Estimating Transition Probabilities from Published Evidence: A Tutorial for Decision Modelers

Risha Gidwani[1,2,3] · Louise B. Russell[4,5]

Abstract

This tutorial presents practical guidance on transforming various typesof information published in journals, or available online from government and other sources, into transition probabilities for use in state-transition models, including cost-effectiveness models. Much, but not all, of the guidance has been previously published in peer-reviewed journals. Our purpose is to collect it in one location to serve as a stand-alone resource for decision modelers who draw most or all of their information from the published literature. Our focus is on the technical aspects of manipulating data to derive transition probabilities. We explain how to derive model transition probabilities from the following types of statistics: relative risks, odds, odds ratios, and rates. We then review the well-known approach for converting probabilities to match the model's cycle length when there are two health-state transitions and how to handle the case of three or more health-state transitions, for which the two-state approach is not appropriate. Other topics discussed include transition probabilities for population subgroups, issues to keep in mind when using data from different sources in the derivation process, and sensitivity analyses, including the use of sensitivity analysis to allocate analyst effort in refining transition probabilities and ways to handle sources of uncertainty that are not routinely formalized in models. The paper concludes with recommendations to help modelers make the best use of the published literature.

1 Introduction

A set of health states, or events, and the probabilities of transitioning from onestate to others during a specified period of time ("transition probabilities") are the fundamental building blocks of decision models. A state-transition model to evaluate cancer interventions, for example, might start with the Cancer-Free state and proceed through Local, Regional, and Metastatic disease to Death. Transition probabilities would describe the probabilities of moving from Cancer-Free to Local Cancer, from Local to Regional, from Regional to Metastatic, and from any of those states to Death, over, say, 1 year. Different probabilities would be needed to describe the natural (untreated) course of the disease versus its course with treatment. The yearly time period, called the cycle length, would repeat until an appropriate stopping point was reached. The total number of years (cycles) represents the model's time horizon. The usual recommendation is that the time horizon should be long enough to capture all significant health outcomes, which often requires modeling the remaining lifetimes of patients [1, 2].

There are two common challenges a modeler faces when deriving transition probabilities for use in a decision model. One challenge is that the data from the published and publicly available literature, such as data published by the United States (US) Census Bureau or the Centers for Disease Control and Prevention, are often not reported as probabilities. Rather, evidence that is relevant to the decision model may be in the form of counts, rates, relative risks (RRs), or odds ratios (ORs) that need to be converted into probabilities. A second challenge is that when the evidence

✉ Risha Gidwani
  rishag@rand.org

1 Department of Health Management and Policy, UCLA School of Public Health, Los Angeles, CA, USA

2 Health Economics Resource Center, VA Palo Alto Health Care System, Menlo Park, CA, USA

3 Center for Innovation To Implementation, VA Palo Alto Health Care System, Menlo Park, CA, USA

4 Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA

5 Center for Health Incentives and Behavioral Economics and Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, USA

is expressed as probabilities, the published probabilities will often not match the cycle length of the decision model. For example, annual probability data may be published, while the decision model has a 3-month cycle length.

This tutorial grew out of a popular VA Health Economics Resource Center (HERC) cybercourse, which generated many requests for the slides and suggestions that they be written up. As an example, during the first week the World Health Organization defined the COVID-19 crisis as a pandemic, modelers from a national US agency read the slides and reached out to us for guidance in transforming model probability inputs. Responding to those requests, this tutorial presents practical guidance for transforming published estimates into appropriate transition probabilities. Much of the guidance is already available in peer-reviewed journals. Our purpose here is to collect it in one location to serve as a stand-alone resource for the decision modeler. Our intended audience is decision modelers who find most or all of their information in the published literature. The principles presented to transform summary statistics into probabilities apply more widely, but for simplicity, we talk about state-transition models, often called Markov models. We focus on the technical aspects of manipulating published data to derive transition probabilities and touch only lightly on how to select the most appropriate evidence.

The paper begins by outlining the types of evidence available in the published literature. We first discuss how to derive transition probabilities from common types of summary statistics, such as RRs, odds, and ORs, and issues to keep in mind when using data from different sources in the derivation process. We then discuss how to derive transition probabilities from probabilities whose time frames do not match the model's cycle length. We explain the well-known approach for deriving transition probabilities when the model, or model node, has only two state transitions, and discuss how to handle three or more possible transitions, for which the two-state formulas are not appropriate. Lastly, we discuss how to use sensitivity analyses to allocate analyst effort to refine probabilities and ways to handle sources of uncertainty that are not routinely formalized in models. The paper concludes with recommendations to help the modeler make the best use of the published literature.

2 The Published Evidence

The International Society for Pharmacoeconomics and Outcomes Research (ISPOR)—Society for Medical Decision Making (SMDM) Modeling Task Force recommends that "[t]ransition probabilities and intervention effects should be derived from the most representative data sources for the decision problem" [3]. Although modelers may find all the required data in a single published report, more commonly the data come from multiple studies. Some of it may involve samples that are small, unrepresentative, or both. Some studies may be 20 or 30 years old. The task force recommends conducting a systematic literature review when multiple studies are available to inform the same parameter(s). The systematic review can be used to select the study that best fits the model, or it could be used as the first step in a meta-analysis producing a quantitative pooled estimate of the individual study-level outcomes [4].

The task force suggests that transition probabilities for the natural history of a condition, sometimes called the "Do Nothing" model arm, are best drawn from population-based epidemiological studies. Studies with longer follow-up times are preferred since they allow realistic modeling of the disease for a larger portion of the model's time horizon. In circumstances where the model contains interventions from all arms of a single randomized controlled trial (RCT), the trial's control arm can be used to represent natural history, but that is less desirable because trial participants are often selected using criteria that make them unrepresentative of the population to which an intervention will be applied in practice.

