

# Analysis of adaptive platform trials using a network approach

Ian C Marschner  and I Manjula Schou

Clinical Trials  
2022, Vol. 19(5) 479–489  
© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/17407745221112001

[journals.sagepub.com/home/ctj](http://journals.sagepub.com/home/ctj)



## Abstract

**Background:** Adaptive platform trials allow randomized controlled comparisons of multiple treatments using a common infrastructure and the flexibility to adapt key design features during the study. Nonetheless, they have been criticized due to the potential for time trends in the underlying risk level of the population. Such time trends lead to confounding between design features and risk level, which may introduce bias favoring one or more treatments. This is particularly true when experimental treatments are not all randomized during the same time period as the control, leading to the potential for bias from non-concurrent controls.

**Methods:** Two analysis methods addressing this bias are stratification and adjustment. Stratification uses only comparisons between treatment cohorts randomized during identical time periods and does not use non-concurrent randomizations. Adjustment uses a modeled analysis including time period adjustment, allowing all data to be used, even from periods without concurrent randomization. We show that these competing approaches may be embedded in a common framework using network meta-analysis principles. We interpret the stages between adaptations in a platform trial as separate fixed design trials. This allows platform trials to be viewed as networks of direct randomized comparisons and indirect non-randomized comparisons. Network meta-analysis methodology can be re-purposed to aggregate the total information from a platform trial and to transparently decompose this total information into direct randomized evidence and indirect non-randomized evidence. This allows sensitivity to indirect information to be assessed and the two analysis methods to be clearly compared.

**Results:** Simulations of platform trials were analyzed using a network approach implemented in the netmeta package in R. The results demonstrated bias of unadjusted methods in the presence of time trends in risk level. Adjustment and stratification were both unbiased when direct evidence and indirect evidence were consistent. Network tests of inconsistency may be used to diagnose inconsistency when it exists. In an illustrative network analysis of one of the treatment comparisons from the STAMPEDE platform trial in metastatic prostate cancer, indirect comparisons using non-concurrent controls were inconsistent with the information from direct randomized comparisons. This supports the primary analysis approach of STAMPEDE, which used only direct randomized comparisons.

**Conclusion:** Network meta-analysis provides a natural methodology for analyzing the network of direct and indirect treatment comparisons from a platform trial. Such analyses provide transparent separation of direct and indirect evidence, allowing assessment of the impact of non-concurrent controls. We recommend time-stratified analysis of concurrently controlled comparisons for primary analyses, with time-adjusted analyses incorporating non-concurrent controls reserved for secondary analyses. However, regardless of which methodology is used, a network analysis provides a useful supplement to the primary analysis.

## Keywords

Adaptive design, indirect treatment comparison, mixed treatment comparison, multi-arm multi-stage study, network meta-analysis, platform trial

## Background

Platform trials provide a framework for undertaking randomized controlled treatment comparisons of potentially many treatments, while flexibly allowing design adaptations as the study progresses.<sup>1,2</sup> Adaptive design changes include adding new treatments,

NHMRC Clinical Trials Centre, University of Sydney, Sydney, NSW, Australia

### Corresponding author:

Ian C Marschner, NHMRC Clinical Trials Centre, University of Sydney, Sydney, NSW 2006, Australia.

Email: [ian.marschner@sydney.edu.au](mailto:ian.marschner@sydney.edu.au)

dropping existing treatments, altering randomization allocations, enrichment to selectively recruit participants likely to respond, and many others. The times of design adaptations define a multi-stage structure for platform trials which motivates an alternative name, multi-arm multi-stage studies (MAMS).<sup>3</sup>

Despite the flexibility of platform trials, they have been criticized because the adapting design may introduce bias if there are time trends in the population risk level.<sup>4,5</sup> Indeed, if treatments are not all randomized during identical time periods, the non-concurrent randomization leads to a mixture of randomized and non-randomized evidence.<sup>4</sup>

Although the design of a platform trial may adapt over time, the multi-stage structure provides a sequence of fixed design trials. The synthesis of information from the multiple stages of a single platform trial is therefore analogous to a meta-analysis of fixed design trials. Our goal is to take advantage of this structure by using meta-analytic principles to overcome some of the limitations of platform trials, and to provide transparent delineation between randomized and non-randomized evidence.

The primary limitation of adapting the study design over time is that key design features become confounded with time. This means that if the background risk of participants is related to time of randomization, comparisons of risk between treatments may be biased. For example, if the underlying risk level of the population decreases over time, and the allocation ratio of an experimental treatment is increased over time relative to the control, then this will bias the controlled comparison in favor of the experimental treatment.

There have been two primary approaches for dealing with this confounding. The first approach, which we will call *adjustment*, uses a modeled analysis including an adjustment for randomization time. An example of a platform trial following this strategy is the REMAP-CAP trial in COVID-19.<sup>4,6</sup> In this trial, the primary analysis uses a time epoch adjustment allowing all data to be used in treatment comparisons, even for periods in which there was no concurrent randomization to the treatments being compared. The second approach, which we will call *stratification*, involves using the stages of a platform trial as time strata, and only using those strata that include concurrent randomization to the treatments being compared. An example of a platform trial using this strategy is the STAMPEDE trial in prostate cancer.<sup>3,7</sup> Advocates of adjustment argue that it allows more information to be extracted, while advocates of stratification argue that it preserves randomization and is, therefore, less susceptible to bias and modeling assumptions.

