The Challenge of Early Crossover in Oncology Trials

Background Paper 1: Treatment Switching and Adjustment Methods

October 2014

Nicholas R Latimer[1]
Helen Bell[1]

[1]School of Health and Related Research, University of Sheffield, UK

ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
Tel: +44 (0) 114 222 0821
Email: n.latimer@shef.ac.uk

Contents

1. Introduction

This paper provides preparatory reading, to facilitate discussion during the meeting on "The Challenge of Early Crossover in Oncology Trials" to be held in Adelaide, Australia, in October 2014. The paper is not exhaustive, and does not cover every issue associated with treatment crossover (also called treatment switching) in detail. However it aims to provide an overview of the key issues associated with treatment crossover in the context of oncology randomised controlled trials (RCTs).

In Section 2 of this paper we will define what we mean by treatment switching. In Section 3 we will describe why treatment switching causes a problem for the analysis of trials, from the perspective of a range of stakeholders. This will take into account clinical development programmes and the challenges of designing these in the face of international variations in clinical, regulatory and coverage practice with respect to evidence requirements and expectations. In Section 4 we will introduce approaches that may be taken to adjust for treatment switching and in Section 5 we will summarise the performance of these methods in simulation studies. This paper is supplemented by five additional papers: Background Paper 2 provides details on case studies submitted by Workshop participants; Background Papers 3-5 provide relevant guidance and recommendations on the use of switching adjustment methods made by regulatory and reimbursement agencies from around the world. Brief background and introduction to these papers are provided in Sections 6 and 7 of this paper. Background Paper 6 presents the proposed confidentiality rules for

the Workshop. In the final section of the current paper, we highlight areas that have not been addressed by currently available guidance documents.

2. Treatment Switching – definitions

In this paper treatment switching is generally defined as the switch from control treatment to experimental treatment by patients randomised to the control group of an RCT. It is worthy of note that some authors use the term "treatment crossover" rather than "treatment switching" – here we have used "switching" because "crossover" may evoke crossover trials, which are a different entity. As defined here, treatment switching does not involve experimental group patients switching onto the control treatment, or patients randomised to either group receiving other post-study treatments. This is in line with definitions of treatment switching previously given in the literature.[1] However, what is classified as treatment switching may differ depending upon the perspective taken.

Previous definitions of treatment switching given in the literature have focused upon an economic evaluation context.[1,2] Generally an economic evaluation seeks to compare a state of the world in which the novel intervention is used and is given to a cohort of indicated patients, to a state of the world where the novel intervention is not used and standard treatments are received. If an experimental group patient discontinues the novel therapy and receives a standard treatment (either that received in the control group or a separate standard treatment) this is likely to have occurred due to treatment failure, toxicity, tolerability, or adverse events. Such events and subsequent treatment

switches are likely to occur in reality and therefore they form a relevant part of the analysis of outcomes in the state of the world in which the new treatment is available. Hence, in general, for an economic evaluation we would not wish to adjust for these treatment changes. Similarly, if control (or experimental) group patients received standard post-study therapies that do not include the experimental treatment, this reflects a realistic treatment pathway and we would not wish to adjust for this in an economic analysis. Even if differential proportions of patients receive different post-study therapies this may reflect appropriate treatment pathways given the initial treatment.

However, from a clinical perspective, these treatment changes might be regarded as treatment switching and there may be a desire to adjust for them. It may be relevant to estimate the treatment effect specific to the experimental treatment, excluding the impact of subsequent treatments (even if the subsequent treatment received is commonly available). In addition, from a clinical *or* an economic perspective, if subsequent treatments received in a trial represent other novel agents that are not part of the standard treatment pathway, or include non-standard treatments that are of the same class as the experimental agent, adjustment may be required because the trial results will neither isolate the survival impact of the new treatment, or demonstrate the impact of adding the new treatment to the pathway of care.

Therefore, the treatments received during an RCT should be carefully considered when assessing what should be defined as treatment switching and the scope for adjustment. This should include an assessment of what constitutes "standard care". In situations where several existing treatments are commonly available it

is likely to be difficult to ascertain the long-term effect of a novel therapy in isolation from potential confounding from all other available treatments. Hence, generally it may be considered that switches to standard therapies do not represent switches that require attention and adjustment.

Switching from the control treatment onto the experimental treatment represents a clear case where adjustment is required, but it may also be justifiable to adjust for switching onto other therapies in some situations – for instance, where these therapies are novel and are not commonly available. However it is important to note that if switches to "other" therapies are to be adjusted for, it may be necessary to make adjustments to survival estimates in the control group *and* the experimental group: it may appear counterintuitive to adjust for switches to non-study drug x in the control group but not for similar switches in experimental group patients.

In this paper we generally define treatment switching as switching from control to experimental treatment, but we include within this definition switching in either the control or the experimental group to other non-standard treatments.

3. Problems caused by treatment switching

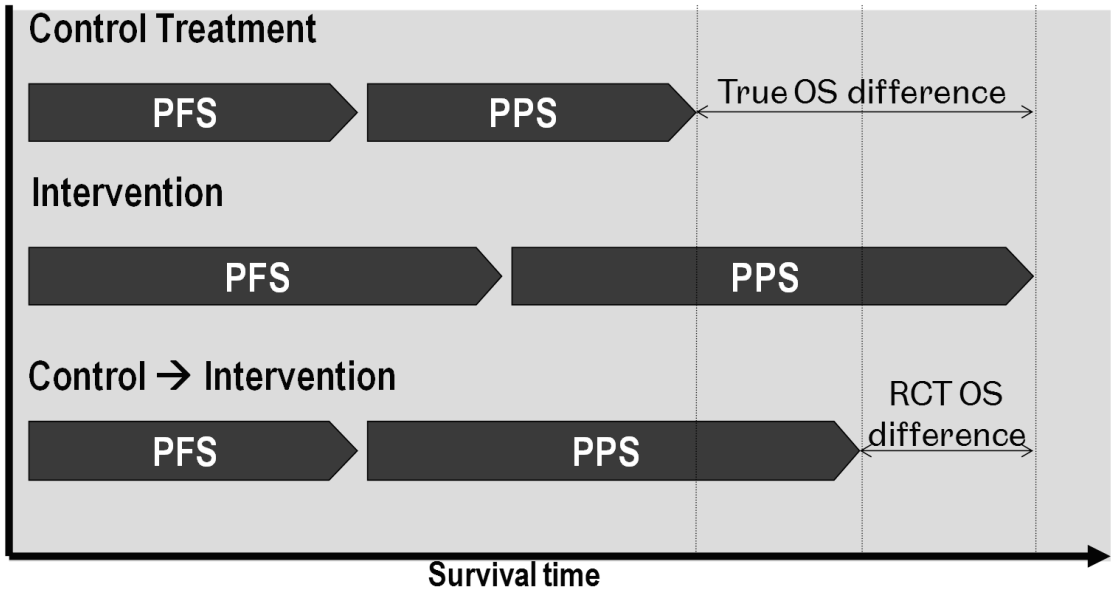
Treatment switching causes issues for various decision makers involved in the development, regulation and use of drugs, each of whom may have different concepts of the value delivered by drugs and hence the evidence they seek from clinical trials.

These issues all stem from the fundamental problem associated with treatment switching: when patients in the control group switch to – and benefit from – the experimental treatment a standard intention to treat (ITT) analysis (a comparison of groups as randomised) will underestimate the "true" survival benefit associated with the new treatment – that is, the benefit that would have been observed if switching had not been permitted. In this paper we define bias as the difference (error) between the estimated treatment effect and the effect that would have been observed in the absence of treatment switching. The bias that may be created by treatment switching and the theoretical problems that it creates for clinical and economic analyses are illustrated in Figure 1 (also presented in [1]).

The first two rows ("Control Treatment" and "Intervention") illustrate the "perfect" trial, where no treatment switching occurs. Survival time is on the x-axis, and in this example the new intervention extends PFS and post progression survival (PPS). This results in the "True OS difference" identified in the diagram. In this case, a standard ITT analysis will usually give us the information that we need (ignoring any need for extrapolation). However, the third row ("Control → Intervention") demonstrates what may happen to survival in the control group if treatment switching is permitted (in this case, after disease progression). PPS is extended compared to the "Control Treatment" comparator, under the assumption that some control group patients switch and benefit from the new intervention after disease progression. The result of this is that the OS difference observed in the RCT ITT analysis (labelled "RCT OS difference" in Figure 1) is smaller than the true OS difference that would have been observed if no

treatment switching had occurred. The simple ITT analysis will result in bias equal to the difference between the "true OS difference" and the "RCT OS difference" when treatment switching occurs. The extent of this bias will be unknown, as the true OS difference will be unobserved. However it is clear that provided switching patients benefit to any extent from the new intervention, some bias will exist. The extent to which estimates of the treatment effect on OS are likely to be confounded is likely to depend upon the crossover proportion, and survival times post progression.[3]

Figure 1: The potential impact of treatment switching illustrated



Notes: PFS = Progression Free Survival; PPS = Post Progression Survival; OS = Overall Survival; RCT = Randomised Controlled Trial

It is worthy of note that here we focus on a situation where treatment switching results in the observed treatment effect being smaller than the "true" treatment effect – assuming that treatment switchers benefit from switching treatment. This may not always be the case – if a novel therapy has little benefit, or is not useful after disease has progressed, it is possible to envisage a situation where treatment switching could inadvertently benefit the experimental arm of a

trial.[4] This highlights the importance of an intimate understanding of the mechanism of action of the new intervention, and the disease. This does not alter the fundamental problem associated with treatment switching – the observed treatment effect is still likely to be different to the "true" treatment effect.