For the model's intervention arms, RCTs represent the highest-quality evidence of efficacy, since properly conducted randomization balances measured and unmeasured confounders across treatment and control groups. The generalizability problem, however, arises here as well. RCTs may not represent effectiveness in real-world practice for at least two, possibly offsetting, reasons: (1) trialists work hard to maintain the quality and consistency of an intervention and to keep patient adherence high, while compliance in actual practice

Table 1 Common forms of published data and their definitions

Statistic	Evaluates	Range
Probability/risk	$\dfrac{\text{\#of events that occurred in a time period}}{\text{\#of people followed for that time period}}$	0–1
Rate	$\dfrac{\text{\#of events that occurred in a time period}}{\text{Total time period experienced by all subjects followed}}$	0 to ∞
Relative risk	$\dfrac{\text{Probability of outcome in exposed}}{\text{Probability of outcome in unexposed}}$	0 to ∞
Odds	$\dfrac{\text{Probability of outcome}}{1 - \text{Probability of outcome}}$	0 to ∞
Odds ratio	$\dfrac{\text{Odds of outcome in exposed}}{\text{Odds of outcome in unexposed}}$	0 to ∞

may be lower, reducing the intervention's effectiveness; and (2) control groups may benefit from the placebo effect of participating in a trial, raising the possibility that the intervention will be more effective in practice than in the trial.

In a model that compares interventions with each other, but has no "Do Nothing" or placebo arm, model data should maintain randomization within RCTs. For example, suppose the model compares interventions A, B, and C and there are separate RCTs providing treatment efficacy data for each intervention. If each RCT compares active treatment to placebo, using the relative treatment effect of A versus placebo from the first RCT when populating efficacy for A, the relative treatment effect of B versus placebo from the second RCT when populating efficacy for B, and so on, maintains randomization within trials. This approach is still subject to some bias, as patients have not been randomized *across* studies.

The best approach to maintain randomization within RCTs is to conduct a network meta-analysis (also sometimes referred to as an indirect treatment comparisons analysis) to derive appropriate transition probabilities. A network meta-analysis can generate the relative treatment effects of two or more interventions, such as A, B and C, when the interventions have not been directly compared in a single RCT: A versus B, A versus C, and B versus C. If conducted in a Bayesian framework, it provides probabilities for direct use in a model. Network meta-analytic estimates are less subject to selection bias than estimates from observational data because relative treatment effects are calculated within RCTs prior to the comparison across treatments; again, within-trial randomization is maintained. For more information about network meta-analyses, see the following excellent introductory articles: Sutton et al. [5], Jansen et al. [6], Jansen et al. [7], Hoaglin et al. [8], and Welton et al. [9].

Table 1 shows the most common ways of reporting data from single studies and their definitions. The evidence may be published as probabilities, but need to be converted to the model's cycle length. Or it may be published in another form that can be used to derive transition probabilities—RRs, ORs, and the like—in which case the modeler needs to know best practices for deriving probabilities and be aware of pitfalls and limitations. Note that some valuable information that is available when all the data come from a single study and individual-level data are available—such as correlations between probabilities—is not available when the data come from different studies.

3 Model Time Horizon and Cycle Length

The modeler must choose an appropriate cycle length for the model. Shorter cycles yield more accurate estimates of life expectancy and costs, but at higher computational cost. When the required cycle length changes over the disease/intervention pathway, say because events can occur quickly in the first month after diagnosis or surgery but are much less frequent over the longer term, the model may begin with a decision tree, to represent events that can occur within days or weeks, and shift to Markov nodes based on a cycle length of, for example, 1 year to represent longer-term events [2].

The discount rate recommended by the Second Panel on Cost-Effectiveness in Health and Medicine is 3% per year [10]. When the model does not have an annual cycle length, the discount rate must be modified using the formula $(1 + \text{annual rate})^{1/t}$, where t is the number of model cycles in a year, so that it remains 3% across a 1-year period [11]. If the model has a 1-month cycle length, for example, and a 3% annual discount rate, the formula yields a monthly discount rate of 0.247% [derived from $(1 + 0.03)^{1/12}$].

Often, the model's time horizon exceeds the follow-up time of published studies. For example, a clinical trial evaluating drug efficacy may have a follow-up period of only 2 years, while the model has a 30-year time horizon. There is a large amount of literature on the many methods available for extrapolating beyond the available data and a small amount of literature on the success of such extrapolations [12–16]. For modelers working with publicly available

data, it may be possible to extrapolate using life tables, or mortality rates by cause, available from national vital statistics systems. An important issue for such extrapolations, to which we return in the section on uncertainty analysis, is the need to increase the variance around the model's estimates to include extrapolation-based error. In addition, in models with long time horizons, transition probabilities, including probabilities of adverse events, which are too often omitted, usually need to be modified as the cohort ages.

4 Relative Risks and Odds Ratios

As noted, published information about disease burden and treatment efficacy is often summarized in the form of RRs or ORs. Here we discuss how to derive transition probabilities from these statistics.

4.1 Using Relative Risks to Derive Transition Probabilities for the Treated Group

Investigators often compare the probability of an event in people exposed to an intervention or condition to the probability in those not exposed—the probability of lung cancer in smokers versus nonsmokers, for example, or the probability of heart attacks in those who take statins versus those who do not. A relative risk (RR) (also called a risk ratio) is the ratio formed by the probability of the event in the exposed group divided by the probability of that same event in the unexposed group:

$$RR = \frac{p_1}{p_0}, \tag{1}$$

where p_1 is the probability of the event in exposed persons, and p_0 is the probability of the event in unexposed persons.