In this article, we present an approach to analyzing platform trials that embeds both approaches within a common framework. We use network meta-analysis principles, providing an established methodology for

combining direct and indirect comparisons.<sup>8</sup> We show how a platform design may be viewed as a network of direct and indirect comparisons, which may be decomposed into randomized and non-randomized evidence. This allows both the adjusted and stratified approaches to be conveniently compared in a common framework that allows transparent assessment of their consistency.

## Methods

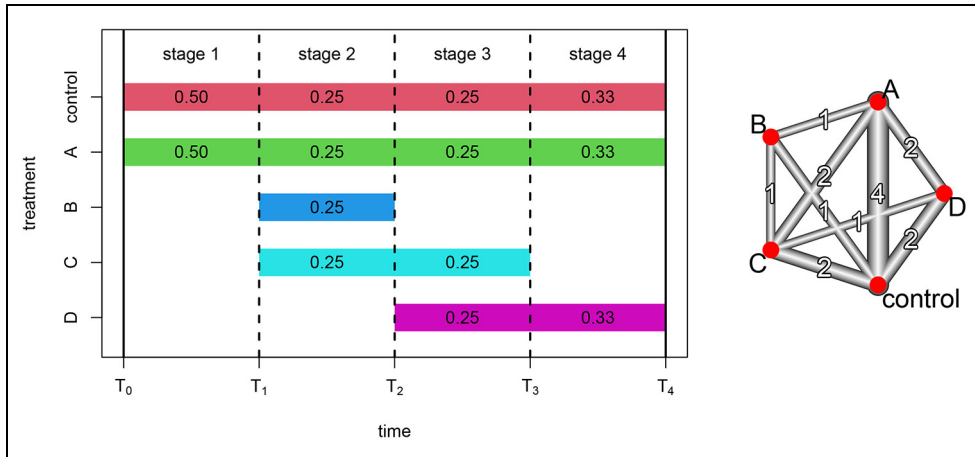
Our proposed analysis method is based on viewing a platform trial as a network of direct and indirect treatment comparisons. The stratified and adjusted approaches described above correspond, respectively, to restricting the analysis to direct comparisons or aggregating the information from the direct and indirect comparisons. The aggregation of direct and indirect treatment comparisons is called a mixed treatment comparison.<sup>9,10</sup> Methods for undertaking mixed treatment comparisons are well-known in meta-analysis and may be re-purposed for the analysis of a single platform trial.

### Mixed treatment comparisons

Figure 1 depicts an illustrative platform trial with four stages and five treatment arms, including one control and four experimental treatments. The study is initiated at time  $T_0$  and adaptations occur at times  $T_1$ ,  $T_2$ , and  $T_3$  followed by termination of the study at time  $T_4$ . The adaptations at times  $T_1$ ,  $T_2$ , and  $T_3$  involve adding and dropping treatment arms and changing the randomization probabilities.

The MAMS structure in Figure 1 is equivalent to four distinct randomized trials with four different fixed designs: (1) control versus A; (2) control versus A, B, and C; (3) control versus A, C, and D; and (4) control versus A and D. Within each stage-specific design, randomization is fully preserved, and the design does not change. Thus, within stages, the concurrent randomization prevents any bias resulting from changes in background risk. However, between stages there are design changes that are confounded with time, so comparisons between stages are susceptible to bias from changes in background risk.

Figure 1 yields a collection of direct and indirect treatment comparisons. A direct comparison compares treatment cohorts concurrently randomized within the same study stage, whereas an indirect comparison compares treatment cohorts randomized during different stages with a reference treatment common to both stages. For example, stage 2 provides a direct comparison of control and treatment B, yielding an effect estimate  $\hat{\theta}_{B0}^{direct}$ . These treatments may also be compared indirectly, using stages 1 and 2 with treatment A as the common reference. If  $\hat{\theta}_{A0}^{direct}$  is the effect estimate from the direct comparison of control and treatment A in stage 1, and  $\hat{\theta}_{AB}^{direct}$  is the effect estimate from the direct



**Figure 1.** Platform trial with four treatments compared to a common control. Design adaptations occur at three timepoints denoted by vertical dashed lines, yielding four stages. Colored bands denote the stages that each treatment is available for randomization, with the corresponding stage-specific randomization probabilities. The network diagram depicts the network and number of direct treatment comparisons available within the platform.

comparison of treatments A and B in stage 2; then, the indirect effect estimate of control versus treatment B is  $\hat{\theta}_{B0}^{indirect} = \hat{\theta}_{A0}^{direct} - \hat{\theta}_{AB}^{direct}$ .<sup>10</sup> The synthesis of the direct estimate  $\hat{\theta}_{B0}^{direct}$  and the indirect estimate  $\hat{\theta}_{B0}^{indirect}$  into a single summary of the comparison between control and treatment B is then a mixed treatment comparison. For appropriately chosen weights  $w_1$  and  $w_2$  based on the variability of the treatment effect estimates, combination of the two estimates then leads to a mixed treatment comparison of treatment B and control

$$\hat{\theta}_{B0}^{mixed} = w_1 \hat{\theta}_{B0}^{indirect} + w_2 \hat{\theta}_{B0}^{direct} = w_1 \hat{\theta}_{A0}^{direct} - w_1 \hat{\theta}_{AB}^{direct} + w_2 \hat{\theta}_{B0}^{direct} \tag{1}$$

From equation (1), the mixed treatment effect estimate is a linear combination of the three direct treatment effect estimates. As we will see, network meta-analysis generalizes this notion to complex networks of treatment comparisons by using linear combinations of direct treatment effect estimates to optimally combine direct and indirect information.

