5 Transformation of Evidence

5.1 Transformation of Clinical Evidence

Key Recommendation: Clinical trials should be analysed using data from the intention-to-treat (ITT) population. All statistically significant clinical events should be included in base-case analyses. For clinical events with a $p$ value close to 0.05, consideration should be given to the magnitude of effect; whether the results are likely to be clinically significant; the relevance and validity of composite measures; and whether statistical significance has been demonstrated in an independent study. The exclusion of any event from an analysis should be justified.

It is important to make sure the outcomes most relevant to the condition are included in the CUA and that they reflect the perspective and scope of the model. This will often require incorporating information on relative treatment effects (usually obtained from clinical trials) with baseline health events.

Outcomes included in the model may include (but are not limited to):

- probability of success or failure
- relapse
- adverse events
- discontinuation/loss to follow-up
- death.

These outcomes should be well defined, mutually exclusive, and generally long-term or final outcomes.

5.1.1 Use of Surrogate versus Clinically Important Outcome Measures

Economic analysis should ideally be based on studies that report clinically important outcome measures. These are valid outcomes that are important to the health of the patient.

In some cases, only surrogate outcomes may be available. These are a substitute for a clinically meaningful endpoint, and measure how a patient feels, functions or survives.

Surrogate measures should only be used in CUAs where no alternative health outcome data are available. Caution must be used when using surrogate measures, as these may not necessarily translate into clinically relevant and effective outcomes.

5.1.2 Analysing Data from Clinical Trials

Clinical trials should be analysed using data from the intention-to-treat (ITT) population, rather than per protocol (PP), in order to take into account outcomes of all patients irrespective of whether they received treatment. For further information on data sources to be used when estimating relative treatment effects, refer to Chapter 4.

Where ITT analysis has not been reported, the effectiveness rates should ideally be recalculated by adding to the ‘on treatment’ participant population for the group (ie the denominator) all of the patients who withdrew, dropped out, or were otherwise lost to follow-up. This is the group’s true ITT starting participant population.

CUAs should not include last-observation-carried-forward (LOCF) analysis due to the large bias this incorporates in economic models. LOCF assumes that a patient who drops out of the study will continue to be in the same state as the last
time they were assessed. In studies where patients’ health is deteriorating, this may overestimate the effects of a treatment (17 (https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref17)).

5.1.3 Relative Clinical Effectiveness Data to Include

PHARMAC recommends that all statistically significant clinical events be included in a cost-utility analysis. Statistical significance is defined here as the \( p \) value being less than 0.05[5](PFPA%202.2%20z%20compiled%20chapters%20and%20Appendices.docx#ftn5).

For clinical events with a \( p \) value close to (but still larger than) 0.05 (ie the event is close to but does not reach conventional statistical significance), the following issues should be considered.

Table 7: Issues to Consider when Evaluating Statistically Insignificant Events

<table>
<thead>
<tr>
<th>Issue</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnitude of effect</td>
<td>Is the treatment effect size substantial given size of study?(^{(6)})</td>
</tr>
<tr>
<td></td>
<td>(PFPA%202.2%20z%20compiled%20chapters%20and%20Appendices.docx#ftn6)</td>
</tr>
<tr>
<td>Clinical significance</td>
<td>Is the outcome patient focused with clinically meaningful effects on longevity or quality of life and with good evidence for causality(^{(7)})</td>
</tr>
<tr>
<td></td>
<td>(PFPA%202.2%20z%20compiled%20chapters%20and%20Appendices.docx#ftn7)</td>
</tr>
<tr>
<td>Independent study</td>
<td>Has statistical significance been demonstrated in more than one independent study (or in a meta-analysis of relevant studies), with no evidence of statistical heterogeneity?</td>
</tr>
<tr>
<td>Composite events</td>
<td>Are similar events statistically significant when combined(^{(8)})</td>
</tr>
<tr>
<td></td>
<td>(PFPA%202.2%20z%20compiled%20chapters%20and%20Appendices.docx#ftn8)</td>
</tr>
</tbody>
</table>

Accounting for clinical factors and magnitude of effect means that, in some cases, a result considered to be ‘statistically non-significant’ (ie \( p \) value equal to or greater than 0.05) should still be used. This is because the magnitude of clinical relevance overrides the statistical aspects. Likewise, in some cases a result considered to be statistically significant (\( p \) value less than 0.05) should not be used, because it has no meaningful clinical effects.

When analysing multiple events without significant effects individually, it is preferable to use raw data and conduct suitable statistical tests (eg F-test). When only summary data are available, it is important to also take into account the likelihood of the same patient being included in multiple groups.