When the RR is multiplied by the probability of the event in unexposed persons, p_0, the denominator of the RR cancels out, leaving the probability of the event in the exposed, p_1:

$$p_1 = RR \times p_0 = \left(\frac{p_1}{p_0}\right) \times p_0. \tag{2}$$

Use of Eq. 2 necessitates knowing p_0. Often, the probability of the event in the unexposed is reported in the same article that reports the RR. In other cases, this information will come from a different source. For example, the probability, p_0, that an untreated diabetic person develops diabetic retinopathy may come from one source (such as the control arm of an RCT) and the RR of diabetic retinopathy for treatment versus no treatment from another, such as an epidemiologic study. The modeler needs to decide whether p_0 and the RR come from sufficiently similar populations (or whether there is reason to believe the RR is similar in all populations) for

the resulting modeler-derived estimate of p_1 to be valid and applicable to the population being modeled.

Of note, when the RR reported in the study has been adjusted for covariates and the probability of the event in the unexposed group has not, the denominator of the RR does not cancel out:

$$p_1 \approx \left(\frac{p_{1\_adjusted}}{p_{0\_adjusted}}\right) \times p_{0\_unadjusted}. \tag{3}$$

The modeler may, for lack of other data, use Eq. 3 anyway, recognizing that there is an unknown degree of error in the resulting estimate of p_1. A plausible range of values for p_1 should be tested in sensitivity analyses to determine the likely importance of this error to the analysis.

4.2 Using Relative Risks to Derive Subgroup Transition Probabilities

Probabilities are often available for a population, but not for subgroups that are important for the model. RRs can help in this situation because a population probability is a weighted average of subgroup probabilities and RRs provide the weights. To illustrate, an analysis of the cost-effectiveness of maternal immunization to prevent pertussis in infants, which compared maternal immunization plus routine infant vaccination with routine infant vaccination alone, required probabilities of pertussis death in infants by vaccination status and age group [17]. Probabilities of pertussis death for infants aged 0–1, 2–3, 4–5, 6–8, and 9–11 months were available from Brazilian mortality and hospitalization data systems. Probabilities that infants in each of those age groups had received no, one, or two to three doses of vaccine were modeled from survey data [18]. But probabilities of pertussis death by vaccination status, needed to estimate the impact of vaccination on pertussis mortality in infants, were not available.

To derive probabilities by vaccination status, the overall probability of pertussis death in each age group was expressed as a weighted average of the probabilities of death by vaccination status:

$$pm_{total} = (p0 \times pm_0) + (p1 \times pm_1) + (p23 \times pm_{23}), \tag{4}$$

pm_{total} is the known probability of dying of pertussis for the age group as a whole; pm_0, pm_1, and pm_{23} are the unknown probabilities of dying of pertussis for infants who received no dose, one dose, or two to three doses of vaccine. p_0, p_1, and p_{23} are the known proportions of children of that age who had received no dose, one dose, or two to three doses of vaccine.

Multiplying the right-hand side by pm_0/pm_0 allows the equation to be restated in terms of the RRs of death by vaccination status:

$$pm_{total} = pm_0 \times \left(p0 + \left(p1 \times \left(\frac{pm_1}{pm_0} \right) \right) + \left(p23 \times \left(\frac{pm_{23}}{pm_0} \right) \right) \right),$$

or,

$$pm_{total} = pm_0 \times \left(p0 + \left(p1 \times RR_1 \right) + \left(p23 \times RR_{23} \right) \right),$$

RRs were available from Juretzko et al. [19] (albeit for acellular pertussis vaccine, not the whole-cell vaccine used in Brazil), and p_0, p_1, and p_{23} were obtained from Clark using the methods described in Clark et al. [18], so the equation could be solved for pm_0. Once pm_0 is known, the RR equations can be used to solve for pm_1 and pm_{23},

$$RR_1 = 0.32 \quad \text{and} \quad RR_{23} = 0.05.$$

Then, employing Eq. 2:

$$pm_1 = 0.32 \times pm_0,$$

and,

$$pm_{23} = 0.05 \times pm_0.$$

Another example is given in Black et al. [20], where the method is used to derive probabilities of survival for non-smokers, former smokers, and current smokers from the 2009 mortality tables published by the US National Center for Health Statistics.

Combining evidence from studies that investigated different populations, at different times, and/or under different conditions may produce unrealistic results, a problem that is not always easy to detect. As an example, for an analysis of smoking cessation, survival probabilities for smokers, by sex, ethnicity, and age, were derived from US life tables. Survival probabilities for quitters were initially estimated by applying quitter/smoker survival ratios by age from another study [21] to the probabilities for smokers, but the resulting estimates exceeded 1.0 at some ages in some sex–ethnicity groups.

4.3 Deriving Relative Risks from Odds Ratios

The odds of an event, defined as the probability of the event, p, divided by 1 minus the probability of that event, can also be used to derive a probability:

$$Odds = \frac{p}{(1-p)}. \tag{5}$$

For example, the odds that a mother reported post-partum depression after a live birth in the USA were 0.13 in 2012. Thus, the probability of reporting post-partum depression was 0.115 [22].

