW, Australia (https://www.aihw.gov.au/reports-data) • Blakely et al., New Zealand (14)	Blakely et al. (14) data can be used to supplement AIHW data on health expenditure, to disaggregate costs into death-year and non-death year • Health system cost by sex, age and proximity to death in New Zealand

To source appropriate BAU inputs, it is important to understand the relevant risk factors and disease pathways that interventions may act on. For example, if an intervention impacts the severity of a disease, it needs to be considered how this will be measured – i.e. is severity measured directly via the disease's disability weight and distribution of cases that are classified as mild vs. moderate vs. severe, or can mortality risk be used as the proxy for severity?

When obtaining demographic data, it is important to consider what the base or reference year is for modelling. This is also required for costing – costs obtained from the literature need to be updated to the base year, taking into account inflation. The units for costing should also be considered – do costs needs to be converted to another currency (USD commonly used) using standard exchange rates or purchasing power parity (PPP) exchange rates? (15)

Key sources for inputs for modelling of the Australian population, from both Australian-specific and international data sources are described in further detail below.

AUSTRALIAN BUREAU OF STATISTICS & AUSTRALIAN INSTITUTE OF HEALTH AND WELLFARE

The Australian Bureau of Statistics (ABS) is a key source of demographic data in Australia obtained periodically through various national surveys. The ABS obtains data on the prevalence of different risk factors, such as smoking, alcohol consumption and physical activity. Information on risk factors and health outcomes is also stratified into Aboriginal and Torres Strait Islanders and non-Indigenous populations, and by SES using socio-economic indexes by geographic area. (16)

A variety of data available from the ABS that is not directly health-related, but may be necessary for particular research questions, includes population data on housing and education. Economic data can also be extracted, including gross domestic product (GDP), and average household income.

The Australian Institute of Health and Welfare (AIHW) contains further data on health. Similarly to the ABS, data is collated from a variety of datasets, to create comprehensive information on different health outcomes, including risk factors and diseases, and trends over time. Health information is also available by different strata, including: level of remoteness, Aboriginal and Torres Strait Islander population compared to non-Indigenous population, socioeconomic status, and age.

The Australian Burden of Disease study (ABDS) is produced by the AIHW, and utilises methods adapted from the GBD. (9) It is important to note that, due to some differences in methodology, disability-adjusted life years (DALYs) and related estimates produced by the ABDS may differ from GBD results for Australia. (9) The choice between using results from either of these sources may depend on whether it is important to make comparisons to other countries for a particular research project.

GLOBAL BURDEN DISEASE STUDY

The GBD is a key resource for disease modelling, containing data on disease incidence rates, prevalence, and case fatality rates, for different countries by age and sex. These inputs can be retrieved from Institute for Health Metrics and Evaluation (IHME) online tool. (5) The GBD also provides disability weights (DW) for calculation of disease burden (i.e. disease morbidity, or years lived with disability [YLDs]), which may be an output/component of an output in SHINE modelling. (8)

AOTEAROA/ NEW ZEALAND COST DATA

Statistics New Zealand collates linked population data from various sources through its Integrated Data Infrastructure (IDI). (17) This data is only available on request from researchers, but has been separately utilised by Blakely et al. to calculate income loss by disease. (13) Currently, data of this specificity is not available directly in Australia. This data for the Aotearoa/New Zealand (A/NZ) population can be used to approximate income loss by disease in Australia.

Blakely et al. has also estimated general (non-disease specific) health system costs by age and sex, additionally disaggregated into death year and non-death year, in the A/NZ population. (14) This data can be combined with AIHW data on health system costs.

PART 3: INPUT PARAMETERS THAT MIGHT NEED A RAPID REVIEW

The primary inputs usually requiring rapid reviews are those relating to interventions. This includes the effect size for the relationship between interventions and an outcome, being either a risk factor or disease, and cost information for interventions.

An important consideration for effect size data, is determining at what point in the causal pathway the intervention is acting on an outcome. For examples: is this an intervention that aims to reduce the prevalence of a specific risk factor (such as tobacco use); does the intervention reduce the case fatality rate for a specific disease; or does the intervention act at multiple points, for example COVID-19 vaccinations impacting the risk of COVID-19 infection, hospitalisation and death? This will impact what disease-specific epidemiological information is required to input into the model (as previously mentioned). For the COVID-19 vaccination example, this required data to be collected on the hospitalisation rate, infection rate and infection fatality rate for COVID-19 in Victoria for modelling purposes. (1)

The availability of intervention-related data, both cost and effect sizes, is often limited by whether this intervention has been implemented and evaluated in other settings, and how novel the idea is. Newer interventions may have only been implemented or evaluated in populations external to that being modelled, or in different contexts (e.g. observing a different outcome to that of interest); novel interventions are also more likely to require consensus approaches to determine effect sizes if there is a lack of published evidence. The process for determining the best data for interventions, or any inputs identified as requiring a literature review, is detailed in the next section.