Given this fundamental problem, it is useful to consider how this impacts upon patients, present and future, the decision making of manufacturers, regulators and health technology assessment (HTA) agencies.

Patients

Patients, both present and future, are fundamentally affected by treatment switching. Patients who consent to take part in clinical trials do so on the understanding that the research being undertaken will be of benefit to future patients. If evidence on the overall survival benefit associated with a new treatment is required to obtain regulatory approval and reimbursement, it may be argued that those future patients are best served by a trial that does not permit treatment switching. In contrast, it may be argued that switching should be permitted, given that it may be possible to make good predictions of OS benefits by commissioning additional observational studies, through analysis of registry data, or by collecting enough relevant information during the trial to enhance the likelihood that robust adjustments can be made using statistical methods. In situations where novel therapies show strong effects on progression-free survival at interim analyses, it is ethically problematic to refuse patients randomised to the control treatment access to the new therapy.[5,6,7]

The case for permitting switching might be affected by the properties of the control treatment – arguments in favour are likely to be particularly strong if the comparator is placebo, whereas the need for switching might be less if other active treatments are available. However, from the patient perspective the issue remains the same – if the experimental treatment is shown to be superior to the control treatment at interim analyses it is in the trial participants' interest to be permitted to switch treatments, if randomised to the control group.

Manufacturers

Treatment switching must be considered by manufacturers during drug development, and during the analysis phase, when the results of confounded RCTs are presented to regulatory and reimbursement agencies.

For each new therapy under development, the issues associated with treatment switching are likely to first become apparent at the design stage of RCTs. A manufacturer may decide to take one of two approaches when designing a trial that may or may not permit treatment switching:

a) Despite the ethical and practical rationale for treatment switching, the trial will not permit treatment switching.
b) Despite the problems associated with allowing treatment switching, the trial will permit treatment switching.

Under case (a) treatment switching is not permitted and therefore trial results will not be confounded by it. However, taking this position may be ethically and practically difficult. Ethically, when there are no other non-palliative treatments

available it may be deemed inappropriate to deny control group patients the new treatment if interim analyses indicate a positive treatment effect. Practically, it may be difficult to recruit to a trial that does not allow treatment switching. In addition, pharmaceutical companies have responded to incentives associated with the acceptance of progression free survival (PFS) as a primary endpoint for drug regulatory approval by agencies such as the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA).[8,9] Adopting PFS as the primary outcome measure in an RCT means that there is less motivation for pharmaceutical companies to ensure that randomised groups are maintained beyond disease progression.

Given these issues, approach (b) may be taken. Once in this position the manufacturer is faced with the likelihood that trial results will be confounded, and should therefore consider how to design the whole clinical development programme in such a way that will allow strong evidence on the effectiveness of the therapy to be produced, even in the presence of treatment switching.

This may involve ensuring that the primary endpoint chosen for the trial is not affected by treatment switching – for instance by dictating that switching can only occur after disease progression (allowing an unconfounded analysis of PFS. Aside from this, several steps may be taken to improve the likelihood that useful estimates of overall survival can be obtained.

Firstly, trial designers may attempt to ensure that sufficient data are collected during the trial to enhance the probability that adjustment methods can be successfully applied when the trial data are analysed: as will be described in

Section 4, different adjustment methods have different data requirements and some are particularly data-intensive, necessitating data collection during the trial that may be more burdensome than usual. Ensuring suitable data collection and pre-specifying which adjustment methods will be used represent important tasks for the designers of RCTs that will permit treatment switching.

An alternative approach that may be taken to lessen the problems associated with treatment switching may be to conduct a second randomisation upon disease progression – such that it is random as to which control group patients switch onto the experimental treatment and which continue on a control therapy.[10,11] One of the potential methods for adjusting for the impact of treatment switching is to consider the trial follow-up period in two segments: an initial period that is randomized, and a second period following disease progression that essentially resembles an observational study. The "two-stage" adjustment method described in Section 4 attempts to address the switching problem by first estimating the treatment effect associated with switchers in this second, observational period, and then by estimating how long switchers would have lived for if they had not switched. This is a useful approach, but it is difficult to account for the selection bias that may arise when treatment switching is at the discretion of the treating clinician. A second randomisation after disease progression could avoid this problem, and robust estimates of overall survival could be obtained from control group patients who were not randomised to switch. However, conducting a secondary randomisation in this way may not satisfy the ethical objectives of the trial, because some control

group patients would not be offered the opportunity to receive the novel therapy.

Finally, manufacturers may decide to commission additional studies to provide supplementary data to aid future adjustment attempts. For instance, observational data might be collected on similar patient cohorts who receive standard therapies, in order to provide historical case-matched estimates of survival in the absence of the novel therapy.

Regulators

Treatment switching causes confounding in the estimate of the clinical effect of novel therapies – therefore, it is clearly relevant for regulators. However, as noted above, PFS has been accepted as a primary endpoint for drug regulatory approval by agencies such as the FDA and the EMA,[8,9] and therefore treatment switching that only occurs after disease progression may not pose a serious problem from a regulatory perspective. However, it is important to note that it is by no means certain that PFS will be accepted as evidence sufficient for licensing. The FDA state that it is critical to show direct evidence of clinical benefit or improvement in an established surrogate endpoint for clinical benefit – whether an improvement in PFS is sufficient depends on the magnitude of effect and the risk–benefit profile of the new treatment compared with available therapies.[8] The EMA states that precise estimates of OS may not be needed for approval in situations in which a large effect on PFS, an extended expected survival after progression, or a clearly favourable safety profile is observed.[12] Clearly, this does not mean that accurate OS estimates are never required.

Treatment switching will represent an important problem for regulators in situations where PFS is not a well-established surrogate for OS (it is worthy of note that the FDA states that data are usually insufficient to allow for a robust evaluation of the correlation between effects on PFS and on OS); where PFS effects are not deemed to be of sufficient magnitude; or where PFS estimates are themselves confounded by treatment switching. In these situations it is likely to be relevant for regulatory agencies to consider adjusted estimates of the treatment benefit associated with the new treatment. The impact of treatment switching on a selection of cases seen by regulatory agencies is highlighted by the case studies presented in the supplementary background paper, summarised in Section 6 of this report.

Reimbursement / Health Technology Assessment Agencies

Treatment switching is likely to cause more serious problems for health technology assessment (HTA) agencies than for licensing bodies. HTA agencies generally use economic evaluation to provide estimates of the cost-effectiveness of new treatments – allowing decisions to be made upon whether novel therapies represent value for money. To reflect the cost-effectiveness of treating an entire disease population with a novel treatment, and to take into account all the potential benefits and costs associated with providing the new treatment, a life-time horizon is generally advocated in economic evaluations, especially for interventions that impact upon survival.[13,14,15,16] Therefore, whilst providing accurate estimates of an OS advantage may not be essential for gaining a license, accurate estimates of the treatment effect on OS are almost always required by HTA agencies.

In the presence of treatment switching, an economic evaluation that relies upon this ITT analysis would produce inaccurate cost-effectiveness results. In the case where control group patients benefit from the experimental treatment the survival advantage of the novel therapy would be underestimated and the incremental cost effectiveness ratio (ICER) would likely be over-estimated (although this would depend upon whether the costs of the experimental treatment are incorporated for treatment switchers in the economic model, and the cost-effectiveness of the experimental treatment in switchers). As a result, inappropriate resource allocation decisions may be made. Therefore, in the vast majority of cases in which treatment switching occurs, it is likely to be relevant for reimbursement/HTA agencies to consider adjusted estimates of the treatment benefit associated with the new treatment, and to assess the sensitivity of cost-effectiveness results to these adjustments. The impact of treatment switching on a selection of cases seen by HTA agencies is highlighted by the case studies presented in the supplementary background paper, summarised in Section 6 of this report.

4. Treatment switching adjustment methods

In this section we introduce treatment switching adjustment methods. We begin with relatively simple methods, before moving on to more complex methods. The simpler methods are more commonly used in HTA, but more complex methods are beginning to be used more regularly.[1] Here we discuss the key assumptions and limitations of the key methods. We focus on the key principles of the methods rather than their mathematics – though further details on the more complex methods are provided in Appendix A.

Simple methods

Intention to treat

An ITT analysis does not attempt to adjust for treatment switching. Groups are compared as randomised, and thus the randomisation-balance of the trial is respected. The ITT analysis represents a valid comparison of randomised groups, but in the presence of treatment switching this is unlikely to be what is required for an economic evaluation because the "true" survival benefit associated with the novel intervention will be diluted due to the switching of control group patients onto the novel therapy.

Per protocol – excluding and censoring switchers

Per protocol (PP) analyses have been commonly used in previous HTAs.[1] Data from patients that switch are either excluded entirely from the analysis, or are censored at the point of the switch. Such analyses are prone to selection bias because the randomisation balance between groups is broken if switching is associated with prognostic patient characteristics – for instance, if patients with either good or poor prognosis are more likely to switch.[17,18] This is highly likely in the case of treatment switching in clinical trials – clinicians decide whether it is appropriate for individual patients to switch and this decision will be made based upon patient characteristics rather than being random.