Use of the direct effect estimate  $\hat{\theta}_{B0}^{direct}$  alone is equivalent to a stratified analysis approach for platform studies. Furthermore, as we will see later in the simulation study, use of the mixed comparisons estimate synthesizing  $\hat{\theta}_{B0}^{indirect}$  and  $\hat{\theta}_{B0}^{direct}$  is analogous to an adjusted analysis approach. In the next section, we discuss this synthesis process before going on to explore the difference between the two analysis approaches.

### Network analysis

Figure 1 also displays a network representation of the direct comparisons available within the illustrative platform trial. Vertices represent treatments while edges

represent direct comparisons between the associated treatments. The width and number associated with each edge represent the number of stages providing each direct comparison. Where two vertices are not directly connected, if they can be reached from each other by passing through a third vertex, then an indirect comparison is available with the third treatment as reference. For the network displayed in Figure 1, direct comparison of treatments B and D is unavailable. However, since vertex D can be reached from vertex B via each of the other three treatments, indirect comparisons are available with each of the other three treatments as reference. We now discuss the synthesis of such direct and indirect evidence using network meta-analysis and demonstrate its application to the network representation of platform trials.

Network analysis methods are more straightforward to describe when we assume that there is a common control arm in all stages. This is not an essential assumption but will often be true in platform trials and can be relaxed with additional technical details.<sup>11</sup>

For each stage  $s = 1, \dots, S$ , we have a  $C_s \times K$  design matrix  $X_s$  specifying the treatment contrasts for that stage, where  $C_s$  and  $K$  are the number of treatment contrasts possible in stage  $s$  and the number of non-control treatments, respectively. We also have a  $C_s \times 1$  column vector  $\hat{\theta}_s$  of treatment contrast estimates for stage  $s$  and a  $K \times 1$  column vector of parameters  $\theta$  for the treatment contrasts with the common control. This leads to a total of  $C = \sum_s C_s$  stage-specific treatment

comparisons for the entire study. For example, in Figure 1, there are  $K = 4$  contrasts compared to the control, with  $\theta = [\theta_{A0} \ \theta_{B0} \ \theta_{C0} \ \theta_{D0}]^T$ . Across all stages, there are  $C_1 = 1$ ,  $C_2 = C_3 = 6$ ,  $C_4 = 3$ , and  $C = 16$  contrasts. For stage 1, the estimate and design

matrix are then  $\hat{\theta}_1 = [\hat{\theta}_{A0}]$  and  $X_1 = [1 \ 0 \ 0 \ 0]$ , while for stage 2, they are  $\hat{\theta}_2 = [\hat{\theta}_{A0} \ \hat{\theta}_{B0} \ \hat{\theta}_{C0} \ \hat{\theta}_{AB} \ \hat{\theta}_{AC} \ \hat{\theta}_{BC}]^T$  and

$$X_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}.$$

Using this notation, the linear model for a general platform trial is then specified in terms of a  $C \times 1$  vector of estimates  $\hat{\theta} = [\hat{\theta}_1 \ \hat{\theta}_2 \ \dots \ \hat{\theta}_S]^T$  and a  $C \times K$  design matrix  $\mathbf{X} = [X_1^T \ X_2^T \ \dots \ X_S^T]^T$  using the linear relationship  $\mu(\theta) = E(\hat{\theta}) = \mathbf{X}\theta$ . This linear model may be fitted by weighted least squares, which is implemented by minimizing the weighted sums of squares

$$Q(\theta) = [\hat{\theta} - \mu(\theta)]^T W [\hat{\theta} - \mu(\theta)] \quad (2)$$

for some appropriately chosen  $C \times C$  weight matrix  $W$ . This yields the network estimates of the control contrast parameters  $\theta$ ,

$$\hat{\theta}^{network} = (\mathbf{X}^T W \mathbf{X})^{-1} \mathbf{X}^T W \hat{\theta} \quad (3)$$

from which the effect estimates associated with other contrasts may also be obtained.

Equation (3) is a weighted linear combination of the direct effect estimates and is a generalization of the basic mixed comparison approach exemplified in equation (1). The weight matrix employed in equation (3) is chosen by using the variance–covariance structure implied by the stage-specific treatment effect standard errors and the stage-specific designs, and it is analogous to the inverse variance weights used in regular meta-analysis. This weighing process gives greater weight to effect estimates having smaller standard errors, but it is also necessary to accommodate correlations between treatment comparisons from the same stage. These correlations stem from the fact that a comparison of treatment A with control shares the same control participants as the comparison of treatment B with control, within the same stage. Thus, the two treatment effect estimates are correlated and collectively provide less information than if they had been obtained from different study stages. Full details of this variance–covariance weighting process are described by Schwarzer et al.<sup>11</sup> and references therein. The weight matrix may also be used to quantify the proportion of the overall evidence that is provided by direct treatment comparisons, which is referred to as the direct evidence proportion. This allows decomposition of the full network analysis into its randomized and non-randomized components, facilitating a comparison of the two sources of information.<sup>12</sup>

An important assumption of network meta-analysis is that the underlying difference between two treatments is identical for all direct and indirect comparisons between the two treatments. The concept of inconsistency in network meta-analysis refers to the violation of this assumption and is a useful addition to platform trial analysis methodology. Inconsistency is an extension of the concept of heterogeneity in regular meta-analysis. Generalizations of Cochrane's Q-statistic and  $I^2$  for assessing heterogeneity are available for network meta-analyses and are based on the weighted sums of squares in equation (2).<sup>11</sup> In the context of platform trials, heterogeneity translates to differences in treatment effects between stages that compare the same treatments, while inconsistency translates to an interaction between treatment comparisons and the stage-specific design. This occurs when the magnitude of a treatment effect is modified by the presence (or absence) of other treatments in a particular stage of the platform trial. The detection of such inconsistency may reflect conflict between the direct and indirect evidence in a platform trial and provides information about whether the adjusted or stratified analysis approaches are more appropriate. Comparison of effect estimates based only on direct comparisons with those based only on indirect comparisons is an important aid for this assessment. Application of these network meta-analysis tools to platform trials will be illustrated further in the Results section.