PART 4: RAPID REVIEW STRATEGY FOR MODEL INPUTS

Rapid reviews are a form of evidence synthesis, generally applied in the context of direct policy assessment. (18) They are beneficial in being able to simplify the processes of a systematic review, which can take upwards of a year to complete. (19)

Existing rapid review guidelines suggest a time frame of around 8-12 weeks to complete this type of review, with many including conducting a meta-analysis. (18) However, these guidelines generally pertain to rapid reviews of single intervention questions with the purpose of direct to policymaker decision-making. Given that input parameterisation is only one component of the complex modelling process, it is not feasible to spend 12 weeks extracting every model input. Rapid reviews for SHINE model inputs need to be undertaken with a more efficient, streamlined approach. Only one protocol appears to exist in which guidelines for searching the literature for modelling inputs is provided, produced by the BODE^3 in 2011. (20) The following strategy builds on a fundamental aspect of the BODE^3 process, in that the goal is not to obtain all existing literature on a research question, but to find the best or most appropriate data to be inputted into the model. (20)

When conducting a rapid review to specify input parameters, it is necessary to first consider the likely importance of the input being searched. This will guide the intensity of the initial literature search, and therefore the time required. If an input parameter is likely to have minimal impact on the overall output of the model, then only a brief (<1 day) search of the literature may be necessary. In comparison, key inputs that may have a large impact on model output require a more thorough search (approximately 7 days). For inputs that are specified using the low intensity approach, the actual impact of the input can also be tested in a univariate sensitivity analysis, presented as a Tornado plot. This tests the sensitivity of the main outcome by varying each input by its uncertainty interval and can be used to determine the appropriateness of high vs. low intensity rapid reviews for each input. This will likely show that many of the input parameters do not generate too much uncertainty in the outputs (health-adjusted life years [HALYs], costs) – these inputs are appropriate to leave as is, defined by the low intensity rapid review process. Conversely, it is likely to identify a handful of input parameters with uncertainty that propagates through to the most uncertainty in the outputs. These are the input parameters that will be necessary to re-evaluate by conducting a more thorough rapid review. This univariate sensitivity analysis approach enables us to loop back to conduct a more thorough literature search one those handful of input parameters as necessary. Put another way, there is a strong case for *very* rapid literature searches initially, so long as one uses univariate sensitivity analyses and Tornado plots to loop back and be more thorough. Figure 2 presents these decisions in a flow-chart.

We now turn to the steps for a rapid literature search.

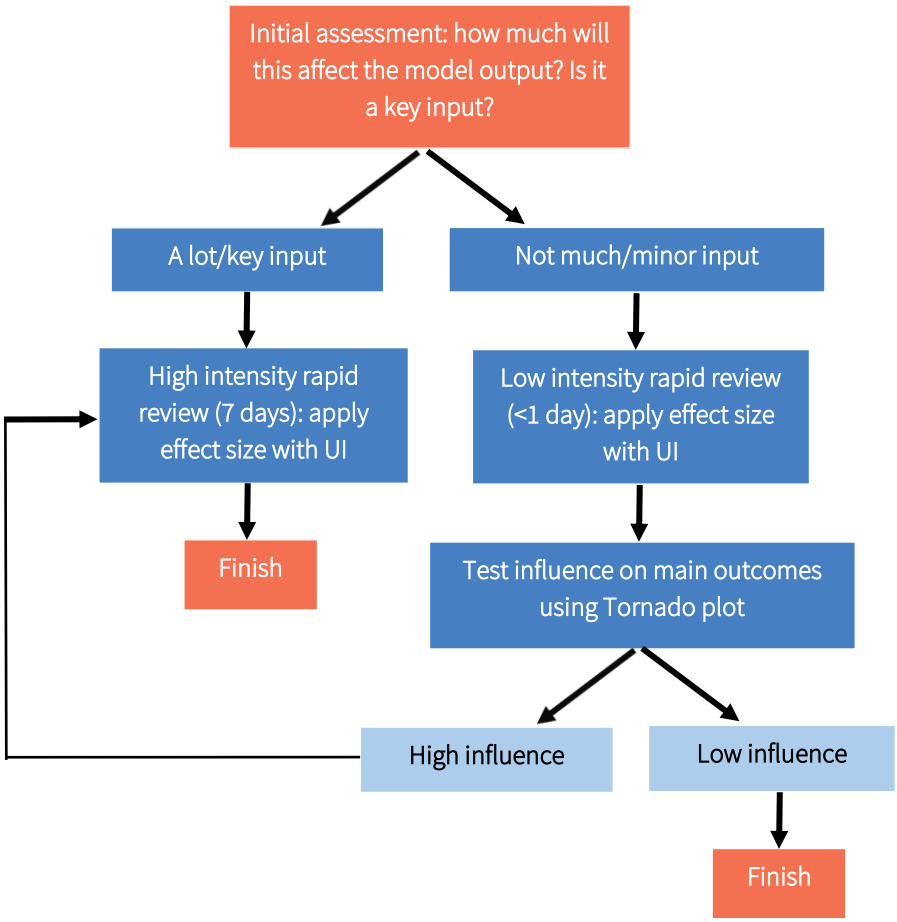

Figure 2. High vs. low intensity rapid review flowchart

UI = Uncertainty Interval.