Complex methods

Inverse Probability of Censoring Weights

The IPCW method has been used in recent HTAs.[19,20] It represents an approach for adjusting estimates of a treatment effect in the presence of any type of informative censoring. In the context of treatment switching, patients are artificially censored at the time of switch, and remaining observations are weighted based upon covariate values and a model of the probability of being censored. This allows patients who have not been artificially censored to be weighted in order to reflect their similarities to patients who have been censored in an attempt to remove selection bias.

The key assumption made by the IPCW method is the "no unmeasured confounders" assumption – that is, data must be available on all baseline and time-dependent prognostic factors for mortality that independently predict informative censoring (switching) and models of censoring risk must be correctly specified.[21] In practice, this is unlikely to be perfectly true, but the method is likely to work adequately if the "no unmeasured confounders" assumption is approximately true – that is, there are no important independent predictors missing. If this is the case, the selection bias associated with the dependence between censoring and failure can be corrected for by replacing the Kaplan-Meier estimator, log-rank test, and Cox partial likelihood estimator of the hazard ratio (HR) with their IPCW versions.[21]

The "no unmeasured confounders" assumption represents a key limitation of the IPCW method. It cannot be tested using the observed data [22,23] and is particularly problematic in an RCT context. The IPCW method represents a type of Marginal Structural Model (MSM), which were originally developed for use with observational data.[24,25] Typically RCT datasets are much smaller than

observational datasets and when fewer data are available (particularly on control group patients who do not switch) the IPCW method may become less stable and confidence intervals may become wide. In addition, some key predictors of treatment switching are usually not collected in RCTs (such as patient preference for switching) and often data collection on key indicators is stopped at some point (e.g. upon treatment discontinuation or disease progression), which hampers the applicability of the IPCW method. Also, the IPCW method cannot work if there are levels of any covariates which ensure (that is, the probability equals 1) treatment switching will occur.[23,24,25]

Although the "no unmeasured confounders" assumption cannot be tested, an assessment of the measured covariates alongside findings from previous studies in similar disease areas combined with an elicitation of expert clinical opinion may provide valuable information. The treatment switching mechanism within the trial of interest should also be explored in order to ascertain how and why treatment switching decisions were made, as this may provide information upon whether data on key switching indicators were collected. Linked to this data issue is that of sample size and event numbers. The IPCW method bases its adjustment on the survival experiences of control group patients who do not switch treatments; if almost all patients switch, and/or very few events are observed in patients who do not switch, the method is unlikely to perform reliably.

Rank Preserving Structural Failure Time Model

The RPSFTM method was designed specifically for an RCT context and has been used recently in HTAs.[19,20,26] It uses a counterfactual framework to estimate the causal effect of the treatment in question,[27] where counterfactual survival times refer to those that would have been observed if no treatment had been given. It is assumed that counterfactual survival times are independent of treatment group and g-estimation is used to determine a value for the treatment effect which satisfies this constraint.

The RPSFTM does not rely upon the "no unmeasured confounders" assumption and identifies the treatment effect using only the randomisation of the trial, observed survival and observed treatment history. It is assumed that the treatment effect (an "acceleration factor", or "time ratio") is equal (relative to the time for which the treatment is taken) for all patients no matter when the treatment is received (the "common treatment effect" assumption), and that the randomisation of the trial means that there is only random variation between treatment groups at baseline, apart from treatment allocated – untreated survival times must be independent of the randomised treatment group.[27] This randomisation assumption should be reasonable in the context of an RCT, but the potential remains for important differences at baseline in small and in larger trials.[28] It is therefore relevant to note that it is possible to adjust for baseline covariates within an RPSFTM analysis, which is useful to increase power.[29]

The clinical and biological plausibility of the "common treatment effect" assumption is more problematic. If patients who switch on to the experimental treatment part way through the trial receive a different treatment effect

compared to patients originally randomised to the experimental group, the RPSFTM estimate of the treatment effect received by patients in the experimental group will be biased. Given that treatment switching is often only permitted after disease progression – at which time the capacity for a patient to benefit may be different compared to pre-progression – the "common treatment effect" assumption may not be clinically plausible. As for the "no unmeasured confounders" assumption, it is unlikely that the "common treatment effect" assumption will ever be exactly true. However, of more concern is whether the assumption is likely to be approximately true – that is, that the treatment effect received by switchers can at least be expected to be similar to the effect received by patients initially randomised to the experimental group. The problems associated with the "common treatment effect" assumption are compounded by the fact that it is not possible to definitively test the assumption: while some assessment may be made using trial data (by, for example, estimating the treatment effect received by switchers compared to non-switchers) such analyses are likely to be prone to time-dependent confounding and are therefore unreliable. Hence understanding the mechanism of action of the intervention and eliciting clinical expert opinion on its likely effectiveness at different points of the disease progression pathway is important.

In an attempt to relax the "common treatment effect" assumption, analysts have attempted to apply a multi-parameter version of the RPSFTM. However these have not been successful, with meaningful point estimates for causal effects difficult to determine.[22,30,31]

Use of the RPSFTM method is also made problematic if the comparator treatment used in the RCT is active. The RPSFTM counterfactual survival model requires that patients are either "on treatment" or "off treatment" at any one time. If patients in the control group receive an active treatment followed by supportive care upon treatment failure the "off treatment" category represents more than one type of treatment and the counterfactual survival model is not appropriate unless additional causal parameters are added to the model, but, as stated above, attempts to apply multi-parameter RPSFTMs have not been successful. A standard RPSFTM could still be applied, but several important assumptions about treatment strategies and their effectiveness in the experimental and control groups would be required.[1] Linked to this, the RPSFTM counterfactual survival model assumes that the treatment effect is only received while a patient is "on treatment" – it disappears as soon as treatment is discontinued. The clinical plausibility of this assumption should be considered. If a continuing treatment effect is expected the RPSFTM or IPE methods could be applied assuming a lagged treatment effect, or on a "treatment group" basis – where patients in the experimental group are always considered to be "on treatment" and patients that switch remain "on treatment" from the time of switch until death. This analysis ignores treatment discontinuation times and requires there to be a common treatment effect associated with the sequence of treatments received by patients randomised to the experimental group and the sequence of treatments received by switchers after the point of switch. Any benefits associated with post study treatments will be attributed to the experimental treatment, though similarly any benefits from post-study treatments received by control group non-switchers would be attributed to the

control group. If the post study treatments received in all groups represent realistic treatment pathways this approach may appropriately address the economic evaluation decision problem – particularly if the costs of the post-study treatments are also incorporated within the economic model. Hence such an approach might be considered if the comparator is active, or if a continuing treatment effect is expected.

An additional limitation associated with the RPSFTM method involves recensoring. White *et al.* demonstrate that recensoring is required in order to avoid bias in the estimation of counterfactual survival times.[30] Recensoring is required because a positive or negative treatment effect may increase or decrease the probability that the survival time of an individual is censored, and, where treatment switching occurs, treatment received is likely to be associated with prognosis. This means that counterfactual censoring times may be related to prognosis and may therefore be informative (see Appendix A for more details).[30] Recensoring involves data being recensored at an earlier time-point to avoid informative censoring and is therefore associated with a loss of longer-term survival information. Some observed events will become censored if the recensoring time is shorter than the counterfactual event time. The time-point at which recensoring occurs is related to the magnitude of the estimated treatment effect – the larger the treatment effect the earlier the recensoring time-point. Loss of long-term information is likely to be detrimental to the extrapolation of survival data, which is of particular importance in the context of HTA and economic evaluation due to the requirement to estimate the mean survival advantages associated with novel interventions.[13,14,15,16,32,33] In

addition, recensoring may lead to biased estimates of the "average" treatment effect in circumstances where the treatment effect changes over time, because longer term data on the effect of treatment may be lost.

Finally, it is worthy of note that the RPSFTM typically loses power in the presence of treatment switching, like the ITT analysis. By design, the RPSFTM maintains the significance level associated with the ITT analysis, and therefore confidence intervals are often relatively wide. In comparison, the IPCW is not restricted in this way, but IPCW confidence intervals may also be wide if data are relatively sparse.

Iterative Parameter Estimation algorithm

Branson and Whitehead (2002) extended the RPSFTM method using parametric methods, developing a novel iterative parameter estimation (IPE) procedure.[34] The same accelerated failure time model is used, but a parametric failure time model is fitted to the original, unadjusted ITT data to obtain an initial estimate of the treatment effect. The failure times of switching patients are then re-estimated using this, and this iterative procedure continues until the new estimate is very close to the previous estimate, at which point the process is said to have converged.[34]

The IPE procedure makes similar assumptions to the RPSFTM method – for example the randomisation assumption is made, as is the "common treatment effect" assumption. Therefore, similar limitations exist, including problems with multi-parameter models, active trial parameters and recensoring. An additional assumption is that survival times follow a parametric distribution, and thus it is

important to identify suitable parametric models, which in itself can be problematic.[32]

Alternative "two-stage" methods

In addition to the "standard" adjustment methods described so far, "two-stage" methods might be considered. To our knowledge, these have not yet been used in HTA, but have been described recently in the literature.[1] These methods involve first estimating a treatment effect specific to switching patients, and then using this to derive a counterfactual dataset unaffected by switching. Then a treatment effect specific to patients randomised to the experimental group can be estimated. Robins and Greenland (1994) and Yamaguchi and Ohashi (2004) have previously used such an approach, making use of a structural nested failure time model (SNM) with g-estimation to estimate the treatment effect in switchers.[22,23] The SNM is essentially an observational version of the RPSFTM and attempts to account for time dependent confounding using the "no unmeasured confounders" assumption. It therefore has similar limitations to the IPCW.