Odds are not commonly reported, but are the basis for a frequently encountered summary statistic, the OR, because
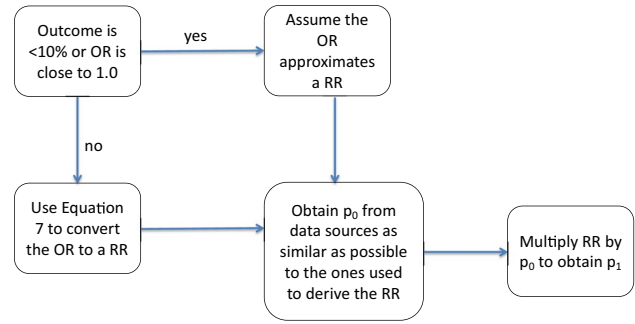

Fig. 1 Deriving a transition probability from a reported OR. *OR* odds ratio, p_0 probability of the event in unexposed persons, p_1 probability of the event in exposed persons, *RR* relative risk

the coefficients of logistic regressions are logged ORs, which can be exponentiated to get ORs. The OR is the odds of the event in one group, A, divided by the odds of the same event in another group, B:

$$Odds\ ratio = \frac{odds_A}{odds_B}. \tag{6}$$

ORs can be converted into probabilities using one of two methods (Fig. 1). If one of the outcomes is rare (< 10%) and/or the OR is close to 1.0, the OR is a reasonable approximation to the RR [23] and can be inserted directly into Eq. 2, in place of the RR, to derive the probability. To see how well an OR approximates an RR, readers are referred to Zhang and Yu [23] or Grant [24].

If the OR cannot be used to approximate the RR, the RR can be derived from the OR using the following equation [23, 24]:

$$RR = \frac{OR}{\left(1 - p_0 + \left(p_0 \times OR \right) \right)}, \tag{7}$$

where p_0 is the probability of the event in the unexposed group.

As the equation shows, the same OR produces different RRs depending on p_0, the probability of the event in the unexposed group. For example, an OR of 1.5 yields an RR of 1.429 when p_0 is 0.1, 1.250 when p_0 is 0.4, 1.154 when p_0 is 0.6, and 1.071 when p_0 is 0.8 [24]. Thus, as the probability of the event in the unexposed group increases, the OR becomes a poorer approximation of the RR.

When the OR comes from a multivariable logistic regression, as it often does, there is no single baseline risk. Within a single regression, the baseline risk depends on the values of the covariates, and there are as many baseline risks as there are combinations of covariate values; for example, one baseline risk for a smoker with low blood pressure and another baseline risk for a smoker with high blood pressure. Regressions based on the same dataset but with different

covariates will yield different baseline risks. And, of course, regressions that use different datasets will produce different baseline risks [25]. In principle, it is possible to calculate an average baseline risk for a specific regression by inserting the means of the covariates into the published logistic regression [24], or to calculate a baseline risk that best represents the model population by specifying appropriate values for the covariates. In practice, this is usually not possible because even when authors publish the coefficients for all covariates, they rarely include the value of the intercept, which is also needed.

If the modeler does not have access to the complete original regression, Grant recommends establishing a range of baseline risks and calculating the corresponding range of RRs, then conducting one-way sensitivity analysis to determine the influence of that range on the model's results [24]. A plausible range can be based on previous published research or on expert opinion.

5 Converting Probabilities to the Model's Cycle Length

Once the evidence is in the form of probabilities, it may need to be converted to the model's cycle length. For example, a trial may report outcomes at 2 years' follow-up, while the model has an annual cycle length. For a model node with only two branches, that is, two possible state transitions, the relationship between probabilities and rates provides a simple way to derive probabilities that match the model's cycle length. Recall that a probability is the number of events in a time period divided by the total number of people followed for that time period, and ranges from 0 to 1.0. A rate is the number of events divided by the total time at risk experienced by all people followed, and ranges from 0 to infinity. Thus, probabilities and rates for the same event are differentiated by their denominators: the calculation of a rate takes into account the time spent at risk, while the calculation of a probability does not [26]. See Appendix for a detailed example and the assumptions involved in the formula.

5.1 The Probability-Rate Equations when there are Two State Transitions

Equations 8 and 9 show the relation between a probability (p), rate (r), and time (t) [11, 26–28].

$$r = \frac{-\ln(1 - p)}{t}, \tag{8}$$

$$p = 1 - \exp(-rt). \tag{9}$$

To convert a probability from one time frame to another, the modeler can use Eqs. 8 and 9, which are the ones most frequently found in published articles, or the equivalent Eq. 10 [11].

$$p = 1 - (1 - p)^{1/n}. \tag{10}$$

The Appendix demonstrates both approaches for a hypothetical example.

As a real-world example, consider the 12-month probability, 10.8%, that a child under age 6 living in Milwaukee, Wisconsin is newly diagnosed with elevated blood lead levels (defined as ≥ 5 mcg/dL of blood) in 2016 [29]. If the model has a 3-month cycle length, a 3-month probability is needed. Using Eq. 8, we convert the 12-month probability to a 12-month rate. Since the time period does not change, the denominator is 1,

$$12 \text{ month rate} = \frac{-\ln(1 - 0.108)}{1} = 0.114289.$$

Next, using Eq. 9, we convert this 12-month rate to a 3-month probability,

$$3 \text{ month probability} = 1 - \exp\left(-0.114289 \times \frac{1}{4}\right)$$
$$= 1 - 0.97183 = 0.0282.$$

The 3-month probability is thus 0.0282 (alternatively, we could have converted the 12-month probability to a 3-month rate, and then the 3-month rate to a 3-month probability). Using Eq. 10 yields the same 3-month probability:

Table 2 Markov model of elevated lead levels

| | A | B | C | D |
	Children tested	Elevated lead levels	Cumulative elevated lead levels	Sum of persons in all health states by cycle (A+C)
Baseline	10,000.00	0	0	10,000.00
End of cycle 1	9718.32	281.68	281.68	10,000.00
End of cycle 2	9444.58	273.75	555.42	10,000.00
End of cycle 3	9178.54	266.03	821.46	10,000.00
End of cycle 4	8920.00	258.54	1080.00	10,000.00