RAPID REVIEW STEPS

1. DEFINE THE RESEARCH QUESTION

Research questions should be specific, and defined in PICOT format (Population, Intervention [or exposure], Comparator, Outcome, Time). Limiting the scope of the research question as much as appropriate can help to simplify the subsequent search of the literature. (18) If limited literature is available under these parameters, the scope of the search may then need to be widened and a second search undertaken.

2. DEFINE THE RESEARCH QUESTION

Key inclusion and exclusion criteria should be defined a priori and reflect the specificity of the research question. It is important to consider the population of interest, appropriate comparator groups, and outcomes assessed, when defining these criteria. However, it may also be necessary, especially when searching for evidence on relatively new interventions, to widen inclusion criteria to other populations, and include parallel evidence (discussed further below).

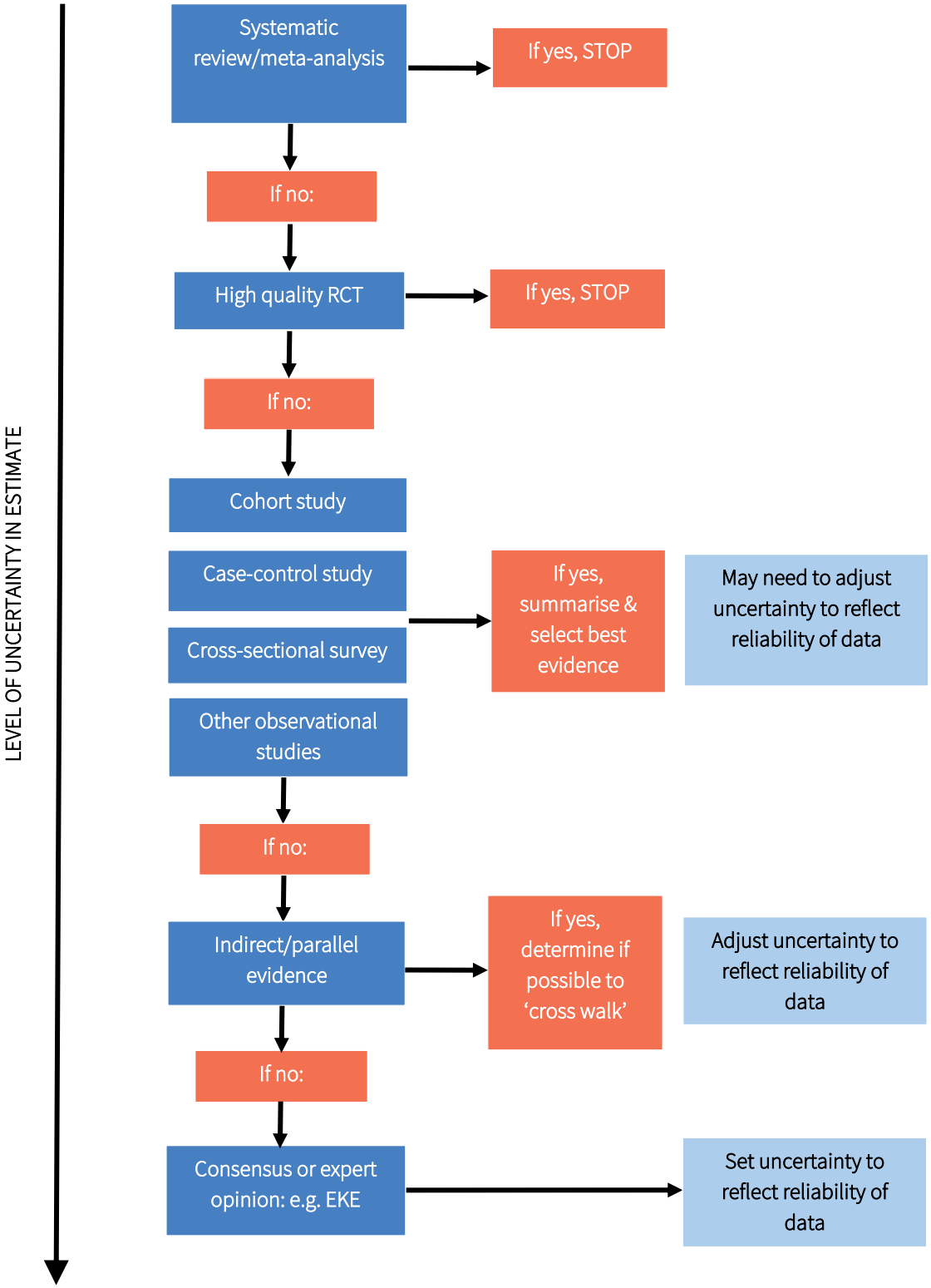

Figure 3. SHINE rapid review literature search process

EKE= Expert Knowledge Elicitation; RCT = Randomised Controlled Trial.