### Implementation

Various software packages are available for network meta-analysis and may be used to implement the analysis methods for platform trials described here. One such package is the netmeta package in R, which we will use to illustrate our proposed analysis approach.<sup>13</sup>

For binary endpoints, data from a platform trial would be organized as events and sample sizes by treatment arm for participants randomized in each stage of the study. In Table 1, we provide illustrative binary data in this format for the illustrative platform design presented in Figure 1. This binary data structure is pre-processed by the function pairwise and then inputted to either the netmetabin or netmeta function for analysis. For other types of endpoints, the stage-specific effect estimates and standard errors are inputted directly to the netmeta function. Data from the STAMPEDE trial will be analyzed in this way using log hazard ratios (HRs). Importantly, participants contribute only to the stage-specific effect estimate from the stage in which they were randomized, regardless of whether the endpoint involves follow-up during a subsequent stage.

Other analytical and graphical functions are also available in the netmeta package. For the purpose of investigating the difference between adjusted and stratified analyses, decomposition of the analysis into direct

**Table 1.** Example data for the platform design shown in Figure 1.

Stage	Treatment	Number randomized	Percent randomized	Events	Risk	RR
1	Control	103	51.5%	54	52.4%	1
1	A	97	48.5%	25	25.8%	0.49
2	Control	83	20.8%	47	56.6%	1
2	A	97	24.3%	27	27.8%	0.49
2	B	111	27.8%	39	35.1%	0.62
2	C	109	27.3%	54	49.5%	0.87
3	Control	95	23.8%	42	44.2%	1
3	A	99	24.8%	21	21.2%	0.48
3	C	109	27.3%	53	48.6%	1.10
3	D	97	24.3%	19	19.6%	0.44
4	Control	102	34.0%	50	49.0%	1
4	A	96	32.0%	27	28.1%	0.57
4	D	102	34.0%	29	28.4%	0.58
All	Control	383	29.5%	193	50.4%	1
All	A	389	29.9%	100	25.7%	0.51
All	B	111	8.5%	39	35.1%	0.70
All	C	218	16.8%	107	49.1%	0.97
All	D	199	15.3%	48	24.1%	0.48

RR: relative risk.

Percent randomized is the percentage of randomizations to each treatment within that stage. Risk is events divided by number randomized.

randomized evidence and indirect non-randomized evidence is particularly important.<sup>12</sup> This facility is provided by the function `netsplit`, applied to the output from a `netmetabin` or `netmeta` analysis, together with a test of inconsistency between the direct and indirect evidence using the separate indirect and direct evidence (SIDE) back-calculation method of König et al.<sup>12</sup> Graphical functions for constructing a network diagram of the design and presenting forest plots of the analysis results are also available and will be illustrated in the Results section. Further details of the `netmeta` package are presented by Schwarzer et al.<sup>11</sup>

## Results

### Simulated analyses

We begin by providing the results of simulated analyses using the four-stage platform design depicted in Figure 1, under various scenarios. The purpose of the simulations is to investigate the bias of an unadjusted analysis in the presence of time trends in background risk, and whether adjustment and stratification can address this bias.

In the first scenario, platform trials were simulated with a binary endpoint that occurs with a risk (probability) of 0.5 in the control arm. The RR compared to control was 1.0 for arms B and C, and 0.5 for arms A and D. The randomizations to each arm occurred with probabilities shown in Figure 1. We refer to this scenario as the *homogeneity* scenario.