A clear exception, where events that are not significantly different between groups can be omitted, is when there is no difference in survival and any difference in the mean (point estimate) of events favours the comparator (eg if the new intervention has fewer adverse events but statistical significance is not reached).

In general, the exclusion of any statistically significant event from an analysis should be justified, and the impact of a decision to include or exclude certain parameters should be included and tested in the sensitivity analysis. However, for rapid analyses, statistically non-significant events should only be included if they are likely to change the results of the analysis.

5.1.4 Incorporation of Relative Treatment Effects with Baseline Events

A common approach is to model risk factors or interventions as having an additive or multiplicative effect on baseline probabilities, mortality or disease incidence. This is done by deriving relative risks (or hazard or odds ratios) between treatment options in clinical trials, and then ‘superimposing’ these estimates onto baseline probabilities derived from other sources (usually population based) (9 (https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref9), 23 (https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref23)).

Once the baseline probabilities have been determined, a relative risk can be applied to the proposed treatment group. This may include a relative risk reduction if the proposed treatment reduces the risk of exacerbation, relapse, mortality, etc.

For example, disease-specific mortality can be used with all-cause mortality. All-cause mortality should be derived from
New Zealand life tables (24), unless an alternative source can be justified. In general, it is not necessary to correct for the fact that all-cause mortality includes disease-specific mortality in the general population, unless the disease represents a major cause of death in the population (23). The choice of functional form for disease-specific mortality should be specified and justified.

More detailed information on the incorporation of relative treatment effects can be found at http://www.pbs.gov.au/info/industry/listing/elements/pbac-guidelines (external link).

5.2 Extrapolation of Data

**Key Recommendation:** The methodology, limitations, and any possible bias associated with extrapolating data should be clearly described in the report and explored through sensitivity analysis. This includes extrapolating data from clinical trials to the longer term (or to final outcomes); generalising results from clinical trials to the New Zealand clinical setting by taking into account non-compliance; and undertaking indirect comparisons of trials. It is recommended that in the absence of conclusive data, conservative assumptions be used in the analysis.

Data from clinical trials and other sources need to be translated into an appropriate form for incorporation into a model. Modelling may require:

- extrapolating data to the longer term
- translating surrogate (intermediate) endpoints to obtain final outcomes affecting disease progression, overall survival and/or quality of life
- generalising results from clinical trials to the New Zealand clinical setting
- using indirect comparisons where the relevant trials do not exist.

The methodology, limitations, and any possible biases associated with extrapolating and incorporating data should be clearly described in the report and explored through sensitivity analysis.

In the absence of conclusive data, conservative assumptions should be applied in the analysis. This may include cases where there is uncertainty about the:

- long-term benefit of treatment (ie beyond the period of the trial(s))
- correlation between surrogate measure and clinical outcomes
- effectiveness of treatment (ie if evidence is of low-quality, such as non-randomised trials)
- relevance of evidence to New Zealand clinical practice (ie poor external validity of trials)
- incremental effectiveness of treatment (ie if indirect treatment comparison data are used).

5.2.1 Extrapolation to Longer Terms

Many trials have endpoints that may be too early to show the full impact of the treatment. Therefore, it may be necessary to use intermediate outcomes to obtain final endpoints by extrapolating data beyond the period observed in the clinical trials, and comparing the extrapolated outcomes with expected long-term outcomes from observational studies (or any clinical trials in other settings with long-term outcomes that are relevant). This often requires explicit assumptions about the continuation of treatment effect once treatment has ceased (8, 25). If there is any uncertainty about long-term benefit, it is recommended that conservative assumptions are applied in the analysis (eg it may be assumed that the benefit reduces or wanes entirely over time). Alternative scenarios should also be included to compare the implications of different assumptions around extrapolation beyond the clinical trial, for example scenarios where the treatment benefit in the extrapolated phase is nil, is the same as during treatment phase, or diminishes in the long term.

5.2.2 Translating Surrogate Endpoints to Final Outcomes

Available evidence may be limited to surrogate endpoints rather than clinically important outcome measures that affect
disease progression, overall survival and quality of life. Therefore, it may be necessary to translate surrogate endpoints to clinically important outcomes, using data from observational studies that relate the surrogate outcome to the clinically important endpoints (or any clinical trials in other settings with clinically important outcomes that are relevant).

If there is uncertainty about the clinical significance of endpoints or the correlation between surrogate measure and clinical outcomes, conservative assumptions should be applied in the analysis regarding their impact (short and/or long term) on survival and/or health-related quality of life. In the absence of conclusive data, conservative assumptions should be included in the analysis.