3. CONDUCT A RAPID SEARCH OF THE LITERATURE

A rapid review of the literature, while specific, should be conducted systematically and include documentation of the process undertaken (detailed later).

The search itself, and importantly the end or stopping point, should follow the guide of Figure 3, and is discussed further below. Key steps where there are differences for low vs. high intensity rapid reviews are highlighted.

Search for existing systematic reviews

According to the traditional hierarchy of evidence (Figure 4), systematic reviews hold the highest validity. (21) Key systematic review databases to search include Cochrane Library, Medline (via PubMed or Ovid), Scopus, or Web of Science.

Low intensity rapid review: 1 database

High intensity rapid review: 2-3 databases

Key questions to ask at this point include:

- Is a relevant population being considered?
- How recently was the systematic review conducted? Is there likely to be more up-to-date evidence available?
- Is the review examining an outcome relevant to the research question/model?
- Does the meta-analysis have significant heterogeneity? (20)

If an effect size and confidence interval can be extracted from an appropriate systematic review, the rapid review can be stopped. If a systematic review is not available or is inappropriate, a rapid search of primary studies should be conducted.

Search for primary studies

One or two key databases should be selected for this process, for example Medline (PubMed), and Web of Science. Specific databases exist for different areas, such as National Health Service (NHS) Economic Evaluation Database.

A formal quality and risk of bias assessment may be too time consuming. A rapid assessment should be conducted, considering the following factors of appropriateness and quality of studies reviewed:

- Hierarchy of evidence, as shown in Figure 4. However, this traditional view may not always indicate the most appropriate data source: e.g., costs of interventions may be best sourced from surveys. (21)
- Sample size
- Risk of bias: is there likely a high risk of bias towards or away from the null? (See Appendix Table 2 for a list of key biases, and other important considerations, in different study types)
- Population of interest: is the study undertaken in the relevant setting?
- Applicability of the intervention to the model
- Study inclusion/exclusion criteria: are the study participants representative of the relevant population?
- Are results stratified by relevant factors (e.g., age, ethnicity)

Low intensity rapid review: quick assessment of above criteria

High intensity rapid review: in-depth assessment of above criteria (including use, if deemed necessary, of formal quality assessment criteria for a given study type; see appendix)

Figure 4: Traditional hierarchy of evidence

Obtained from WHO (21).

If multiple potential 'candidates' are obtained from part b of the rapid review, the above factors can be utilised to compare and assess the best data to be inputted into the model. If no suitable candidates can be found from this search it may be appropriate to widen the literature search or consider an expert knowledge elicitation approach.

Widening the literature search

Loosening the restrictions imposed by search terms and inclusion/exclusion criteria that were previously defined can widen the literature search. This allows for potential parallel or indirect evidence to be sourced. The use of parallel and indirect evidence, when higher quality evidence is not available, is based on definitions by Vos et al., as part of the Assessing Cost-Effectiveness in Prevention (ACE-Prevention) study, and previous related works by Haby et al., and Swinburn et al. (22-24)

Parallel evidence designates that which examines the intervention/exposure of interest on a related, but different, outcome. For example, using evidence from the effectiveness of a social marketing campaign on tobacco control to estimate the impact of a similar strategy on obesity prevention. (24) Indirect evidence is that which does not specifically look at the exposure-disease relationship of interest but indicates the positive or negative effect of the

exposure/intervention. (24) Swinburn et al. use the example of continued advertisement of particular food products being an indirect source of evidence that such marketing campaigns increase the purchasing of these foods. (24)

Evidence may also be obtained from other contexts by widening the search criteria, for example on the exposure-disease relationship of interest but conducted in a different population with notable differences to the population that is the subject of modelling, or on a specific group of people such as healthcare workers that are not representative of the wider population of interest.

By widening a search to capture indirect or parallel evidence, or transporting evidence from different contexts, an effect size for the population and outcome of interest can then be estimated, and if possible 'cross-walked' to the relevant setting (discussed further later).

4. DEFINE UNCERTAINTY

Four types of variability or uncertainty are defined in relation to modelling: stochastic uncertainty (micro-simulation only), model structure uncertainty (e.g. whether to treat diabetes as a risk factor for other diseases or not; e.g. how to model the polyp pathway to colorectal cancer; beyond the scope of this protocol), heterogeneity (not so much uncertainty, but true variation in the population that we may want to explicitly model, as described in part one), and input parameter uncertainty. The latter is the focus of this protocol (although rapid reviews can and should be used to determine model structure). It is important to specify an uncertainty distribution around input parameters, as the model can then be 'run' many thousands of times, for example using a Monte Carlo simulation, with each run randomly sampling a value for each input from a probability distribution defined by its uncertainty interval.