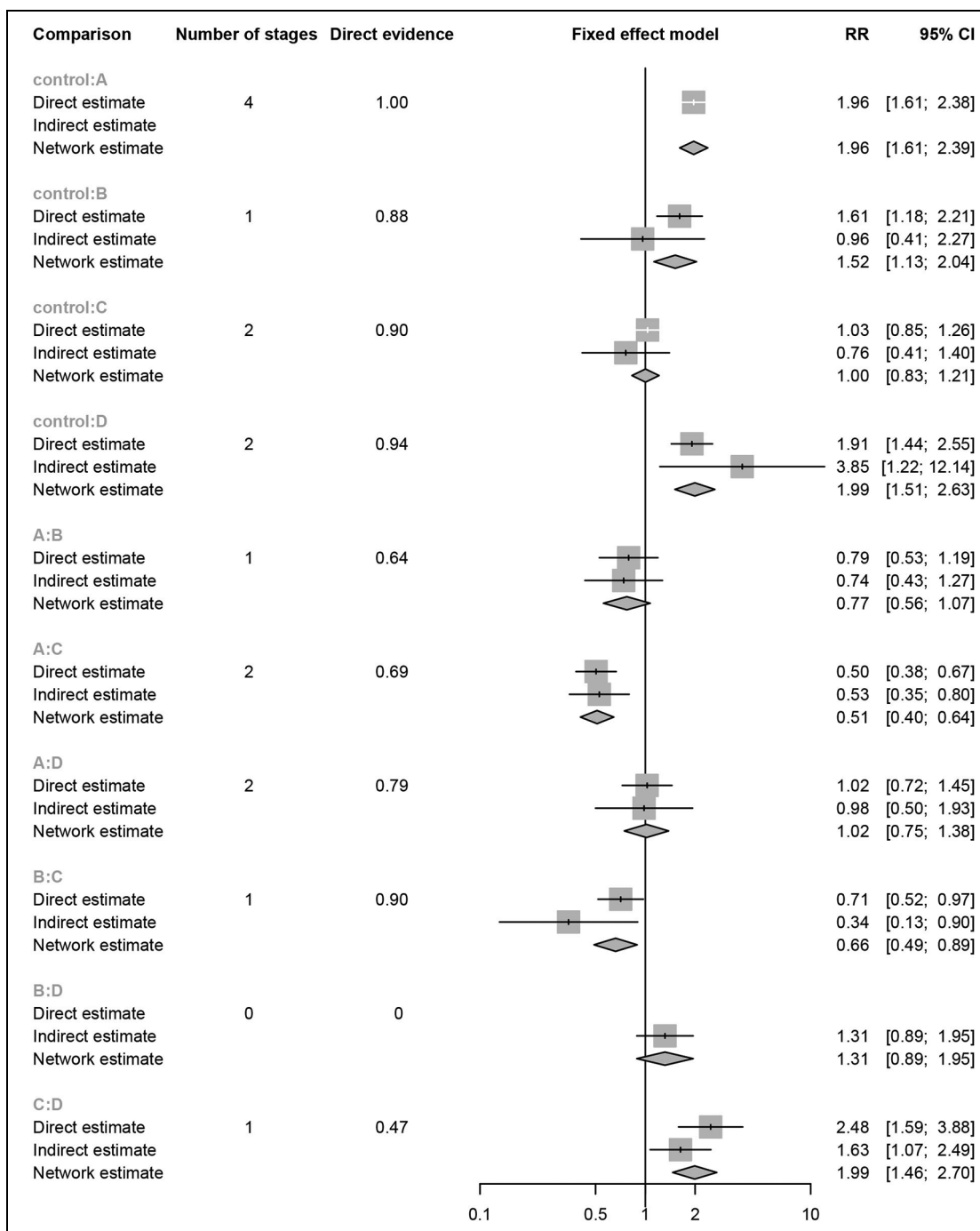
The homogeneity scenario was then altered to provide the *time trend* scenario, in which the risk in the

control arm was assumed to have a decreasing trend over time, beginning with 0.5 in stage 1, followed by 0.4, 0.3, and 0.2 in stages 2 to 4, respectively. In this scenario, the RR for each treatment remained the same as in the homogeneity scenario. For each scenario, 1000 trials were simulated, first with an average of  $n = 100$  randomizations per treatment per stage, and then again with  $n = 200$ .

While the homogeneity and time trend scenarios are the primary simulations discussed here, various other simulations were also conducted. These include a *differential time trend* scenario in which the RR decreases over time and an *inconsistency* scenario in which the RR for one treatment was inconsistent between stages. Full specification of these scenarios is provided in the Supplemental Material.

For each simulated platform trial, the data were analyzed using the `netmeta` package described in the Methods, producing a direct randomized RR estimate for each comparison, an indirect non-randomized estimate, and a combined network estimate. In addition, we computed unadjusted and adjusted RR estimates, the latter from a log-link binary regression model adjusting for a stage effect, which is typically called relative risk regression.<sup>14</sup>

Table 1 provides the full data for the first simulation of the homogeneity scenario with  $n = 100$ . An example network analysis is shown in Figure 2, which is the graphical output from the `netmeta` package applied to the data in Table 1. The most important aspect of Figure 2 is the decomposition of total information across the platform network, into a direct randomized component