5.2.3 Impact of Operator Skills and Experience: External Validity of Trials

The benefit of some pharmaceuticals, in particular many medical devices, is linked to how that pharmaceutical is applied. The efficacy of such a medicine or device in clinical practice may therefore differ from trials, due to the experience and skill of the operator. For example, if only experienced operators take part in the trial, the efficacy of the pharmaceutical in clinical practice may be lower in the first few years as operators gain the necessary experience and skills. During this ‘learning curve’, errors and adverse outcomes are potentially more likely.

In cases where there is evidence of reduced efficacy or safety in clinical practice compared with the trial, the analysis should adjust the efficacy/safety of a pharmaceutical in the first few years, and assume increased efficacy/safety over time as operators gain experience.

5.2.4 Product Modifications: Relevance of Trial Data over Time

Medical devices frequently undergo product modifications, some of which may impact on efficacy. Modifications are often incremental, based on emerging clinical evidence or use in clinical practice. Clinical trial data may become less relevant over time as the pivotal clinical trials may have been undertaken at an early stage in the technology’s evolution.

In cases where products have been modified since the reported clinical trials, it is recommended that the assessment be based on a synthesis of the trial data (to evaluate overall efficacy of product group) and any further evidence available on the impact of product modifications on the efficacy of the device. Any reported improvements in efficacy and safety should be assessed according to the grades of evidence. For example, any improvements reported by observational studies should be modelled conservatively because observational studies are a lower grade of evidence. If there is no evidence available on the efficacy of the modification, the assessment should be based solely on the initial trial evidence and should not assume any improvements to efficacy and/or safety due to modifications.

5.2.5 Extrapolation of Clinical Trial Data to the New Zealand Clinical Setting

It is important that the effectiveness and cost data included in the economic model are applicable to the New Zealand health sector. Clinical practice in New Zealand may differ from that in clinical trials in terms of the level of resources available (eg staffing), patient management (eg frequency of consultation), and type of patient. These may in turn impact on compliance rates and therefore change the effectiveness of treatment in clinical practice.

Some types of treatment non-compliance and non-adherence are listed in Table 8.

Table 8: Types of Non-compliance
### Types of Non-compliance

<table>
<thead>
<tr>
<th>Types of Non-compliance</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary non-compliance</td>
<td>Failing to initiate treatment – equivalent to no treatment.</td>
</tr>
<tr>
<td>Drug regimen non-compliance</td>
<td>Treatment ‘holidays’, inadequate treatment dose, administration timing variations, treatment withdrawal.</td>
</tr>
<tr>
<td>Premature discontinuation</td>
<td>Failing to complete a recommended course of treatment, and/or non-redemption of repeat prescriptions.</td>
</tr>
</tbody>
</table>

PHARMAC recommends that non-compliance be included in the model when there is evidence that non-compliance rates may be material to the point that they may impact the effectiveness and cost of treatment. Observational data can be used to estimate levels of non-compliance. Non-compliance can be modelled by incorporating different discontinuation rates into the model, and by adjusting the subsequent probability of treatment success for non-compliant and compliant patients. Non-compliance can also cause additional costs, such as hospitalisations and comorbidities.

In cases where non-compliance is likely, but there is absence of evidence for it, the possible effects should be tested in the sensitivity analysis by varying both effectiveness data and costs.

### 5.2.6 Indirect Comparisons of Trials

Many trials may not use the most relevant treatment comparator for the New Zealand clinical setting, or they may not include multiple comparators needed for analysis in the New Zealand setting. In such cases, it may be necessary to synthesise a head-to-head comparison. For example, a difference in clinical effect between medicine A and medicine B can be modelled by obtaining separate estimates from trials comparing medicine A versus placebo, and medicine B versus placebo.

When undertaking indirect comparisons, there is greater uncertainty in the effectiveness of one treatment over the other. This is because the trials that are being compared may contain very different groups of patients, which may alter the overall treatment effect. If indirect comparisons are required in an analysis, conservative assumptions should be applied and these assumptions need to be clearly stated.

For information regarding how results from trials should be synthesised, please refer to the Pharmaceutical Benefits Advisory Committee (PBAC) and Canadian Agency for Drugs and Technologies in Health (CADTH) guidelines.

### 5.3 Subgroup Analyses

If treatment can be targeted to those who are most likely to benefit, subgroup analyses may be necessary.