In primary studies and meta-analyses, random error is presented in the form of a confidence interval around an effect size. For SHINE modelling, the default position is to widen the confidence interval, to more generously specify uncertainty about intervention effect sizes. This is largely because studies do not explicitly adjust for likely residual confounding, misclassification or selection bias errors. Ideally, a quantitative bias analysis (QBA) would be conducted on all inputs to 'fold in' residual possible systematic error to the random error (i.e., that carried in the confidence interval) to give a total uncertainty interval. (25,26) In the absence of resources to conduct QBA on all inputs, a default to specify 'generous uncertainty' as below is used. This additional or 'generous' uncertainty is more important for observational studies with a lower quality of evidence, such as those with a moderate/high risk of systematic bias (i.e. one or more of confounding, measurement error or selection biases). A generous level of uncertainty also needs to be parameterised around those inputs that have been obtained from parallel/indirect evidence and cross-walked, transported from other context (time, country, type of people), or for those determined through consensus approaches (see below section for further detail).

While the confidence interval for effect sizes from relevant, high-quality systematic reviews and RCTs may be applied unchanged as measure of input parameter uncertainty, this is not always the case. It is important to consider that in meta-analyses, a common approach is to present a narrower confidence interval based on inverse variance pooling of estimates, whereas it may instead be more appropriate to specify a wider uncertainty interval *given* the heterogeneity observed across studies. This is a practice now being used in the GBD with their 'burden of proof' approach. (27)

The general rule for applying uncertainty to input parameters, developed in the BODE[3] program, and adopted by SHINE, is as follows:

- Low uncertainty: +/- 5% of the expected or median value of the input parameter as the standard deviation (SD) for the input's probability distribution
- Moderate uncertainty: +/- 10% SD
- High uncertainty: +/- 20% SD
- Very high-level uncertainty: +/- 30%-40% SD
- Transported evidence: double the uncertainty defined in study.

Note that with this schema, three characteristics of the input parameter need consideration. First, if a value less than (or sometimes greater than) zero is impossible, and one is positing large uncertainty (e.g., >20% SD), it may be appropriate to log or otherwise transform the variable so that a value of less than zero is impossible to draw. Second, and related, the reference value may be one (e.g., relative risks), requiring the analyst to consider if uncertainty applies to the difference to one or can cross one. If the former, then a beta distribution may be appropriate (as it samples values from 0-1). If the latter, a log-normal distribution for a ratio may be appropriate. Lastly, when drawing from the uncertainty distribution of a given parameter, it is necessary to decide whether to correlate this variable with the draws from other inputs – for example, in modelling mask effectiveness against COVID-19 infection, draws of respirator effectiveness (as an odds ratio) were correlated 100% with draws of surgical mask effectiveness (meaning that, if the 50th percentile was drawn for respirator effectiveness from its uncertainty distribution, then the 50th percentile was also drawn for surgical mask effectiveness from its uncertainty distribution). (1) It is important to consider correlations between input parameters at the outset (see Appendix Table 1 for examples).

DOCUMENTATION

Rapid reviews are often poorly documented. (19) A lack of documentation on the process of sourcing input parameters can undermine the validity of the model as a whole. It is therefore important to document the rapid review search strategy. A record of the literature retrieved from each database, when the search was undertaken, the number duplicates removed etc. should be kept, following the PRISMA flow diagram formula (noting that there will be differences to PRISMA flow diagram guidelines, which is for systematic reviews). (28)

If multiple records are obtained for a particular literature search, those meeting basic inclusion and exclusion criteria should also be documented systematically, highlighting key attributes, risks of bias, and the overall decision rule. An example table that can be used to document this process can be found in the Appendix (Appendix Table 3).

CROSS WALKING

There is no commonly used definition of 'cross-walking' in epidemiology to our knowledge. For the purposes of this SHINE protocol, we defined two definitions.

Firstly, 'One scale-to-another.' Here one has or develops a method of making comparable metrics across two or more studies that use different measures of the exposure, outcome, or important covariates. The most common application of cross-walking in the existing epidemiology/health economics literature is mapping between different quality of life metrics. For example, linking between different metrics used for depression-related quality of life, such as PHQ-9, BDI-II

and PROMIS Depression, has been undertaken by developing cross-walk tables, which can also be used to establish a standard metric of depression-related quality of life. (29,30) Similarly, Charlson et al. applied a cross-walk method to map depression prevalence data obtained using DSM criteria, to estimate prevalence that would have been obtained from the preferred ICD-10 criteria (31). Examples also exist for more general disease measurement, such as mapping quality of life metrics (e.g. SF-36 instrument) to GBD disability weight estimates. (32) Another example of this cross-walk is an approach used by Finucane et al. when estimating global prevalence of high body mass index (BMI). (33) Some studies used for high BMI estimation in this review reported on the prevalence of overweight or obesity, rather than the preferred study outcome. A regression model was developed by the authors to cross-walk these outcome measures to an estimate of BMI, to allow inclusion of studies reporting on overweight and obesity. (33)