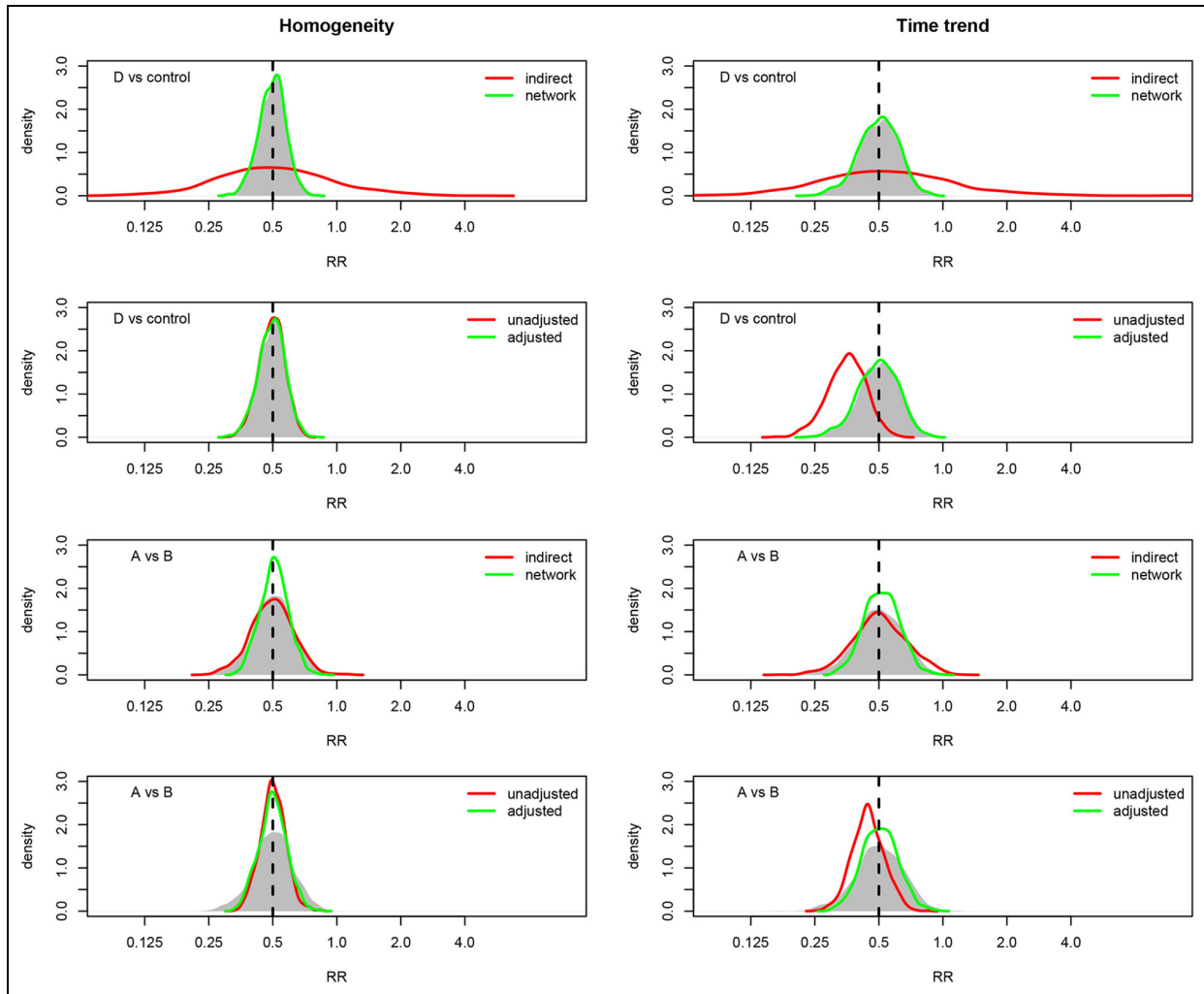


**Figure 2.** Network analysis of the platform trial depicted in Figure 1, with data provided in Table 1. For each possible treatment comparison, the direct randomized estimate, the indirect non-randomized estimate, and the mixed network estimate of the relative risk (RR) and 95% confidence interval (CI) are provided. Also provided for each treatment comparison is the number of stages with direct comparisons and the proportion of overall evidence coming from direct comparisons.

and an indirect non-randomized component. The proportion of information obtained from direct comparison is 1 for control versus A and 0 for B versus D, while all other comparisons have a mixture of direct and indirect evidence.

Figure 3 provides a selection of simulated treatment effect distributions from the various scenarios. First, all

estimation methods are unbiased except for the unadjusted method which has substantial bias when there is a time trend in the control risk. Importantly, this bias is eliminated by the other methods, which is the intention of the stratified and adjusted approaches. It may also be seen that the comparison of control and treatment D is a situation where the indirect information



**Figure 3.** Results of 1000 simulations of the platform design depicted in Figure 1, for the treatment comparisons D versus control and A versus B. For each comparison, the true relative risk (RR) is 0.5 and each treatment has an average of 100 randomizations per stage. In all panels, the shaded area depicts the distribution of the direct (stratified) estimates, with other methods shown in the legend. The left column assumes constant control risk over time, whereas the right column assumes a decreasing trend in the control arm risk over time.

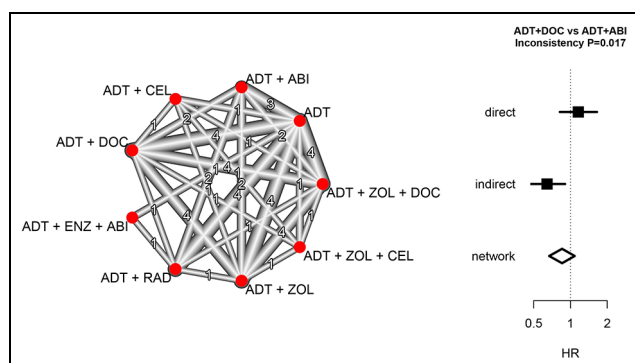
adds very little, so that the network and adjusted methods are almost identical to the direct method. However, for comparison of treatments A and B, the indirect information is more substantial, leading to greater efficiency for the network and adjusted methods compared to the direct method. A complete tabulation of all simulation results is provided in the Supplemental Material where, as well as the above observations, we observe that the adjusted regression and the network approaches are essentially equivalent across all comparisons and scenarios.

Overall, the primary conclusion of the simulation study is that time trends in the background risk level

may lead to bias in unadjusted analyses, while stratification and adjustment may be used to address this bias.

### STAMPEDE trial

The STAMPEDE trial is a multi-stage platform trial of various treatments for metastatic prostate cancer with a primary endpoint of overall survival.<sup>3,7,15,16</sup> The trial was initiated in 2005 with a control arm of androgen-deprivation therapy (ADT) which continued until 2015, after which ADT was dropped as the control therapy. During the period in which ADT was the control, an additional eight experimental therapies were initiated



**Figure 4.** Network analysis of the STAMPEDE multi-stage platform trial of prostate cancer therapies during the years 2005–2015. The network diagram depicts the network and number of direct treatment comparisons between combinations of androgen-deprivation therapy (ADT), docetaxel (DOC), zoledronic acid (ZOL), celecoxib (CEL), abiraterone (ABI), radiotherapy (RAD), and enzalutamide (ENZ). For the comparison of ADT + DOC and ADT + ABI, the direct, indirect, and network estimates of the hazard ratio (HR) for overall survival are shown with 95% confidence intervals.

and dropped at various times. These adaptation times define seven stages within which the design remained fixed. The precise periods between adaptations have been published in various sources, see for example the study's protocol<sup>17</sup> (Table 6, page 10) in which the seven stages correspond with the first seven protocol changes after activation. The complete network of available randomized treatment comparisons for the era in which ADT was the control therapy is shown in Figure 4. This network diagram demonstrates the potential complexity of direct and indirect treatment comparisons available in a large platform trial, such as STAMPEDE.