Subgroup analyses comprise two inter-related elements:

#### 5.3.1 Variability in absolute baseline risk

Variability in baseline risk occurs when differences between patients in aspects such as disease severity cause differences in treatment outcomes. This relatively common effect is best summarised as a constant relative reduction in treatment effects across the trial population of varying baseline (expected) risks. This enables application of the overall trial data to specific subgroups with greater expected absolute risks of future events (i.e., poorer prognosis) and hence greater likelihood of benefiting from a new treatment. The absolute or incremental treatment effect can then be calculated by multiplying the expected absolute risks across the eligible population by the estimated overall relative treatment effect.
5.3.2 Variability in relative treatment effects

Variability in relative treatment effects occurs due to differing characteristics of the patient, the intervention(s), or the disease, causing varying relative reductions in the risk of clinical outcomes across the trial population. The population may also include sub-groups with different absolute baseline risks. In this case, which is far less common, analysis is required to identify statistically significant heterogeneity (variation) in the treatment effects across the subgroups. Such evidence is needed to help justify any calculations of absolute treatment effect that apply the estimated relative treatment effect for the subgroup to the expected risk for the subgroup (22 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref32]).

When examining variability in treatment effects, in order for the results of subgroup analyses to be reliable, the subgroups in the clinical trial (or meta-analysis of clinical trials) should be defined a priori on the basis of known biological mechanisms or in response to findings in previous studies. The choice of subgroup and expected direction of difference should ideally have been justified in the trial protocol (32 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref32]).

Where subgroups are defined retrospectively, information should be interpreted cautiously. This is because it is more likely that differences in effect in subgroups of patients are due to chance, given the smaller patient numbers. There is also an increased probability of either falsely ascribing ‘significant differences’ as a result of over-testing or producing false-negative results (33 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref33]). Because of these concerns, it may be more appropriate to use data from a retrospective subgroup of patients in the sensitivity analysis rather than the base-case analysis.

In addition, statistical tests of interaction (34 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref34], 35 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref35]) should be used to assess whether a treatment effect differs among subgroups (ie evidence of heterogeneity)[12] (PFPA%202.2%20z%20compiled%20chapters%20and%20Appendices.docx#ftn12). However, even when there is heterogeneity between subgroups, results of subgroup analyses should still be interpreted with caution. The outcomes of subgroup analyses should be checked to ensure that they were pre-specified and that treatment effects are both statistically strong and pharmacologically, biologically and clinically plausible (33 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref33]).

When examining variability in treatment effects, subgroup analysis in the CUA can be acceptable if justified by a formal and reliable subgroup analysis of the trial data (22 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref22]) that adequately considers the above elements of plausibility, timing of the underlying hypothesis (a priori), and statistical heterogeneity (22 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref22], 32 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref32]). Otherwise, subgroup analysis should generally not be used when a trial reports statistically significant treatment effect(s) in subgroup(s) or secondary endpoint(s) yet there is no overall treatment effect in the intention-to-treat (ITT) population[13] (PFPA%202.2%20z%20compiled%20chapters%20and%20Appendices.docx#ftn13) or primary endpoint (33 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref33], 36 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref36]).

5.4 Assessment of Vaccines

5.4.1 Adjustments to Vaccine Trial Efficacy Data

The following points should be considered when modelling vaccine efficacy (37 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref37]):

- Proportion of vaccinated people who will be protected – a proportion of vaccinated people experience the intended effects, and the remainder of vaccinated people do not. For example, a vaccine with 90% ‘take’ would then produce the intended effect in 90% of vaccinated people, and not in the remaining 10%.
- Degree of protection – vaccinated people in whom the vaccine ‘takes’ may experience the intended effects to a certain degree (ie not 100% protection). For example, a vaccine with 90% ‘degree’ would produce the intended effect in 90% of vaccinated people in whom the vaccine ‘takes hold’.
- Length of protection – efficacy may remain constant over lifetime or wane as a function of time.
- Age at administration – the immune system shows different responsiveness based on the vaccinated person’s age.
- Adherence with the vaccination schedule (compliance and time between doses) – this especially needs to be considered for vaccines where compliance with a full schedule is problematic.
• Adverse reactions – some people have adverse reactions to a vaccine, which should be taken into account if significant.
• Potential loss of potency – this can be due to heat and cold exposure; however, it only needs to be considered if relevant data are available.
• Herd immunity – whether the vaccine is likely to provide indirect protection to unvaccinated people through appropriate coverage, as in section 5.4.2 below (further details provided below).

5.4.2 Herd Immunity and Vertical Transmission

Key Recommendation: Include herd immunity in CUA models if vaccine coverage is likely to be high enough to achieve herd immunity and if the inclusion is likely to affect the relative cost-effectiveness of the intervention.