Secondly, 'One context-to-another.' Here one may be using parallel evidence (see above), and positing a strength of association for the exposure/intervention of interest using studies on similar (but not identical) mechanisms, exposures, or outcomes. An example of this was applied in a recent project estimating the burden of disease attributable to long COVID in Australia. (34) At the time, detailed long COVID symptom data was not available among the population of interest for this study, being Omicron-infected and vaccinated cases. Therefore, available data on the incidence of specific long COVID symptoms among those infected with pre-Omicron variants of SARS-CoV-2 were adjusted using separate estimates of the odds of any long COVID symptom among Omicron cases compared to pre-Omicron causes, and the odds of any long COVID symptom among vaccinated compared to unvaccinated individuals. Adjusting 'base case' estimates of symptom prevalence allowed for an indirect estimate of the incidence of these symptoms among the cases of interest (Omicron-infected, vaccinated cases).

EXPERT KNOWLEDGE ELICITATION

Expert knowledge elicitation (EKE) is a scientific method that allows quantification of unknown statistical parameters, using expert judgement. (35) Multiple formal protocols exist for eliciting information in an unbiased manner; therefore, we do not detail these methods here. (35) This is a time-consuming process and is therefore should only be used in SHINE modelling for model inputs that have a high impact on measured outputs. An example of EKE is presented in the supplementary material of recently published SHINE tobacco modelling paper in A/NZ. (2)

REFERENCES

1. Szanyi J, Wilson T, Howe S et al. Epidemiological and economic modelling of optimal COVID-19 policy: public health and social measures, masks and vaccines in Victoria, Australia. Lancet Reg Health West Pac. 2023;32:100675. doi:10.1016/j.lanwpc.2022.100675.
2. Ait Ouakrim D, Wilson T, Waa A et al. Tobacco endgame intervention impacts on health gains and Māori: non-Māori health inequity: a simulation study of the Aotearoa-New Zealand Tobacco Action Plan. Tob Control. 2023:tc-2022-057655. doi:10.1136/tc-2022-057655.
3. Singh A, Mizdrak A, Daniel L et al. Estimating cardiovascular health gains from eradicating indoor cold in Australia. Environ Health. 2022;21(54). doi:10.1186/s12940-022-00865-9.
4. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 2005.
5. GBD 2019 Demographics Collaborators. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. Lancet. 2020 Oct 17;396(10258):1160-1203. doi:10.1016/S0140-6736(20)30977-6.
6. Australian Bureau of Statistics (ABS). Life Tables (reference period 2019-2021) [Internet]. Canberra: ABS; 2022 [cited 2022 Nov 10]. Available from: https://www.abs.gov.au/statistics/people/population/life-tables/latest-release.
7. Institute for Health Metrics and Evaluation (IHME). GBD results [Internet]. Seattle, WA: IHME; 2022 [cited 2022 Nov 10]. Available from: https://vizhub.healthdata.org/gbd-results/.
8. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Disability Weights [Internet]. Seattle, USA: Institute for Health Metrics and Evaluation (IHME); 2020 [cited 2022 Nov 11]. Available from: https://ghdx.healthdata.org/record/ihme-data/gbd-2019-disability-weights.
9. Australian Institute of Health and Welfare (AIHW). Australian Burden of Disease Study: Impact and causes of illness and death in Australia 2018. Canberra: AIHW; 2021.
10. Australian Institute of Health and Welfare (AIHW). National Drug Strategy Household Survey 2019. Canberra: AIHW, Australian Government; 2020. doi:10.25816/e42p-a447.
11. Global Burden of Disease 2019 Risk Factor Collaborators. Globan burden of 87 risk factors in 204 countries and territores, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet. 2020;396(10258):1223-1249. doi:10.1016/S0140-6736(20)30752-2.
12. National Hospital Cost Data Collection (NHCDC) Public Hospitalis Report – Round 26 (financial year 2021-22). Sydney: IHACPA; 2022 Jul 30 [cited 2022 Nov 12]. Available from:

https://www.ihacpa.gov.au/resources/national-hospital-cost-data-collection-nhcdc-public-hospitals-report-round-26-financial-year-2021-22.