A previously unused two-stage approach that does not rely upon g-estimation may provide a good fit to the treatment switching mechanism often observed in oncology RCTs. When switching is only permitted after disease progression, the time of progression can be used as a secondary baseline. Using this secondary baseline a parametric accelerated failure time model (such as a Weibull model) that includes covariates measured at the time of progression can be fitted to the post-progression control group data. This model could then be used to estimate

the effect of switching to the treatment post-progression, by contrasting post-progression survival times in those control group patients who switch post progression with those who do not. The resulting acceleration factor can then be used to "shrink" the survival times of switching patients in order to derive a counterfactual dataset unaffected by switching.

This method effectively recognises that the clinical trial is randomised up until the point of disease progression, but beyond that point it essentially becomes an observational study. This is a simplification of the method used by Robins and Greenland[22] and Yamaguchi and Ohashi[23] and is theoretically inferior to these because no attempt is made to adjust for time-dependent confounding between the point of disease progression and the time of treatment switch – and therefore a strong assumption is made that there is no time-dependent confounding between these time-points. However, this simplified method remains relevant due to the convergence issues which have been shown to lead the more complex two-stage SNM methods to perform poorly in simulation studies, as will be described in Section 5 of this paper. In addition, if switching is likely to happen soon after disease progression any time-dependent confounding associated with the lag between disease progression and treatment switch would be small.

The simplified two-stage method may not be generalisable because it is reliant on the ability to identify a secondary baseline. The key limitation of the method relates to its assumption of no time-dependent confounding between the point of disease progression and the time of treatment switch, but in addition to this other potential limitations exist. Firstly, the method requires that the "no

unmeasured confounders" assumption holds at the point of the secondary baseline – hence data must be available on all baseline and time-dependent prognostic factors for mortality that independently predict switching. In addition, because counterfactual survival times are estimated recensoring is required, which may lead to inaccurate estimates of the average long-term treatment effect if the treatment effect changes over time. However, in the method's favour, it does not require data to be collected on prognostic factors or predictors of switching to be collected at time-points other than baseline and the secondary baseline, and hence data-collection requirements are less burdensome than those associated with the IPCW. Finally, the method does not require the "common treatment effect" assumption, the key limitation associated with RPSFTM and IPE methods.

Use of external data

In some instances it might be possible to estimate OS based upon external data, rather than relying upon confounded RCT data. External trials that incorporated the comparator treatment and that were not confounded by treatment switching may exist, or long-term registry data for the disease in question may be available. While such data sources are valuable, the use of external data may be associated with important limitations. Patient populations may differ between different trials due to inclusion criteria, and standards of care may differ if the trials were undertaken at different times and in different locations. Definitions of disease events may also differ, making it difficult to draw appropriate comparisons between trials. These issues are likely to be exacerbated further if the external data source is a registry rather than a clinical trial. If patient-level data are

available from the external datasets it may be possible to adjust for differences in patient characteristics, allowing more accurate estimates of what counterfactual survival would have been in the control group of the RCT under investigation (such an approach was taken in Technology Appraisal 171 conducted by NICE on lenalidomide for multiple myeloma[35]). However, this requires that all important prognostic variables are available from both the novel clinical trial, and the external trial(s). In the absence of these, different trial populations cannot be adjusted appropriately for comparison. Finally, it may be the case that relevant external datasets do not exist, or that the patient-level data associated with these are not available, hence using external data to adjust survival time estimates in the presence of treatment switching is unlikely to represent a generalisable approach.

However, given the problems created by treatment switching, and the limitations associated with statistical adjustment approaches, attempting to obtain external data to inform estimates of long-term survival represents a worthwhile approach. Manufacturers may consider designing their clinical development programme to facilitate the collection of such data. If steps were taken to enhance the probability that suitable external data were available, the use of such data to estimate what survival times would have been in the absence of treatment switching could become a much more reliable and acceptable approach.

5. Results of simulation studies

Several simulation studies have been undertaken to assess the performance of treatment switching adjustment methods in a range of scenarios. In this section we briefly summarise the key results of these.

Morden *et al.* demonstrated that in circumstances where the "common treatment effect" assumption holds, RPSFTM and IPE methods perform very well, producing very little bias in their estimation of the "true" treatment effect.[2] In comparison, naïve adjustment methods that involved either censoring crossover patients at the point of treatment switch, or excluding them entirely from the analysis, often produced very high levels of bias. Although useful, the study conducted by Morden *et al.* was limited because it only considered scenarios in which the "common treatment effect" assumption held, did not consider time-dependent treatment effects, and did not include IPCW or two-stage methods.

In an extension to the Morden *et al.* study Latimer *et al.* conducted a more extensive simulation study that evaluated the performance of treatment switching adjustment methods across a wide range of scenarios.[1,36] A joint longitudinal and survival model was used to simultaneously generate a time-dependent prognostic covariate and survival times. Parameter values were selected such that simulated survival times were reflective of the type of data often observed in metastatic cancer trials. Different levels were tested for switching proportion, treatment effect, and censoring, as were different switching mechanisms. Results confirmed those found by Morden *et al.* – that is, RPSFTM and IPE methods perform very well when the "common treatment

effect" assumption holds, producing bias equivalent to approximately 0.05-1.02% of the treatment effect, while simple censoring and exclusion methods produce very high levels of bias (up to 171.69% of the treatment effect). In addition to this Latimer *et al.* provided evidence on the comparative performance of relevant methods in scenarios in which their key assumptions did not hold.[36]

The IPCW method represented a substantial improvement compared to simple methods (bias as a proportion of the treatment effect often in the region of 5-9%), but produced higher bias than RPSFTM and IPE methods when the "common treatment effect" assumption held.[36] This was likely to be due to the error associated with applying an observational-based method to a relatively small RCT dataset (with simulated sample size 500), and was in line with findings in a previous simulation study reported by Howe *et al.* that focused upon the IPCW.[37] Bias associated with the IPCW method became much higher (bias as a proportion of the treatment effect often in the region of 30-45%) in scenarios in which the proportion of control group patients that switched treatments increased to approximately 90%, leaving approximately 20 patients in the control group who did not switch.[36] It was also found that excluding a covariate that influenced the probability of treatment switching (thus violating the "no unmeasured confounders" assumption) only had a minimal impact on the bias produced by the method – however, this was likely to be due to the high level of correlation between the simulated prognostic covariates. The IPCW method resulted in substantially lower bias than the simple censoring method, which demonstrated the importance of the "no unmeasured confounders"

assumption, as the IPCW reduces to simple censoring when all confounders are unmeasured.

In scenarios in which the treatment effect received by switchers was approximately 15% lower than the average effect received by patients initially randomised to the experimental group (violating the "common treatment effect" assumption) it was found that the RPSFTM, IPE and IPCW methods produced similar levels of bias in their estimates of the treatment effect, with all producing bias equivalent to approximately 5-10% of the treatment effect.[36] In scenarios where the treatment effect received by switchers was approximately 25% lower than the average effect received by patients initially randomised to the experimental group the IPCW method produced lower bias than the RPSFTM and IPE methods (which often produced bias of over 10%) and in these scenarios the ITT analysis often produced least bias (0-5%) if the treatment effect was relatively low (equivalent to a hazard ratio (HR) of approximately 0.75 in experimental group patients).[36] This is logical, because in these scenarios patients who switch receive very little benefit from the experimental treatment.

In addition to the "standard" treatment switching adjustment methods described so far, Latimer *et al.* tested two "two-stage" methods – a structural nested model (SNM) with g-estimation and a simplified two-stage approach that used a Weibull parametric model. The two-stage SNM performed relatively poorly, particularly when switching proportions were very high.[36] The simple Weibull two-stage method performed much better, producing relatively low bias across all scenarios. Bias was higher in scenarios with very high switching proportions, but even in these rarely exceeded 6%. The simplified two-stage

method generally produced lower bias and was much less sensitive to the switching proportion than the IPCW method – perhaps reflecting its lower data and modelling requirements. While RPSFTM and IPE methods produced marginally less bias than the two-stage Weibull method when the "common treatment effect" assumption held, the opposite was true when that assumption was violated.

However, the authors note that the results associated with the simple two-stage Weibull method should be interpreted with some caution because it was well suited to the switching mechanism incorporated within the simulation study – in particular, switching could only occur soon after disease progression.[36] However, it is noteworthy that the switching mechanism simulated was similar to that often observed in metastatic cancer trials, and thus the good results associated with the simple two-stage Weibull method should not be ignored. This method is appears to be worthy of consideration in situations in which treatment switching can only occur after an identifiable secondary baseline, where switching occurs soon after that secondary baseline, where data on important prognostic factors are available at that secondary baseline and where RPSFTM, IPE and IPCW methods seem inappropriate.