The pre-specified primary analysis for STAMPEDE uses a stratified approach in which treatment comparisons are restricted to patients who were concurrently randomized. The restriction to direct randomized comparisons in STAMPEDE contrasts with, for example, the approach taken in the REMAP-CAP platform trial where both randomized and non-randomized treatment comparisons are combined using an adjusted analysis.<sup>4,6</sup> We now illustrate the potential impact of combining direct randomized evidence with indirect non-randomized evidence using our network approach for one particular comparison, ADT + DOC versus ADT + ABI.

The randomized comparison of ADT and ADT + DOC was based on randomizations during the years 2005–2013, which reached a conclusion that ADT + DOC is superior for overall survival (HR 0.78, CI 0.66–0.93).<sup>7</sup> The randomized comparison of ADT and ADT + ABI was based on randomizations

during the years 2012 and 2013, which reached a conclusion that ADT + ABI is superior (HR 0.63, CI 0.52–0.76).<sup>15</sup> The randomized comparison of the two experimental treatments ADT + DOC and ADT + ABI was based on randomizations during the years 2012 and 2013, which favored ADT + DOC but did not provide evidence of a difference (HR 1.16, CI 0.82–1.65).<sup>16</sup> Nonetheless, the period 2005–2011 provides further information about ADT + DOC that could potentially be used for further exploration of the difference between the two experimental therapies, using ADT as a common reference treatment.

A combined network comparison of ADT + DOC and ADT + ABI may be conducted using randomized comparisons from two periods, 2005–2011 and 2012–2013. For the 2005–2011 stage, there is just one randomized comparison of interest, ADT versus ADT + DOC. For the 2012–2013 stage, there are three randomized comparisons of interest: ADT versus ADT + DOC; ADT versus ADT + ABI; and ADT + DOC versus ADT + ABI. The log HR and associated standard errors for each of these stage-specific randomized comparisons were used in a network analysis implemented in the netmeta package and reproduced in the Supplemental Material.

Figure 4 displays the results of the network analysis. Visually, there is a difference between the direct and indirect treatment comparisons, with an averaging of the two effects in the network analysis, and the test of inconsistency is statistically significant ( $p = 0.017$ ). The analysis also produces an assessment of heterogeneity across the network, which in this case assesses differences between the ADT versus ADT + DOC comparison in the two stages ( $Q = 1.49$ ,  $df = 1$ ,  $p = 0.22$ ).

Figure 4 illustrates the transparent decomposition of the complete network into a direct randomized component and an indirect non-randomized component. The STAMPEDE trial pre-specified the use of randomized comparisons in the primary analysis, and the inconsistency of the two information sources in Figure 4 provides strong support for that approach. The cause of the inconsistency cannot be determined from this analysis but may be due to time trends in the population risk combined with a treatment effect magnitude that is dependent on underlying risk. Regardless of the cause of the inconsistency, this analysis identifies the potential for it to be present and illustrates how it can be transparently reported.

## Conclusion

This article has shown that an adaptive platform trial provides a network of stage-specific treatment comparisons that may be analyzed using the methodology of



network meta-analysis. We have described the network approach at a conceptual level, emphasizing the combination of randomized and non-randomized information using weighted linear combinations of stage-specific direct treatment comparisons. Our approach was guided by the methods and software available in the *netmeta* package in R.<sup>11,13</sup> However, alternative network methods and software exist and could be used in an analogous manner.<sup>18</sup> Rather than advocating specific methods and software, our goal has been to present the general principles of network meta-analysis as a relevant methodology for analyzing platform trials.

An important characteristic that distinguishes different platform trials is the choice of a Bayesian or frequentist approach. Both approaches are common, and the Bayesian REMAP-CAP and frequentist STAMPEDE trials exemplify the two. However, the distinction between Bayesian and frequentist approaches to analyzing platform trials is tangential to the issues discussed in this article. Direct versus indirect evidence, and adjustment versus stratification, have similar relevance to the analysis of platform studies regardless of whether a Bayesian or frequentist framework is used. Software for implementing Bayesian network analyses is also available.<sup>19,20</sup> It is beyond our scope to investigate these methods here; however, it would be of interest for future research to undertake simulation studies of Bayesian network analyses analogous to the simulations presented here.

An issue of growing importance in adaptive trials is treatment effect bias arising from response-driven design adaptations.<sup>21,22</sup> In the context of platform studies, the observed design structure is randomly generated by the outcome of decisions to drop treatments or adaptively change randomization probabilities. Conditional likelihood methods that condition on this observed design have been proposed to adjust for potential estimation biases, and these methods have been studied in the context of traditional group sequential designs,<sup>23</sup> as well as designs with other adaptive features<sup>24</sup> and meta-analyses of group sequential studies.<sup>25,26</sup> It may be possible to extend our network approach for platform studies by using conditional stage-specific treatment effect estimates.

Our work has implications for the design of platform trials. We have assumed that design adaptations occur at discrete times that define the stages of a multi-stage platform design. This means that within a stage the design remains fixed. Frequent adaptation limits the information provided by each fixed design and forgoes the opportunity to assess the consistency between randomized and non-randomized evidence. Instead, we recommend an adaptation frequency guided by the typical timing of interim analyses in a group sequential trial. Furthermore, in assessing the consistency between

direct and indirect evidence, formal tests of inconsistency are available and should be conducted. However, like tests of interaction, such tests may have low power in some circumstances. Since designing a platform trial typically involves extensive simulations to inform design assumptions, it would be possible to assess the power of inconsistency tests as part of these design simulations.