Table 3 Example of the three-state problem

	Health states for 124 patients with severe congestive heart failure who were good candidates for heart transplant		
	Good candidate (gc)[b]	Transplant (t)	Dead (d)
Panel A: The transition probability matrix[a]			
Good candidate (gc)	$1-p_{gct}-p_{gcd}$	p_{gct}	p_{gcd}
Transplant (t)	0	$1-p_{td}$	p_{td}
Dead (d)	0	0	1
Panel B: Study data and (incorrect) transition probabilities derived by the two-state formula for the first row of the transition matrix (panel A), for the study cohort of 124 people			
Study outcomes at 3 years	9	92	23
3-year probability from study	0.0726	0.7419	0.1855
Annual probability, derived using the two-state formula	*c*	*0.3633*	*0.0661*

Projection end of year	Good candidates (A)	Transplants (B)	Cumulative transplants (C)	Dead (D)	Cumulative dead (E)	Sum of health states (A+C+E)
Panel C: Results of a Markov model using the annual two-state probabilities in panel B to project 3 years of outcomes for a cohort of 124 persons[d]						
1	70.7494	45.0543	45.0543	8.1963	8.1963	124
2	40.3667	25.7061	70.7604	4.6765	12.8728	124
3	**23.0316**	14.6669	**85.4273**	2.6682	**15.5411**	**124**
Correct numbers	**9**		**92**		**23**	**124**

Source: Anguita 1993 [31], Fig. 1

[a]The rows represent the state a person can transition "from"; the columns represent the state a person can transition "to"

[b]p_{gct} proportion of good candidates who receive a transplant during cycle t, p_{gcd} proportion of good candidates who die from all-cause mortality during cycle t, p_{td} proportion of transplanted patients who die during cycle t

[c]This probability is not derived from the formula. It is 1 minus the probabilities of transitioning to transplant or dead. See panel A, first row

[d]Applying annual transition probabilities derived by the two-state formula results in incorrect numbers experiencing health-state transitions, as shown by the differences between the last two rows: 23 projected good candidates vs. the correct 9, 85.4 transplants vs. 92, and 15.5 deaths vs. 23 (see the values formatted in bold)

$$\text{3 month probability} = 1 - (1 - \text{12 month } p)^{1/n}$$
$$= 1 - (1 - 0.108)^{1/4} = 1 - 0.892^{1/4}$$
$$= 1 - 0.97193 = 0.0282.$$

The 3-month rate can be verified by using it to calculate the probability that a child will be diagnosed over a year (Table 2, especially column C, end of cycle 4).

5.2 Changing Cycle Length When There are Three or More State Transitions

The conversion procedure for two state transitions (Eqs. 8–10) does not yield correct probabilities when three or more state transitions can occur in a cycle, a common situation [11, 26, 28, 30]. The problem is illustrated in Table 3, which is based on a study of patients with severe congestive heart failure evaluated for a heart transplant [31]. Panel A depicts the transition probability matrix of a Markov model. Among those considered good candidates for heart transplant and followed for 3 years, there are three possible transitions: remain a good candidate, receive a transplant, or die. The two-state formula will give incorrect annual transition probabilities for this row.

Panel B shows the study data applicable to the first row of the transition matrix; study outcomes, 3-year probabilities calculated from the study outcomes, and (incorrect) annual probabilities derived from the 3-year probabilities using the two-state method. Panel C shows the results of a Markov model that used the incorrect annual probabilities to project health outcomes for a baseline cohort of 124 patients (124 chosen to match the source study and facilitate comparisons). The correct numbers, calculated using the 3-year probabilities from the study, are shown in bold in the last row of panel C. The annual probabilities derived using the two-state method substantially overestimate the number of good candidates remaining at 3 years and underestimate the other two health states.

There are three possible solutions to the problem of deriving model transition probabilities when more than two
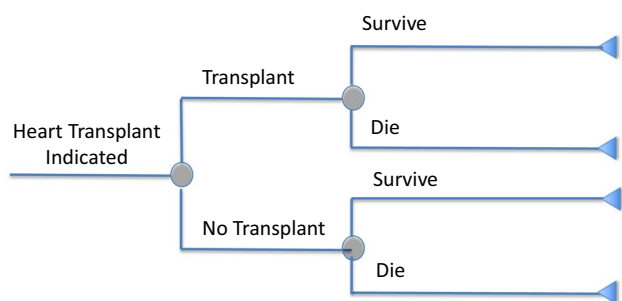
Fig. 2 Conditional nodes for decision models

transitions can occur within a cycle. The first is to revise the model structure so that each node has only two branches (two transitions). This would be easy to do in the heart transplant case if the diagnostic pathway led from "good candidate/heart transplant indicated" to transplant/no transplant, and then each of those branches led to survive/die (Fig. 2).

If the model cannot easily be reduced to a series of two-state branches, because of the nature of the states or because it becomes too "bushy," eigendecomposition methods offer a second possible solution. Eigendecomposition consists of decomposing the transition probability matrix into a set of eigenvectors and eigenvalues [11, 28, 30, 32]. To employ eigendecomposition, a modeler must have a single data source to inform the transitions from an initial health state, or have data from multiple studies that each have the same follow-up time (Chhatwal, personal communication). These conditions are rarely met. For example, a rheumatoid arthritis model in which treatment efficacy and side effects are obtained from a single RCT may still require all-cause mortality data from another source due to the short follow-up time of the trial. If this second source has a different follow-up period, the eigendecomposition approach cannot be applied. For this reason, eigendecomposition may also be inapplicable for models in which scenario analyses are conducted on cohort subgroups, with multiple sources providing probability data. In addition, technical problems having to do with the nature of the data may make eigendecomposition impossible or produce values that do not meet the conditions required of transition probabilities. For example, eigendecomposition may produce negative numbers or complex numbers (combinations of real and imaginary numbers). A full exposition of these complex methods is beyond the scope of this paper. Interested readers can consult the articles cited earlier in this paragraph [11, 28, 30, 32].