In a follow-up to their first simulation study Latimer *et al.* undertook a second study that tested different data generation models for the simulated data, and which considered further scenarios based around reduced switching proportions and sample sizes, and higher censoring proportions, in order to address limitations associated with their initial study.[38] In general, results supported those found in the authors' previous study. New findings were that all

methods generally produced higher bias when the simulated sample size was smaller, when the censoring proportion was higher, and when the switching proportion was lower – although the effects of these changes were often small. The authors note, however, that the adjustment methods are reliant on the absolute number of control group patients who do not switch treatments – and when this number becomes very small adjustment methods become much more prone to bias. In trials with lower sample sizes critically low numbers of non-switching control group patients will be reached with lower switching proportions. For instance, with a control group sample size of 200, a switching proportion of 90% leaves 20 patients in the control group who do not switch. With a control group sample size of 100 the same number of control group non-switchers is left with a switching proportion of only 80%.

Importantly, the authors noted that levels of bias were generally lower in their follow-up study, compared to those found in their initial study; for the RPSFTM, IPE, IPCW and two-stage methods levels of bias rarely exceeded 2-3% of the treatment effect across all scenarios.[38] A key reason for this was the lower switching proportions simulated (20-50%, compared to 50-95% in the initial study), but the authors also noted that differences in the size of the acceleration factor were important. Survival time treatment effects are usually summarised as a hazard ratio (HR), which is the ratio of the hazards of the event of interest in the control group and the experimental group. An HR of lower than 1 means that being randomised to the experimental group reduces the hazard of the event. Survival time treatment effects can also be presented as an acceleration factor (AF), which works on the time scale and denotes the extent to which time to the

event of interest is accelerated. An AF of greater than 1 means that being randomised to the experimental group extends the time until the event occurs.

In both their simulation studies, Latimer *et al.* aimed to investigate two levels of treatment effect – scenarios in which the treatment effect on overall survival was high (equivalent to an average HR of 0.50), and scenarios where the effect was more moderate (equivalent to an average HR of 0.75).[36,38] However, the different data generating models used in the two studies meant that the simulated survival time distributions had different shapes – which makes it possible for the same hazard ratio to be associated with a substantially different acceleration factor. In the first simulation study the average AF across all scenarios varied between 1.44 and 3.58, and was over 2.0 in 60 of the 72 scenarios. In the second study the AF across scenarios ranged between 1.22 and 1.78, despite the fact that data were simulated that produced similar average hazard ratios in the two studies.

Latimer *et al.* note that both of their simulation studies indicate that the performance of each of the switching adjustment methods is affected by the size of the treatment effect – particularly the IPCW, which produced more bias when the treatment effect was higher.[38] The authors suggest that due to this, it is important to assess the size of the treatment effect not only in terms of a hazard ratio, but also in terms of an acceleration factor, since this might more accurately predict the scope for bias associated with the adjustment methods. For instance, in a situation where the hazard ratio is in the region of 0.50, the scope for bias in adjustment methods may appear to be relatively high. However, if the associated

acceleration factor is only around 1.5, the scope for bias is considerably less than if the AF is greater than 2.0.

A consideration of the AF is particularly important for RPSFTM and IPE methods, due to their use of an accelerated failure time counterfactual survival model and their reliance on the "common treatment effect" assumption. These methods result in bias when the absolute difference in the treatment effect received by treatment switchers and patients initially randomised to the experimental group is important. Because the methods use an accelerated failure time model framework, the common treatment effect referred to relates to an acceleration factor. Clearly, the scope for an important violation of the common treatment effect assumption reduces as the AF tends towards 1.0.

To highlight this, in the second simulation study reported by Latimer *et al.* it was found that the RPSFTM and IPE methods generally produced low (in the region of 1-2%) bias, and slightly less bias than the IPCW methods even when the treatment effect received by switchers was 20% lower than that received by patients randomised to the experimental group. This is likely to be because the true AF was relatively low compared to the authors' first simulation study, where it was found that decrements in the treatment effect of 20% or more were associated with very significant increases in bias associated with the RPSFTM and IPE methods.

Summary

To summarise, simulation studies have shown that a number of factors need to be taken into account when assessing the likely performance of treatment

switching adjustment methods.[2.36,37,38] RPSFTM and IPE methods are likely to produce very low levels of bias when the "common treatment effect" assumption holds, and even if this assumption is unlikely to hold these methods may produce low levels of bias if the acceleration factor associated with treatment is likely to be low (in the region of 1-1.8). However, if the acceleration factor is higher (in the region of 2.0-4.0) and switchers are likely to receive a reduced treatment effect, bias associated with the RPSFTM and IPE methods may increase substantially.

The IPCW method is generally prone to higher levels of bias than the RPSFTM methods when applied to relatively small RCT datasets (with sample size approximately 300-500). When switching proportions are very high, leaving less than approximately 20 patients in the control group who do not switch treatments, the IPCW becomes prone to much higher levels of bias.

Simple two-stage methods appear to produce low levels of bias across a wide-range of scenarios, and are much less sensitive to the switching proportion than the IPCW. However, this finding is restricted to scenarios in which switching can only happen very soon after disease progression.

Simple adjustment approaches, such as censoring switchers at the point of switch, or excluding them entirely from the analysis, are prone to extreme bias when switching is associated with prognosis.

6. Case studies

Workshop participants have kindly submitted case studies, to provide real-world examples of the impact of treatment switching. The submitted case studies are summarised briefly here; more details on each are provided in Background Paper 2, which consists of detailed report templates completed for each case study by the workshop participants.

Overview of the submitted case studies

Eight case studies were submitted by the meeting participants to the scientific committee. A brief summary of each case study is given below. Case studies 3, 4 and 8 have been selected for focussed discussion during the meeting. These provide a mix of trial characteristics with respect to the disease, comparator treatment and the switching proportion. These cases also cover a variety of adjustment methods, including some use of external datasets. The issue of switching to "other" post-study treatments is also highlighted, and these studies allow us to see the response of a variety of agencies to the use of adjustment methods.

Whilst cases (iii), (iv) and (viii) will provide the focus of the case study session during the workshop, additional issues highlighted by the other case studies will also be highlighted. These include:

- Difficulties obtaining the clinical data to specifically perform adjustment analysis that had not been pre-specified and would be used only for reimbursement issues (case study (vii)).

- Adjustment methods were presented to the HTA body but not the regulator in the dabrafenib examples (case studies (i) and (ii)).
- Problems associated with trial design including the lack of long term survival data / power to compare survival outcomes, rather than the crossover (case studies (v) and (vi)).

i. Dabrafenib for advanced melanoma. Submitted to Therapeutic Goods Administration (TGA) Australia, September 2012. Decision made August 2013.

Dabrafenib is a treatment for advanced melanoma. In the BREAK-3 trial, patients were randomised 3:1 to dabrafenib (n=187) and dacarbazine (n=63). Around 50% of the patients crossed over onto dabrafenib. The initial trial design allowed patients to crossover after disease progression. In this case study there was no attempt to apply complex statistical methods to correct for bias created by treatment switching. The analysis was based upon PFS as the primary end point and relies upon the assumption that OS benefits are consistent with PFS benefits. The switching protocol was amended during the trial to allow patients to switch before disease progression, and therefore estimated PFS treatment effects may also be subject to crossover bias.

ii. Dabrafenib for advanced melanoma. Submitted to Institut fur Qualitat und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Germany, October 2013. Decision made January 2014

This case study is also based on the BREAK-3 trial (dabrafenib vs dacarbazine) therefore the characteristics of the trial data are identical to case study (i). In this case study, a naïve adjustment which censored patients at the point of crossover was applied. More complex RPSFTM and IPE methods were also used. The IPCW was not applied because no events were observed in the non-switching control group patients. A key issue in this case study was that the crossover adjustment analyses were not accepted by the HTA agency because the underlying methodological assumptions were deemed to be strong and not justified in the manufacturer's dossier.

iii. Everolimus for the second-line treatment of advanced renal cell carcinoma. Submitted to the National Institute for Health and Care Excellence (NICE) (TA219), England and Wales, October 2009. Decision made November 2010

In the RECORD-1 trial, patients were randomised 2:1 to everolimus (n=277) and placebo (n=139). 81% of placebo patients had switched onto everolimus at the most recent analysis. Crossover was allowed after a decision to terminate the double-blind phase of the trial at an interim analysis. IPCW and RPSFTM analyses were performed. The manufacturer favoured the IPCW method, but it was incorrectly applied in the initial analysis and some of the model's key assumptions were not considered. The independent Evidence Review Group preferred the RPSFTM, a preference which was justified on the basis that the "no unmeasured confounders" assumption required by the IPCW could not be satisfied. The fact that the RPSFTM had been used in previous submissions acted as further justification, but the validity of the key assumptions in the context of

this trial were not discussed. The NICE Appraisal Committee agreed that it was appropriate to adjust the ITT analysis to control for crossover, but noted that this increased the level of uncertainty present in the analysis. The Appraisal involved 5 Appraisal Committee meetings, compared to the usual 2.

iv. Vemurafenib for metastatic melanoma. Latest data-cut December 2012. Roche

In the BRIM-3 trial, patients were randomised 1:1 to vemurafenib (n=337) and dacarbazine (n=338). Switching to vemurafenib was permitted after the December 2010 interim analysis showed evidence of efficacy. Although disease progression was not specified as a pre-requisite for switching, most crossover patients had progressed prior to their switch. At the most recent data-cut (December 2012) 25% of control group patients had switched to vermurafenib, and 34% had switched to any BRAF inhibitor (including vemurafenib). Making adjustments for switches to other BRAF inhibitors was a methodological issue highlighted by the manufacturer. The methods applied to adjust for crossover were censoring at time of crossover, a "discount" method, and the RPSFTM. Censoring was argued to be non-informative because the data monitoring board placed no restrictions on when patients could cross over. The discount method involved presenting five OS analyses with a range of assumptions about the treatment effect after switching. The acceptance of the adjustment methods applied varied internationally across different regulatory and HTA bodies.