13. Blakely T, Sigglekow F, Irfan M et al. Disease-related income and economic productivity loss in New Zealand: a longitudinal analysis of individual-level data. PLoS Med. 2021;18(11):e1003848. doi:10.1371/journal.pmed.1003848.
14. Blakely T, Atkinson J, Kvihinadze G et al. Updated New Zealand health system cost estimates from health events by sex, age and proximity to death: further improvements in the age of 'big data'. NZMJ. 2015;128(1422):13-18.
15. OECD. Purchasing power parities (PPP) (indicator) [Internet]. OECD; 2022 [cited 2022 Nov 9]. doi: 10.1787/1290ee5a-en. Available from: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm#:~:text=Purchasing%20power%20parities%20(PPPs)%20are,in%20price%20levels%20between%20countries.
16. Australian Bureau of Statistics (ABS). Socio-economic indexes for areas [Internet]. Canberra: ABS; [last updated 2022 May 6; cited 2022 Nov 10]. Available from: https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa.
17. Stats New Zealand (NZ). Integrated Data Infrastructure [Internet]. Wellington: Stats NZ, New Zealand Government; [last updated 2022 Aug 23; cited 2022 Nov 8]. Available from: https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/#:~:text=The%20Integrated%20Data%20Infrastructure%20(IDI,migration%2C%20justice%2C%20and%20health.
18. Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. Designing a rapid response program to support evidence-informed decision-making in the Americas region: using the best available evidence and case studies. Implementation Sci. 2015;11(117). doi:10.1186/s13012-016-0472-9.
19. Tricco AC, Antony J, Zarin W et al. A scoping review of rapid review methods. BMC Med. 2015;13(224). doi:10.1186/s12916-015-0465-6.
20. Foster R. Literature search and data synthesis methods for estimating inputs for health economics modelling – Burden of Disease Epidemiology, Equity and Cost-Effectiveness Program (BODE^3), Technical Report: Number 7. [Internet] Wellington: Department of Public Health, University of Otago. 2011 [last updated 15 Feb 2013; cited 2022 Oct 27]. Available from: https://www.otago.ac.nz/wellington/otago042138.pdf.
21. World Health Organization (WHO). Evidence, policy, impact: WHO guide for evidence-informed decision-making [Internet]. Geneva: WHO. 2021 [cited 2022 Oct 27]. Available from: https://apps.who.int/iris/handle/10665/350994.

22. Vos T, Carter R, Barendregt J et al. Assessing Cost-Effectiveness in Prevention (ACE-Prevention): final report. 2010. University of Queensland, Brisbane and Deakin University, Melbourne.
23. Haby MM, Vos T, Carter R et al. A new approach to assessing the health benefit from obesity interventions in children and adolescents: the assessing cost-effectiveness in obesity project. Int J Obes (Lond). 2006 Oct;30(10):1463-75. doi: 10.1038/sj.ijo.0803469.
24. Swinburn B, Gill T, Kumanyika S. Obesity prevention: a proposed framework for translating evidence into action. Obes Rev. 2005;6(1):23-33. doi:10.1111/j.1467-789X.2005.00184.x.
25. Lash TL, Fox MP, Fink A. Applying quantitative bias analysis to epidemiological data. New York: Springer; 2009.
26. Lash TL, Fox MP, LacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. Int J Epi. 2014;43(6):1969-1985.
27. Zheng P, Afshin A, Biryukov S et al. The Burden of Proof studies: assessing the evidence of risk. Nat Med. 2022;28(10):2038-2044. doi:10.1038/s41591-022-01973-2.
28. Page MJ, McKenzie JE, Bossuyt PM et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.
29. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol Assess. 2014;26(2):513-527. doi:10.1037/a0035768.
30. Dorans NJ. Linking scores from multiple health outcome instruments. Qual Life Res. 2007;16 Suppl 1:85-94. doi:10.1007/s11136-006-9155-3.
31. Charlson FJ, Ferrari AJ, Flaxman AD, Whiteford HA. The epidemiological modelling of dysthymia: application for the Global Burden of Disease Study 2010. J Affect Disord. 2013;151(1):111-120. doi:10.1016/j.jad.2013.05.060.
32. Vos T, Flaxman AD, Naghavi M et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the global burden of disease study 2010. Lancet. 2012;380(9859):2163-2196). doi:10.1016/S0140-6736(12)61729-2.
33. Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. Lancet. 2011;377(9765):557-567.
34. Howe S, Szanyi J, Blakely T. The health impact of long COVID during the 2021-2022 Omicron wave in Australia: quantitative burden of disease study. MedRxiv. 2022. doi:10.1101/2022.08.01.22278219.
35. O'Hagan A. Expert knowledge elicitation: subjective but scientific. Am Stat. 2019;73:69-81. doi:10.1080/00031305.2018.1518265.

36. Sterne JAC, Savović J, Page MJ et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ .2019; 366: l4898. doi: 10.1136/bmj.l4898.
37. Sterne JAC, Hernán MA, Reeves BC et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. BMJ 2016; 355; i4919; doi: 10.1136/bmj.i4919.
38. Guyatt G, Oxman AD, Akl EA et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of clinical epidemiology. 2011;64(4):383-94.
39. Shea BJ, Reeves BC, Wells G et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017 Sep 21;358:j4008.