Lastly, in cases where (1) there are only three health state transitions possible, (2) two of the published probabilities are very small, and (3) the model cycle length is shorter than the published probability, the error in using the two-state formula to convert probabilities to the appropriate cycle length will be small. If all three of these conditions are met, and the resulting probabilities are not a major driver of the model results, modelers may wish to consider using this approach.

6 Sensitivity Analyses

The purpose of sensitivity analyses is to understand how uncertainty about parameter values, including transition probabilities, affects a model's results. Each parameter has some error, which should be represented and evaluated in the decision model [1]. The error is represented by a plausible range around the base-case value, based on 95% confidence intervals or expert opinion if formal estimates of uncertainty are lacking. Arbitrary ranges should be avoided. If the base-case estimate of a parameter required conversion from the study time period to the model's cycle length, the same procedure can be used to transform the bounds of its range to the appropriate cycle length. For example, the upper and lower bounds of the reported 95% confidence interval for a probability can also be transformed to the model's cycle length using Eqs. 8–10.

The two main ways to evaluate the effects of uncertainty on model results are deterministic sensitivity analyses and probabilistic sensitivity analyses (PSAs). Deterministic sensitivity analyses entail varying the value of one parameter (one-way sensitivity analysis) or a few parameters (multi-way sensitivity analysis) while holding all other parameters at their base-case values. A series of one-way sensitivity analyses plotted in a tornado diagram shows which parameters have the most influence on the model's results. If the ranges in the tornado diagram reasonably represent the parameters' uncertainty, the diagram points to the most influential parameters. These are the parameters that require the most attention and effort to reduce uncertainty and the possibility of bias in the model's results. Thus, sensitivity analyses conducted early in model development can be a useful guide for allocating effort for further refinement of parameters.

PSAs entail replacing the base-case values for all model parameters with probability distributions. Each distribution represents the range of values the parameter can take and the likelihood of each value. The model is run multiple times, say 1000 times, and each iteration plucks a new set of parameter values from the distributions. The results for the 1000 iterations show the uncertainty in the output, such as the cost-effectiveness ratio, due to uncertainty in all parameters. Since PSAs vary all parameters simultaneously, some transition probabilities may need to be linked to ensure that the values selected in a single iteration are congruent. Modelers who are running their own regressions to obtain model input parameters can use the variance–covariance matrix to specify the correlation between two or more model parameters [33]. When the modeler does not have access to the

individual-level data, this linkage can be done through other mechanisms. For example, in a model evaluating multiple lines of treatment for cancer, it may be necessary to link the probability of response to second-line treatment to the probability of response to first-line treatment, perhaps by defining the probability for second-line treatment as a fraction of the probability for first-line treatment. If the two values are left independent, the PSA can produce implausible model iterations.

Another source of uncertainty derives from the time frame of the original statistic. Consider again the probability of new mothers reporting post-partum depression symptoms, 11.5%. This outcome was reported for a mean follow-up time of 125 days, approximately 4 months, after delivery [22]. To derive a transition probability, the modeler may treat the value of 11.5% as the probability for 4 months and use Eqs. 8 and 9, or Eq. 10, to transform it to the cycle length of the model. But because not everyone was followed for at least 4 months (the mean follow-up was 4 months), this is not correct, and the probability has not only the usual sampling error, but also an additional error associated with the time frame. While this remains an area for future research, modelers should test the impact of this measurement error by conducting a series of one-way deterministic sensitivity analyses using values associated with varying the time frame—for example, transition probabilities derived using the mean, median, minimum, and maximum follow-up times reported for the statistic. If the model results are sensitive to the difference, modelers may wish to contact the study investigators for data stratified by follow-up time or explore alternative sources of data.

An important issue, for which there is as yet no good solution, is how to represent the uncertainty involved in extrapolating transition probabilities beyond the time horizon of the available data. When the modeler has access to the original data, the standard approach is to fit a variety of parametric models to the data, and, using a goodness-of-fit statistic such as the Akaike and/or Bayesian information criterion, choose the best-fitting model to extrapolate beyond the original data, even though goodness-of-fit to the observed data is not an appropriate test of the fitted model's ability to extrapolate accurately. Negrin et al. [34] and Latimer [16] suggest conducting sensitivity analyses by comparing, say, the cost-effectiveness results for the best-fitting model with results based on those that fit less well. This approach shows whether an intervention deemed cost-effective (or not) using the best-fitting model remains so when other models are used and focuses attention on the effects of the extrapolation method on decision uncertainty. Another approach is to fit parametric models to different subperiods within the observed data to explore the stability of the estimated parameters and, rather than using the single best-fitting model, use Bayesian model averaging to combine models for extrapolation [34]. Modelers who are limited to the published data will not be able to use these approaches, but should consider widening the range around extrapolated probabilities to reflect the additional uncertainty associated with extrapolation. The importance of the problem is illustrated by an analysis of artificial hips that found extrapolations based on the 8 years of follow-up data available at the time of the original analysis turned out, once 16 years of follow-up became available, to have identified the wrong artificial hip as the most successful and cost-effective [15].

7 Discussion

There are many complex issues to be addressed in the process of developing a decision model. Here, we summarize some best practices for using data from the published literature that may mitigate downstream challenges.