v. Ruxolitinib for myelofibrosis. Submitted to the pan-Canadian Oncology Drug Review (pCODR), Canada, February 2012. Decision made August 2012

This case study describes 2 trials, COMFORT-1 and COMFORT-2. The COMFORT-1 trial compared ruxolitninib (n=155) to placebo (n=154) and COMFORT-2 compared ruxolitninib (n=146) to best available care (n=73). Both trial designs permitted crossover at a specified disease progression stage relative to the baseline. Crossover was also permitted in the COMFORT-2 trial after the efficacy of ruxolitinib had been demonstrated. A naïve crossover adjustment was applied to COMFORT- 1; patients who crossed over were deemed to have had no response to initial treatment, but no adjustment for crossover was made in COMFORT- 2. The HTA agency Methods Panel conducted a critical appraisal of a matched historical control analysis to assess long-term efficacy, survival and safety of ruxolitinib. The panel noted that due to limitations, conclusions from this should be drawn with caution. It appears that the main problem for the HTA agency was the lack of evidence of survival benefit from the key trials. One cause of this was crossover, but also neither trial was designed to detect a survival difference. The Appraisal Committee was unable to conclude that ruxlotinib improves OS.

vi. Everolimus for pancreatic neuroendocrine tumours. Submitted to pCODR, Canada, February 2012. Decision made August 2012

In the RADIANT-3 trial, patients were randomised 1:1 to everolimus (n=207) or placebo (n=203). Crossover was included in the trial design and was not

permitted until the disease had progressed. Around 75% of control group patients switched onto everolimus after disease progression. No adjustment methods were applied to correct for the crossover bias in OS estimates. Other post-progression treatments also presented an issue – 37.7% of patients in the everolimus arm received antineoplastic therapies, compared to 28.6% in the control arm. The HTA agency noted that OS estimates would be confounded by crossover, and also that the short duration of the trial presented limitations in terms of modelling the potential long-term benefits in OS and PFS.

vii. Sunitinib for second line treatment of gastrointestinal stromal tumour. Reimbursement submission to Pharmaceutical Benefits Advisory Committee (PBAC), Australia, 2009. Pfizer

In the A618-1004 trial, patients were randomised 2:1 to sunitinib (n=243) or placebo (n=118). 87% of placebo patients switched onto sunitinib. Crossover was planned to occur following disease progression, but at the time of the first interim analysis the data safety monitoring committee recommended that all patients who remained on placebo should switch on to sunitinib, based upon its superiority. At the first interim analysis, the OS HR was 0.49, compared to 0.88 at the final analysis. In the PBAC submission, the manufacturer presented several analyses. The base case was the interim analysis. This was supported by an RPSFTM analysis (HR=0.51); analyses of the HR at an early time point and at points through time (with crossover proportions gradually increasing); an OS Kaplan-Meier curve presented based upon 15 patients who did not crossover; and external data to compare median OS estimates. This multi-faceted approach appeared to be accepted by PBAC. An interesting practical issue faced by the

manufacturer was obtaining access to patient data for the purposes of analysis solely for reimbursement and performing analyses that were not pre-specified which could potentially appear in the public domain.

viii. Gefitinib for first line treatment of EGFR mutation advanced non-small cell lung cancer. Submitted to NICE, England and Wales, January 2010, guidance made in July 2010

In the IPASS trial, patients were randomised to gefitinib (n=132) or carboplatin-paclitaxel (n=129). 64.3% of control group patients switched onto gefitinib or another tyrosine-kinase inhibitor (TKI), and some patients had multiple lines of therapy in both arms. At the time the clinical trial took place, gefitinib was approved for second-line therapy in all countries that took part in the study. Several methods were used to adjust for the crossover, in order to isolate the effect of gefitinib on overall survival, compared to a carboplatin-paclitaxel treatment pathway that did not include receipt of a TKI. Adjustments were not made for second-line TKIs received in the gefitinib arm of the trial. One analysis excluded crossover patients, and the RPSFTM and IPCW methods were also used. The manufacturer states that the characteristics of the trial and the observed crossover matched the assumptions of the simple exclusion approach well, because there was no evidence of major selection bias. For the IPCW method the main issue was that key time-dependent covariate data were only collected up until the first disease progression. For the RPSFTM method the issue was that the ITT HR was 1.00, described by the manufacturer as representing the method's "blind spot". Historical control data were available to provide evidence on the outcomes associated with 1st line doublet chemotherapy treatment prior

to the introduction of TKIs, and real world observational studies of OS before and after the introduction of gefitinib are now available to serve as extrinsic validation. The analysis in which switchers were excluded has been provided to NICE as part of the ongoing multiple technology appraisal of second-line TKI in advanced non-small cell lung cancer.

ix. Bortezomib for relapsed multiple myeloma. Prepared for re-review by NICE but not needed, 2011

In the APEX trial, patients were randomised 1:1 to bortezomib (n=333) and dexamethasone (n=336). 71% of dexamethasone patients switched onto bortezomib. Crossover was planned to occur after disease progression, but after an interim analysis all control group patients were offered crossover, regardless of disease status. Approximately 75% of crossover occurred prior to the interim analysis. The ITT analysis resulted in an OS HR of 0.77. Post-hoc, IPE, RPSFTM, and IPCW analyses were undertaken, as was a naïve analysis that included treatment as a time-dependent covariate in a Cox regression model. IPE and RPSFTM results were consistent and provided an adjusted HR of 0.59. IPCW analyses were not presented, but it is stated that problems were encountered when applying the method due to heavy censoring and high weights estimated for control group non-crossover patients.

7. Recommendations made by regulatory and HTA agencies

In order to provide an overview of recommendations made by regulatory bodies and HTA agencies around the world on methods to address the treatment switching issue, we requested that workshop participants provide us with any

relevant guidance that they are aware of. At present, we believe specific reference to the treatment switching problem is only made by NICE in its Guide to the Methods of Technology Appraisal.[13] A technical support document on the topic has also been published by NICE's Decision Support Unit (DSU), but this does not constitute NICE guidance.[39] PBAC is currently developing a document on the topic, a draft version of which has been provided to stimulate discussion during the Workshop.[40] Whilst there is considerable agreement between the DSU and PBAC documents, it is notable that the PBAC analysis framework takes a broader perspective, with less emphasis on statistical methods for adjusting for switching (although these retain an important role). Relevant excerpts from the NICE Methods Guide and the DSU and PBAC documents are provided as Background Papers 3, 4 and 5.

It is worthy of note that IQWiG have recently held an "In Dialogue" session on the topic of treatment switching. The aim of these sessions is to offer representatives from science, industry and IQWiG the opportunity for scientific and technical discussion on various topics related to the work of the institute. There has been no output as a result of the session, but presentations made during the session (in German) and an English summary of the session are available from: https://www.iqwig.de/en/events/iqwig-in-dialogue/iqwig-in-dialogue-2014.6046.html

8. Potential gaps – areas not currently covered by guidance

NICE explicitly states that analyses that adjust for treatment crossover can be presented alongside an ITT analysis when the ITT analysis is likely to be

confounded and therefore inappropriate.[13] Using simple censoring or exclusion adjustment methods is discouraged, but "acceptable" methods are not named. It is stated that chosen methods must be justified in relation to the specific characteristics of the dataset in question, taking into account the mechanism of crossover used in the trial, the availability of data on baseline and time-dependent characteristics, and expectations around the treatment effect if the patients had remained on the treatment to which they were allocated.[13] This would appear to allude to the key assumptions associated with RPSFTM, IPE, IPCW and two-stage methods, and that these should be considered on a case-by-case basis. Thus the NICE guidance is focused upon justifying the use of adjustment methods. This is also true of the DSU technical support document and the PBAC guide, whilst the PBAC guide also recommends analyses to assess whether the treatment switching is likely to have materially affected the comparative treatment effect on OS.[39,40] In addition, whilst the DSU document addresses the use of external data, it does not consider this in detail. Hence, it may be considered that guidance on treatment switching is lacking in the following areas:

- How to design clinical development programmes to ensure the treatment switching problem can be appropriately addressed at a later date. This includes:
    - Commissioning of additional studies
    - RCT design, regarding data collection and switching protocols
- How to appropriately use external datasets to address the treatment switching problem.

It is anticipated that these and other issues will be discussed during the Workshop.

References

[1] Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL, Campbell MJ. Adjusting survival time estimates to account for treatment switching in randomised controlled trials – an economic evaluation context: Methods, limitations and recommendations. Med Decis Making, January 21, 2014, doi:10.1177/0272989X13520192

[2] Morden JP, Lambert PC, Latimer NR, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Med Res Methodol. 2011; 11.

[3] Zhang L, Ko CW, Tang S, Sridhara R. Relationship between progression-free survival and overall survival benefit: A simulation study. Therapeutic Innovation & Regulatory Science 2013;47:95.

[4] Fojo T, Amiri-Kordestani L, Bates SE. Potential pitfalls of crossover and thoughts on iniparib in triple-negative breast cancer. J Natl Cancer Inst 2011;103:1738-1740.