APPENDICES

Appendix Table 1: input parameter template

Variable	Source	Format	Correlations (if applicable)	Future trends (if applicable)	Other considerations
Demographics					
E.g. Population of Australia	ABS Population dataset 2022	Population number by age, sex and SES	N/A	N/A (modelling cohort alive in 2022)	
BAU inputs					
E.g. All cause mortality	ABS		N/A	Apply annual percentage change from 1990-2019, for the first 15 years of model run, then hold constant	
E.g. CHD morbidity	GBD data (via IHME GHDx)	YLD by age (5-year age groups) and sex	N/A	N/A (not applying time trend)	
Intervention inputs					
E.g. COVID-19 vaccine effectiveness (VE)	Rapid review: high intensity	Regression equation coefficients OR Peak VE (as an odds ratio) with waning rate	Correlate values of VE against all SARS-CoV-2 variants active in a model run Correlate values of VE against all measured outcomes (e.g. VE against infection and death)	N/A (model only running for 18 months, therefore not applying future trends)	• Separate VE for different available vaccines, or single VE estimate?
E.g. Denicotinisation of cigarettes	Rapid review: high intensity	Change in rate of smoking initiation and cessation			• Specify changes for vaping also • Apply differences by sex/ age/ ethnicity

ABS: Australian Bureau of Statistics; BAU = Business As Usual; SES = Socio-Economic Status.

Appendix Table 2: evidence evaluation

Type of study	Quality/appropriateness considerations	Bias/confounding considerations	Formal quality assessment tools*
Systematic review	• Is the intervention/ exposure the same as that being modelled? • What selection criteria are being used? How relevant are these criteria to the population being modelled? • Consider heterogeneity (I^2)	Assessment of quality/bias for individual studies should be included within the review: • When reading a systematic review, consider if and how authors have assessed bias in individual studies Publication bias: • Evaluate funnel plot, if presented	Assessing individual studies: RoB 2 [36]/ROBINS-I [37] Assessing the review as a whole: GRADE [38]
Randomised controlled trial (RCT)	• Is the intervention/exposure the same as that being modelled? • Is the study population similar enough to that being modelled? • Does the study have a sufficient sample size (power) • Is the study sample appropriately stratified? • What is the comparator/control?	Selection bias: • Loss to follow-up/attrition • Differential non-adherence to study protocol between trial arms (particularly if study uses 'per-protocol analysis') Information bias: • Differential measurement error (e.g. if participants/assessors are not blinded to treatment allocation) • Non-differential measurement error (e.g. due to poor/inappropriate outcome measurement tools) Confounding: • Important confounders not measured • Measured confounders differ at baseline between trial arms (e.g. due to poor randomisation protocol) and are not adjusted for	AMSTAR 2 [39] RoB 2 [36]
Cohort study	• Is the intervention/exposure the same as that being modelled? • Is the study population similar enough to that being modelled? • Does the study have a sufficient sample size (power) • Is the study sample appropriately stratified? • What is the comparator/control?	Selection bias: • Inappropriate comparison group • Non-participation at baseline • Loss to follow-up/attrition Information bias: • Differential measurement error, of the exposure/outcome (i.e. the outcome influences the misclassification of the exposure, or vice versa) • Non-differential measurement error, of the	ROBINS-I [37]

		exposure/outcome (e.g. due to recall bias) Confounding: • Unmeasured/measured confounders not adjusted for	
Case-control study	• Is the intervention/exposure the same as that being modelled? • Is the study population similar enough to that being modelled? • Does the study have a sufficient sample size (power) • Is the study sample appropriately stratified? • What is the comparator/control?	Selection bias: • Inappropriate comparison group • Non-participation at baseline • Loss to follow-up/attrition Information bias: • Differential measurement error, of the exposure/outcome (i.e. the outcome influences the misclassification of the exposure, or vice versa) • Non-differential measurement error, of the exposure/outcome (e.g. due to recall bias) Confounding: • Unmeasured/measured confounders not adjusted for	ROBINS-I [37]
Economic evaluation/costing study	• Is the intervention/exposure the same as that being modelled? • Is the study population similar enough to that being modelled? • Does the study have a sufficient sample size (power) • Is the study sample appropriately stratified? What is the comparator/control?	Economic evaluation alongside RCT: • See above sources of bias/confounding	Drummond checklist [4]

*While SHINE rapid review time constraints generally do not allow for formal quality assessments, the relevant tools can provide a guide as to key sources of potential bias, and study quality.

Appendix Table 3: example of table format for rapid study assessment

Study (year)	Study design	Population	Sample size	Intervention/ exposure	Outcomes measured	Potential biases	Decision rule
Author A et al. (2022)*	Cross sectional study	Adults (18+) in Australia	400 cases 400 controls	PCR-confirmed COVID-19	Any of three defined long COVID symptoms: shortness of breath, fatigue, palpitations	Selection bias: • COVID+ individuals with ongoing symptoms may be more likely to participate • Drop-out high among cases (>40%) Information bias: • Cases aware of COVID-19 diagnosis (differential misclassificati on of the outcome possible)	Exclude: high risk bias

*Example only

SHINE Technical Report Series, No: 1