First, as Miller and Homan warned more than 2 decades ago, statistics are not always correctly described in the original source [27]. Modelers should carefully review whether the reported statistic is actually a probability, an RR, a rate, or something else; sometimes statistics reported as rates are actually probabilities. As noted earlier, one clue to the difference is that a rate has *time at risk* explicitly stated in the denominator, (e.g., ten events per 100 person-years) while a probability does not (e.g., ten events per 100 persons). Another clue is that for a probability, persons must be followed for the entire time period, whereas for a rate, persons are followed only until the event occurs [26].

Once the published data relevant to the decision model are correctly identified, the methods described in this paper can be used to derive transition probabilities appropriate to the model. Our purpose here has been to collect the methods available in the literature in a single place to make the process of derivation easier for modelers. We have described how RRs can be used to derive transition probabilities for disease or for treatment efficacy, or can serve as weights for deriving transition probabilities for population subgroups. We described how to derive probabilities from the frequently-reported OR, including in situations where the event is not rare. Probabilities derived from summary statistics such as RRs or ORs will be affected by the accuracy and suitability of additional elements required by the derivation, such as the probability of an event in the unexposed, and we have discussed how modelers can incorporate the uncertainty introduced by these elements.

We have discussed several types of statistics that are of direct use for estimating transition probabilities for decision models. There are other statistics that, while not directly useful, are excellent leads to sources of the statistics needed for models. The population attributable fraction (PAF) is one such example [35, 36]. Since it shows the maximum

amount of disease or mortality that can be attributed to a condition, such as obesity [37], it is not directly useful for cost-effectiveness analysis models, which evaluate specific interventions and need measures of the effectiveness of those interventions. However, the calculation of PAF is based on prevalence and risk ratios, from which transition probabilities can be derived, so articles about PAFs can lead to good sources for those statistics. They also often provide helpful discussions of the appropriateness and consistency of the statistics for particular populations, so can help the modeler decide which statistics to use in the model.

Modelers will often need to modify a published probability to match the cycle length of the model. When there are only two possible transitions within a model cycle, the conversion is straightforward, as described in Eqs. 8 and 9 or Eq. 10. When three or more transitions can occur within a single model cycle, modelers can avoid the need to use more complex methods for deriving appropriate probabilities by creating two-branch, conditional nodes, as suggested for the heart transplant example. In other cases, such conditional nodes may not be appropriate, or may result in a model with a confusingly large number of branches, and modelers can instead consider eigendecomposition to obtain model transition probabilities for nodes with three or more branches. Regardless of the approach used, the resulting probability values are estimates only and therefore contain uncertainty additional to that which is present in reported probabilities, which should be considered in the analysis. For further reading on the appropriateness of creating conditional two-branch nodes, see Sendi and Clemen, who point out that two-branch nodes can sometimes complicate sensitivity analyses [38].

Aspects of the model's structure can be chosen to accommodate the available data or to simplify the process of using it. Modelers may choose, for example, to match the model's cycle length to the follow-up time for the data considered most important to model results. Whatever the cycle length, the discount rate needs to be adapted to match it.

Sensitivity analyses are a standard part of reporting model results. They can also be useful early in model development to help allocate effort to the refinement of parameters. Deterministic sensitivity analyses and, specifically, the tornado diagram produced from a series of one-way deterministic sensitivity analyses are an excellent mechanism for determining the importance of model parameters to results. An iterative process may be useful in which plausible placeholder values are entered, a tornado diagram is run, and the results are used to allocate further effort in proportion to each parameter's effects on the model. Attempting to derive "perfect" probabilities for every branch in the model delays the model's completion while adding little to its ultimate quality and usefulness.

8 Conclusion

Decision modelers populating their models with transition probabilities based on published data face numerous challenges, ranging from finding only comparative statistics (such as RRs or ORs) from which to derive probabilities to needing to convert published data to match the model's cycle length. We present here guidance, based on current thinking and literature, to help modelers populate their models with high-quality transition probabilities.

Appendix: Probabilities and rates

Suppose a study followed 100 people with congestive heart failure for 4 years. At the end of the 4 years, 40 had died.

The probability of death over 4 years is $40/100 = 0.40$.

A rate takes into account the time each person was at risk. The 60 people who survived were at risk the entire 4 years and contributed $60 \times 4 = 240$ years at risk. Once a person dies, he/she is no longer at risk. When a study does not report time at risk, the conventional assumption is that the events were spread evenly over the time period. Using this assumption, the average time at risk for the 40 who died was 2 years, adding $40 \times 2 = 80$ years at risk. Total time at risk for the cohort of 100 people is 320 person-years, and thus, the *rate* is $40/320 = 0.125$ deaths per person-year at risk.

Suppose the decision model requires an annual probability of death. Starting with the published 4-year probability, use Equation 8 to convert the probability to an annual rate on the assumption that the deaths occurred evenly throughout the study period. (Note that the answer is close to the manually calculated rate above, but not quite the same.)

$$r = \frac{-\ln(1-p)}{t} = \frac{-\ln(1-0.40)}{4} = 0.127706 \text{ deaths per person-year.}$$

Then use Eq. 9 to convert this rate to a probability.

$$p = 1 - \exp(rt) = 1 - \exp(-0.127706 \times 1)$$
$$= 0.119888, \quad \text{annual probability.}$$

Since Eq. 8 already divided by $t = 4$ to adjust to 1 year, $t = 1$ in Eq. 9. (Alternatively, the 4-year probability could first be converted to a 4-year rate and then the 4-year rate to an annual probability. In this case, $t = 1$ in Eq. 8 and $t = \frac{1}{4}$ in Eq. 9. The results are the same.)

Instead of using Eqs. 8 and 9, Eq. 10 yields the same annual probability in a single step,

$$p = 1 - (1-p)^{1/n} = 1 - (1-0.40)^{1/4}$$
$$= 0.119888, \quad \text{annual probability.}$$

That this is the correct annual probability can be verified by using it to calculate the number of survivors at 4 years.