Some pharmaceuticals such as vaccines change the population risk of infection. The general case is herd immunity, but the issues also apply to vaccines intended to reduce vertical transmission.

Herd immunity is defined as the indirect protection of unvaccinated individuals in a largely vaccinated population. When a high percentage of the population is protected against a pathogen, it is difficult for a disease to infect new hosts because there are so few new people to infect. This can effectively stop the spread of disease in the community. The extent of protection through herd immunity, therefore, depends on the amount of infection in the community. Once herd immunity is achieved through appropriate coverage, vaccination will more than proportionally reduce the incidence of infection, increase the average age at infection and increase the length of the inter-epidemic period. Models that do not account for herd immunity may underestimate the true effects of vaccination in a population.

A key parameter in a vaccine economic model is the ‘force of infection’ – the probability that susceptible individuals become infected per unit of time. In a static model, the force of infection is constant over time, whereas in a dynamic model it can change over time. Vaccination reduces the proportion of people in the susceptible stage. Therefore, as more people are vaccinated, the proportion of infectious people will decrease, and hence the probability that a susceptible person will come into contact with an infectious person will also decrease. As a result, the force of infection declines.

In a dynamic model, the force of infection is recalculated each time period. The consequence of a decline in the force of infection due to vaccination is that if susceptible persons are infected, the infections will occur, on average, at a later age. The age at infection continues to shift as long as the probability of infectious persons contacting with susceptible persons continues to decline. Dynamic models are particularly useful if herd immunity is important.

All dynamic models share the same distinguishing feature – that the risk of infection is dependent on the number of infectious agents at a given point in time. In a dynamic model, the probability of an individual acquiring an infection is dependent on:
• the contact patterns of the individual (ie interaction between individuals)
• how infectious the infection is
• the distribution of the infection within the population over time
• vaccination coverage (ie the proportion of the eligible population who receive vaccination).

As outlined above, the age at infection continues to shift as long as the probability of infectious persons coming into contact with susceptible persons continues to decline. This age shift can by itself have beneficial public health effects or detrimental effects (if infection is more severe in adults than in children). Therefore, it is important to assess whether the net effect of herd immunity is positive or negative.

PHARMAC recommends including herd immunity in assessments of vaccines if:
• vaccine coverage is likely to be high, and therefore herd immunity is likely to occur. The level of coverage required...
for herd immunity, which will vary across antigens, therefore needs to be assessed prior to economic modelling.

- the inclusion of herd immunity is likely to have an impact on the relative cost-effectiveness of the vaccine.

Static models may be appropriate if:

- herd immunity does not play an important role (i.e., the additional effectiveness per additional person vaccinated is constant).

### 5.5 Transition Probabilities

**Key Recommendation:** Convert rates to transition probabilities for use in assessments.

#### 5.5.1 Point Estimates versus Probability Distributions

In most assessments, the use of point estimates is sufficient. It is currently recommended that probability distributions be used only in detailed analyses.

#### 5.5.2 Converting Rates to Probabilities to Transition Probabilities

A rate is defined as an instantaneous likelihood of transition at any point in time, whereas a probability is the proportion of the population at risk that makes a transition over a specified period of time. As Markov models track transitions at discrete time intervals, rates should be converted to transition probabilities (42 [https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/pfpa/references/#ref42]).

A rate can be converted to a probability using the following formula:

\[
p = 1 - e^{-rt}
\]

where:

- \( p \) = probability of an event;
- \( r \) = constant rate;
- \( t \) = time

The probabilities included in the model must correspond to the model's cycle length. If the Markov cycle length is changed (e.g., from yearly to monthly), the probability should not simply be divided by the number of cycles (e.g., 12) to obtain the transition probability for the shorter cycle. Rather, the above formula should be used (i.e., \( p = 1 - e^{-rt/12} \)).

Transition probabilities can also be converted to rates using the inverse formula below. A common use is to adjust transition probabilities for a change in cycle length, for example from yearly to monthly. The yearly transition probability should be converted to the corresponding rate, which is then converted back to a monthly probability.

\[
r = \frac{-\ln (1 - p)}{t}
\]

where:

- \( r \) = constant rate;
- \( p \) = probability of an event;
- \( t \) = time

### 5.6 Mortality Rates
Life-expectancy estimates should be based on the age and gender-specific life expectancy of the patient population in New Zealand, taking into account disease-specific mortality. Life expectancy should not be adjusted for potential changes in life expectancy in the future or for possible comorbidities.

Last updated: 10 August 2017