[5] O'Brien S. Ibrutinib CLL Trial: Where is the Equipoise? The Asco Post, May 1, 2013; 4; 7.

[6] Schuster SJ. SIDEBAR: The Ethical Imperative of Clinical Equipoise. The Asco Post, May 1, 2013; 4; 7.

[7] Levis B. Clinical Trials, Crossover, and Clinical Equipoise: A Patient's Perspective. The Asco Post, September 15, 2013; 4; 15.

[8] U.S.Department of Health and Human Services Food and Drug Administration. Guidance for Industry: Clinical trial endpoints for the approval of cancer drugs and biologics. Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research, editors. 2007.

[9] Committee for Medicinal Products for Human Use (CHMP). Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man (CHMP/EWP/205/95 REV.3). Methodological considerations for using progression-free survival (PFS) as primary endpoint in confirmatory trials for registration. 201. European Medicines Agency.

[10] Davis SM, Koch GG, Davis CE, LaVange LM. Statistical approaches to effectiveness measurement and outcome-driven re-randomizations in the clinical antipsychotic (CATIE) studies. Schizophrenia Bulletin 2003;29(1):73-80.

[11] Habermann TM, Weller EA, Morrison VA, Gascoyne RD, Cassileth PA, Cohn JB et al. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. Journal of Clinical Oncology 2006;24;19:3121-3127.

[12] European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man. EMA/CHMP/205/95/Rev.4 Oncology Working Party, December 2012.

[13] National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: NICE, 2013

[14] Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford University Press Inc., New York, 2006

[15] Gold M.R, Siegel J.E, Russell L.B, Weinstein M.C. Cost-effectiveness in health and medicine. Oxford University Press, Inc., New York, 1996

[16] Canadian Agency for Drugs and Technologies in Health, Guidelines for the economic evaluation of health technologies: Canada, 3rd Edition, 2006

[17] Lee Y, Ellenberg J, Hirtz D, Nelson K. Analysis of clinical trials by treatment actually received: is it really an option? Stat Med. 1991;10:1595–1605.

[18] Horwitz R, Horwitz S. Adherence to treatment and health outcomes. Arch Intern Med. 1993;153:1863–1868.

[19] National Institute for Health and Care Excellence. Pazopanib for the first line treatment of metastatic renal cell carcinoma, TA 215. 2011. London, NICE.

[20] National Institute for Health and Care Excellence. Everolimus for the second-line treatment of advanced renal cell carcinoma, TA219. 2011. London, NICE.

[21] Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000; 56(3):779-788.

[22] Robins JM, Greenland S. Adjusting for Differential Rates of Prophylaxis Therapy for PCP in High-Dose Versus Low-Dose AZT Treatment Arms in An AIDS Randomized Trial. J Am Statist Assoc. 1994; 89(427):737-749.

[23] Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials. Part I: Structural nested models and marginal structural models to test and estimate treatment arm effects. Stat Med. 2004; 23(13):1991-2003.

[24] Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. J Am Statist Assoc. 2001; 96(454):440-448.

[25] Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, editors. Statistical Models in Epidemiology: The Environment and Clinical Trials. New York: Springer-Verlag; 1999. 95-134.

[26] Bond M, Hoyle M, Moxham T, Napier M, Anderson R. The clinical and cost-effectiveness of sunitinib for the treatment of gastrointestinal stromal tumours: a critique of the submission from Pfizer. 2009. Peninsula Technology Assessment Group, Universities of Exeter and Plymouth, commissioned by the NHS R&D HTA Programme on behalf of the National Institute for Clinical Excellence.

[27] Robins JM, Tsiatis AA. Correcting for Noncompliance in Randomized Trials Using Rank Preserving Structural Failure Time Models. Commun Stat Theory Methods. 1991; 20(8):2609-2631.

[28] Senn SJ. Covariate imbalance and random allocation in clinical trials. Stat Med. 1989;8(4):467-75.

[29] Hampson LV, Metcalfe C. Incorporating prognostic factors into causal estimators: a comparison of methods for randomised controlled trials with a time-to-event outcome. Stat Med. 2012;31(26):3073-88.

[30] White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: Examples from the Concorde trial. Stat Med. 1999; 18(19):2617-2634.

[31] Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials. Part II: An application in a clinical trial of unresectable non-small-cell lung cancer. Stat Med. 2004; 23(13):2005-2022.

[32] Latimer NR. Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data: Inconsistencies, limitations and a practical guide. Med Decis Making. published online 22 January 2013.

[33] Guyot P, Welton NJ, Ouwens MJ, Ades AE. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. Value Health. 2011;14:640-6.

[34] Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. Stat Med. 2002; 21(17):2449-2463.

[35] Hoyle, M., Rogers, G., Garside, R., Moxham, T., Stein, K. The clinical-and cost effectiveness of lenalidomide for multiple myeloma in people who have received at least one prior therapy: an evidence review of the submission from Celgene. Submission to NICE As Part of STA Program 2008

[36] Latimer, N.R., Abrams, K.R., Lambert, P.C., Crowther, M.J., Wailoo, A.J., Morden, J.P. et al. Adjusting for treatment switching in randomised controlled trials – a simulation study. University of Sheffield Health Economics and Decision Science Discussion Paper No 13/06 2013.

[37] Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of Inverse Probability-of-Censoring Weights in Estimating Survival in the Presence of Strong Selection Bias. Am J Epidemiol. 2011; 173(5):569-577.

[38] Latimer, N.R., Abrams, K.R., Lambert, P.C., Crowther, M.J., Morden, J.P. Assessing methods for dealing with treatment crossover in clinical trials: A follow-up simulation study. University of Sheffield Health Economics and Decision Science Discussion Paper No 14/01 2014.

[39] Latimer N, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. Report by the Decision Support Unit. July 2014.

[40] Pharmaceutical Benefits Division, Department of Health. Examining the consequences of bias caused by treatment switching in randomized trials for submissions to the Pharmaceutical Benefits Advisory Committee. Document prepared for consultation purposes only. October 2014.

[41] Robins JM. Structural Nested Failure Time Models. Andersen PK, Keiding N, editors. Survival Analysis. 4372-4389. 1998. Chichester, UK, John Wiley and Sons. The Encyclopedia of Biostatistics. Armitage, P. and Colton, T.

[42] Mark SD, Robins JM. A Method for the Analysis of Randomized Trials with Compliance Information - An Application to the Multiple Risk Factor Intervention Trial. Controlled Clinical Trials 1993; 14(2):79-97.

Appendix A: Adjustment method details

IPCW

Robins and Finkelstein (2000) recommend using "stabilised" inverse probability of censoring weights, as these are shown to be more efficient.[21] Unstabilised weights are simply the inverse of the conditional probability of having remained uncensored until time t conditional on baseline and time-dependent covariates, whereas stabilised weights are the conditional probability of having remained uncensored until time t given baseline covariates, divided by the conditional probability of having remained uncensored until time t given baseline and time-dependent covariates. The stabilised weight will be equal to 1 for all t if the histories of the included prognostic factors for failure do not impact upon the hazard of censoring at t – thus there would be no informative censoring and treatment switching would be random.[21]

Formally, the stabilised weights applied to each individual for time interval (t), as specified by Hernan et al (2001) are:[24]

$$\widehat{W}(t) = \prod_{k=0}^{t} \frac{Pr[C(k)=0|\bar{C}(k-1)=0,\bar{A}(k-1),V,T>k]}{Pr[C(k)=0|\bar{C}(k-1)=0,\bar{A}(k-1),\bar{L}(k),T>k]} \quad \text{[A1]}$$

where $C(k)$ is an indicator function demonstrating whether or not informative censoring (switching) had occurred at the end of interval k, and $\bar{C}(k-1)$ denotes censoring history up to the end of the previous interval (*k-1*). $\bar{A}(k-1)$ denotes an individual's treatment history up until the end of the previous interval *(k-1)*, and V is an array of an individual's baseline covariates. $\bar{L}(k)$ denotes the history of an individual's time-dependent covariates measured at or

prior to the beginning of interval *k*, and includes *V*. Hence the numerator of [A1] represents the probability of an individual remaining uncensored (not switched) at the end of interval *k* given that that individual was uncensored at the end of the previous interval *(k-1)*, conditional on baseline characteristics and past treatment history. The denominator represents that same probability conditional on baseline characteristics, time-dependent characteristics and past treatment history. When the cause of informative censoring is treatment switching, past treatment history is removed from the model because as soon as switching occurs the individual is censored.

The IPCW adjusted Cox hazard ratio (HR) can be estimated by fitting a time-dependent Cox model to a dataset in which switching patients are artificially censored. The model includes baseline covariates and uses the time-varying stabilised weights for each patient and each time interval. Robust variance estimators or bootstrapping should be used to estimate confidence intervals.[24,25]

RPSFTM

An accelerated failure time counterfactual survival model such as that presented by Robins (1998) is used:[41]

$$U = \int_0^T exp[\psi A_i(t)]dt \quad [A2]$$

where *U* is the counterfactual survival time for each patient, which is a known function of observed survival time (*T*), observed treatment (*A(t)*, where *A(t)* is a

binary time-dependent variable equal to 1 or 0 over time), and the unknown treatment effect parameter ψ.