Year 1 : $100 \times (1 - 0.119888) = 88.011174$

Year 2 : $88.011174 \times (1 - 0.119888) = 77.459667$

Year 3 : $77.459667 \times (1 - 0.119888) = 68.173162$

Year 4 : $68.173162 \times (1 - 0.119888) = 60.$

References

1. Neumann PJ, Sanders GD, Russell LB, et al. Cost-effectiveness in health and medicine. 2nd ed. New York: Oxford University Press; 2016.
2. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–3. Value Health. 2012;15(6):812–20.
3. Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices–overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. Value Health. 2012;15(6):796–803.
4. Borenstein M, Hedges LV, Higgins JPT, et al. Introduction to meta-analysis. Chichester, West Sussex, United Kingdom: Wiley; 2009.
5. Sutton A, Ades AE, Cooper N, et al. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics. 2008;26(9):753–67.
6. Jansen JP, Crawford B, Bergman G, et al. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health. 2008;11(5):956–64.
7. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health. 2011;14(4):417–28.
8. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health. 2011;14(4):429–37.
9. Welton NJ, Caldwell DM, Adamopoulos E, et al. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. Am J Epidemiol. 2009;169(9):1158–65.
10. Basu A, Ganiats TG. Discounting in cost-effectiveness analysis. In: Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, editors. Cost-effectiveness in health and medicine. 2nd ed. New York: Oxford University Press; 2016. p. 277–288.
11. Chhatwal J, Jayasuriya S, Elbasha EH. Changing cycle lengths in state-transition models: challenges and solutions. Med Decis Making. 2016;36(8):952–64.
12. Guyot P, Ades AE, Beasley M, et al. Extrapolation of survival curves from cancer trials using external information. Med Decis Making. 2017;37(4):353–66.
13. Jackson C, Stevens J, Ren S, et al. Extrapolating survival from randomized trials using external data: a review of methods. Med Decis Making. 2017;37(4):377–90.
14. Hawkins N, Grieve R. Extrapolation of survival data in cost-effectiveness analyses: the need for causal clarity. Med Decis Making. 2017;37(4):337–9.
15. Davies C, Briggs A, Lorgelly P, et al. The "hazards" of extrapolating survival curves. Med Decis Making. 2013;33(3):369–80.
16. Latimer NR. Survival analysis for economic evaluations alongside clinical trials–extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Med Decis Making. 2013;33(6):743–54.
17. Russell LB, Pentakota SR, Toscano CM, et al. What pertussis mortality rates make maternal acellular pertussis immunization cost-effective in low- and middle-income countries? A decision analysis. Clin Infect Dis. 2016;63(suppl 4):S227–S23535.
18. Clark A, Sanderson C. Timing of children's vaccinations in 45 low-income and middle-income countries: an analysis of survey data. Lancet. 2009;373(9674):1543–9.
19. Juretzko P, von Kries R, Hermann M, et al. Effectiveness of acellular pertussis vaccine assessed by hospital-based active surveillance in Germany. Clin Infect Dis. 2002;35(2):162–7.
20. Black WC, Gareen IF, Soneji SS, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. N Engl J Med. 2014;371(19):1793–802.
21. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med. 2013;368(4):341–50.
22. Ko JY, Rockhill KM, Tong VT, et al. Trends in postpartum depressive symptoms—27 states, 2004, 2008, and 2012. MMWR Morb Mortal Wkly Rep. 2017;66(6):153–8.
23. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA. 1998;280(19):1690–1.
24. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. BMJ. 2014;348:f7450.
25. Norton EC, Dowd BE. Log odds and the interpretation of logit models. Health Serv Res. 2018;53(2):859–78.
26. Fleurence RL, Hollenbeak CS. Rates and probabilities in economic modelling: transformation, translation and appropriate application. Pharmacoeconomics. 2007;25(1):3–6.
27. Miller DK, Homan SM. Determining transition probabilities: confusion and suggestions. Med Decis Making. 1994;14(1):52–8.
28. Jones E, Epstein D, Garcia-Mochon L. A procedure for deriving formulas to convert transition rates to probabilities for multistate Markov models. Med Decis Making. 2017;37(7):779–89.
29. Christensen K, Coons M, Walsh R. 2016 report on childhood lead poisoning in Wisconsin. Wisconsin Department of Health Services, Division of Public Health, Bureau of Environmental and Occupational Health, P-01202–16. 2017. https://www.dhs.wisconsin.gov/publications/p01202-16.pdf. Accessed 11 Apr 2019.

30. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Econ. 2002;11(1):33–42.

31. Anguita M, Arizon JM, Valles F, et al. Influence of heart transplantation on the natural history of patients with severe congestive heart failure. J Heart Lung Transpl. 1993;12(6 Pt 1):974–82.

32. Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Med Decis Making. 2005;25(6):633–45.

33. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

34. Negrin MA, Nam J, Briggs AH. Bayesian solutions for handling uncertainty in survival extrapolation. Med Decis Making. 2017;37(4):367–76.

35. World Health Organization. Metrics: population attributable fraction (PAF), https://www.who.int/healthinfo/global_burden_disease/metrics_paf/en/. Accessed 20 May 2020.

36. Greenland S. Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities. Annals of Epidemiology 2015;25:xs155–161.

37. Flegal KM, Panagiotou OA, Graubard BI. Estimating population attributable fractions to quantify the health burden of obesity. Ann Epidemiol. 2015;25:201–7.

38. Sendi PP, Clemen RT. Sensitivity analysis on a chance node with more than two branches. Med Decis Making. 1999;19(4):499–502.