Counterfactual survival time is a sum of observed time spent on treatment and observed time spent off treatment, where time spent on treatment is multiplied by the factor $\exp(\psi)$. The value of the treatment effect (ψ_0) is estimated as the value of ψ for which counterfactual survival is independent of randomised groups. A log-rank or Wilcoxon test can be used for the RPSFTM g-test in a non-parametric setting, testing the hypothesis that the baseline survival curves are identical in the two treatment groups, or a Wald test could be used for parametric models.[42] The point estimate of ψ is that for which the test (z) statistic equals zero. Because the RPSFTM is a randomisation-based efficacy estimator (RBEE) the p-value from the ITT analysis is maintained.[30]

White *et al.* demonstrate that censoring is problematic for the RPSFTM.[23] A positive or negative treatment effect may increase or decrease the probability that the survival time of an individual is censored, and, where treatment switching occurs, treatment received is likely to be associated with prognosis. In turn, this means that the censoring of counterfactual survival times may depend on prognostic factors and therefore be informative.[30] Bias associated with this can be avoided by recensoring counterfactual survival times at the earliest possible censoring time given the treatment effect ψ.[30] Thus for each patient in treatment groups at risk of switching the recensored censoring time is the minimum of the observed administrative censoring time (C_i) and the product $\exp(\psi)C_i$. If the a patient experienced an event, but the recensoring time is less

than the event time, that patient has their survival time recensored and their event is no longer observed.

IPE algorithm

This method uses the same accelerated failure time model as the RPSFTM, but a parametric failure time model is fitted to the original, unadjusted ITT data to obtain an initial estimate of ψ. The observed failure times of switching patients are then re-estimated using $\exp(\psi)$ and the counterfactual survival time model presented in equation [A2], and the treatment groups are then compared again using a parametric failure time model. This will give an updated estimate of ψ, and the process of re-estimating the observed survival times of switching patients is repeated. This iterative process is continued until the new estimate for $\exp(\psi)$ is very close to the previous estimate (the authors suggest within 10^{-5} of the previous estimate but offer no particular rationale for this), at which point the process is said to have converged.[34] Bootstrapping is recommended to obtain standard errors and confidence intervals for the treatment effect.[34]

Two-stage method

When switching is only permitted after disease progression, but is likely to happen soon after this time-point, we can use the time of disease progression as a secondary baseline in a two-stage analysis. If we assume that all patients are at a similar stage of disease at the point of disease progression, we can estimate the effect of the new treatment on extending survival from the point of disease progression to death, specifically for control group patients who switch. Disease progression is used as the secondary baseline – that is, a time-point from which

analysis time can be reset to zero – and data for patients in the control group can be treated as an observational dataset. By fitting an accelerated failure time model (such as a Weibull model) to this data (excluding patients in the experimental group) including covariates measured at the secondary baseline and including a time-varying covariate indicating treatment switch, an estimate of the treatment effect received by patients who switched compared to control group patients who did not switch can be obtained. This would be expected to produce a reasonable estimate of the treatment effect associated with switching, provided the model fits the data, there are "no unmeasured confounders" at the point of the secondary baseline and provided switching occurs soon after the secondary baseline. Counterfactual survival times for switchers could then be obtained using:

$$U_i = T_{A_i} + \frac{T_{B_i}}{\mu_B} \qquad [A3]$$

Where T_{A_i} represents the time spent on control treatment, T_{B_i} represents the time spent on the new intervention, and $\mu_B$is the treatment effect (acceleration factor) in switching patients.

Appendix B: Workshop delegate list and committee membership

Table B1. Workshop Delegate List

Name	Job Title	Institution
Zoe Armstrong	Executive Director Clinical Research – Australia/New Zealand and East Asia	MSD
Dr Helen Bell	Research Assistant	University of Sheffield
Iain Bennett	Health Economic Statistician	F Hoffmann-La Roche AG
Andrew Bruce	Director, Health Policy and Reimbursement	AMGEN
Tom Burke PhD	Executive Director, Global Health Outcomes, Oncology	MSD
Deborah Collyar	President	Patient Advocates In Research (PAIR)
Dr Michael Coory	Director, Clinical Evaluation Unit 5	Therapeutic Goods Administration
Dane J Dickson, MD	Director of Clinical Science	MolDX, Palmetto GBA (a CMS Administrative Contractor)
Emeritus Professor Anthony Fields	Chair, pCODR Expert Review Committee	Canadian Agency for Drugs & Technologies in Health (CADTH)
Professor Davina Ghersi	Senior Principal Research Scientist	National Health and Medical Research Council (NHMRC)
Mendel Grobler	Director: Strategic Access and Policy	Pfizer
Dr Katarina Hedman, PhD Med, MSc/BSc Math, EMBA	Head, Statistical Innovation	Advanced Analytics Centre, AstraZeneca
Dr Phil Haywood	Health Economist	Centre for Health Economics Research and Evaluation (CHERE)
Dr Chris Henshall	Consultant	
Dr Suzanne Hill	PBAC Chair	Pharmaceutical Benefits Advisory Committee (PBAC)
Professor Michael James	Chief Medical Scientist, Rheumatology Unit, Royal Adelaide Hospital	Bellberry Limited
Dr Daniel Kalanovic, MD	Medical Director Oncology	Pfizer
Dr Nick Latimer	Senior Research Fellow in Health Economics	University of Sheffield
Dr Wendy Lipworth	Senior Research Fellow	University of Sydney
Dr Jennie Louise	Medical Ethicist / Statistician	University of Adelaide
Clare McGrath	Oncology TA Head, Global Payer Evidence , Pricing and HTA Policy	Global Medicines Development, AstraZeneca
Andrew Mitchell	Strategic Adviser, Evaluation	Australian Government Department of Health

Dr Francesco Pignatti	Head of Oncology, Haematology and Diagnostics	European Medicines Agency (EMA)
Adriana Platona	Assistant Secretary, Pharmaceutical Evaluations Branch, Department of Health	Australian Government Department of Health
Dr Klaus Pugner	Executive Director, Global Health Economics	AMGEN
James Reimann	Global Head, Oncology Biostatistics and PCOR	Roche/Genentech
Professor Andrew Roberts	Metcalf Chair of Leukaemia Research	University of Melbourne, WEHI & Melbourne Health
Emeritus Professor Lloyd Sansom AO	Bellberry Board of Directors, Special Advisor Department of Health & Ageing	Bellberry Limited
Dr Prudence Scott	Co-Director, Oncology/ Haematology Evaluation Unit	Therapeutic Goods Administration
Dr Nick Simpson	Director, Clinical Evaluation Unit 4	Therapeutic Goods Administration
Mike Smith	Head, Health Economics	Astrazeneca Australia
Dr Katherine Soltys MD	Manager, Oncology Division 1, Therapeutic Products Directorate, Health Canada	Therapeutic Products Directorate, Health Products and Food Branch, Health Canada
Tim Spelman	Biostatistician	Burnet Institute
Mike Spencer	International Brand Value Team Leader, Ibrutinib	Janssen-Cilag
Kylie Sproston	Chief Executive Officer	Bellberry Limited
Rebecca Trowman Day	Consultant Technical Adviser	National Institute for Health & Care Excellence (NICE)
Sean Tunis, MD,	President and Chief Executive Officer	Center for Medical Technology Policy (USA)
Janet Wale	Health Technology Assessment International (HTAI) Patients & Citizens Chair	Cochrane Collaboration
Professor Robyn Ward AM	Professor of Medicine and Clinical Associate Dean	University of NSW and Prince of Wales Hospital
Dr Beate Wieseler	Head of Drug Assessment	Institute for Quality & Efficiency in Health Care (IQWiG)
Professor Dr Bernhard Wörmann	Medical Director	DGHO German Society for Hematology and Medical Oncology

Table B2. **Scientific Committee Membership**

Name	**Job Title**	**Institution**
Professor Keith Abrams	Professor of Medical Statistics and Head of Unit, Studies of Health Populations and Society	University of Leicester
Meindert Boysen	Programme Director - Technology Appraisals	National Institute for Health and Care Excellence (NICE)
Mendel Grobler	Director: Strategic Access and Policy	Pfizer
Clare McGrath	Oncology TA Head, Global Payer Evidence, Pricing and HTA Policy	AstraZeneca
Francesco Pignatti	Head of Oncology, Haematology and Diagnostics	European Medicines Agency
James Reimann	Global Head, Oncology Biostatistics and PCOR	Roche/Genentech
Professor Lloyd Sansom AO	Emeritus Professor	University of South Australia
Professor Uwe Siebert	Adjunct Professor of Health Policy and Management	Harvard University
Tim Spelman	Biostatistician	Burnet Institute
Mona Sabharwal	Executive Director	pan-Canadian Oncology Drug Review (pCODR), CADTH
Rajeshwari Sridhara	Director, Division of Biometrics V	Food and Drug Administration
Rebecca Trowman Day	Consultant Technical Adviser	National Institute for Health & Care Excellence (NICE)
Sean Tunis, MD,	President and Chief Executive Officer	Center for Medical Technology Policy (USA)
Professor Bernhard Wörman	Medical Director	DGHO German Society for Hematology and Medical Oncology
In attendance:		
Scientific Secretariat		
Nick Latimer	Senior Research Fellow in Health Economics	University of Sheffield
Helen Bell	Research Assistant	University of Sheffield
Others:		
Chris Henshall	Associate Professor	Brunel University London
Kylie Sproston	Chief Executive Officer	Bellberry Limited