We also recommend that the reporting of platform trials should take advantage of the underlying fixed designs by providing the stage-specific treatment effects, as illustrated in Table 1. These could be valuable for subsequent systematic reviews and re-analyses. For example, the STAMPEDE trial has contributed to subsequent reviews and re-analyses that may have benefited from stage-specific information.<sup>27,28</sup> Nonetheless, in the absence of stage-specific treatment effects, it may still be possible to use innovative ways to extract the information, such as reconstruction of individual patient data from published Kaplan–Meier plots.<sup>29</sup> The Supplemental Material provides an illustration of this approach.

Although the purpose of this article has not been to advocate for one or other of the stratified or adjusted approaches, it is natural to consider which approach is more appropriate. The direct randomized treatment comparisons from a platform trial provide a higher quality of evidence than the indirect non-randomized treatment comparisons, which provides an argument for using the stratified approach in the primary analysis. As discussed in the Background section, the STAMPEDE and REMAP-CAP trials are exemplars of the stratified and adjusted approaches, respectively. Given the inconsistency between randomized and non-randomized comparisons in the STAMPEDE case study, we would recommend the STAMPEDE approach of restricting to concurrent randomizations as the most appropriate primary analysis method, with the REMAP-CAP approach of including non-concurrent randomizations reserved for secondary analyses. Indeed, it could be argued that non-concurrent controls are unnecessary if the trial is adequately powered, and we would recommend every effort be made to power adaptive platform trials based on concurrently randomized treatment comparisons. Nonetheless, we acknowledge the existence of differing viewpoints on the use of non-concurrent controls and advocate further research on their advantages and disadvantages.

In summary, regardless of whether a platform trial plans to use only direct randomized evidence in the primary analysis or plans to use all available information including indirect non-randomized evidence, this article has shown that transparent display of direct and indirect comparisons using a network analysis provides a valuable supplement to the planned primary analysis.

## Acknowledgements

The authors acknowledge the support of the Australian Trials Methodology Research Network (AusTriM).


## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by research grants 1150467 and 1171422 from the National Health and Medical Research Council of Australia.

## ORCID iD

Ian C Marschner  <https://orcid.org/0000-0002-6225-1572>

## Supplemental material

Supplemental material for this article is available online.

## References

- Pallmann PP, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run them. *BMC Medicine* 2018; 16: 29.
- Angus DC, Alexander BM, Berry S, et al. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nature Reviews Drug Discovery* 2019; 18: 797–807.
- Sydes MR, Parmar MKB, James ND, et al. Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer: the MRC STAMPEDE trial. *Trials* 2009; 10: 39.
- Dodd LE, Freidlin B and Korn EL. Platform trials—beware the non-comparable control group. *New Engl J Med* 2021; 384: 1572–1573.
- Proschan M and Evans S. Resist the temptation of response-adaptive randomization. *Clin Infect Dis* 2020; 71: 3002–3004.
- REMAP-CAP Investigators. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *New Engl J Med* 2021; 384: 1491–1502.
- James ND, de Bono JS, Spears MR, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multi-arm, multi-stage, platform randomized controlled trial. *Lancet* 2016; 387: 1163–1177.
- Higgins JPT, Thomas J, Chandler J, et al. *Cochrane handbook for systematic reviews of interventions* (Chapter 11). 2nd ed. Chichester: John Wiley & Sons, 2019.
- Lu G and Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006; 101: 474–459.
- Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683–691.
- Schwarzer G, Carpenter JR and Rücker G. *Meta-analysis with R* (Chapter 8). Geneva: Springer, 2015.
- König J, Krahn U and Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Stat Med* 2013; 32: 5414–5429.
- Schwarzer G. netmeta: network meta-analysis using frequentist methods. R Package Version 1.5-0, 2021, <https://CRAN.R-project.org/package=netmeta>
- Marschner IC. Relative risk regression for binary outcomes: methods and recommendations. *Aust New Zeal J Stat* 2015; 57: 437–462.
- James ND, de Bono JS, Spears NW, et al. Abiraterone for prostate cancer not previously treated with hormone therapy. *New Engl J Med* 2017; 377: 338–351.
- Sydes MR, Spears MR, Mason MD, et al. Adding abiraterone or docetaxel to long-term hormone therapy for prostate cancer: directly randomized data from the STAMPEDE multi-arm, multi-stage platform protocol. *Ann Oncol* 2018; 29: 1235–1248.
- James ND, Brown L, Parker C, et al. Appendices to the STAMPEDE protocol. Version 19.0, 2020, <http://www.stampetrials.org/centres/protocol-amendment-pack/>
- White IR. Network meta-analysis. *Stat J* 2015; 15: 951–985.
- Béliveau A, Boyne DJ, Slater J, et al. BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network meta-analysis. *BMC Med Res Methodol* 2019; 19: 196.
- Seo M and Schmidt C. bnma: Bayesian network meta-analysis using JAGS. R Package Version 1.4.0, 2021, <https://cran.r-project.org/web/packages/bnma/bnma.pdf>
- Marschner IC. A general framework for the analysis of adaptive experiments. *Stat Sci* 2021; 36: 465–492.
- Robertson DS, Choodari-Oskooei B, Dimairo M, et al. Point estimation for adaptive trial designs. *arXiv*, 2021, <https://arxiv.org/abs/2105.08836>
- Marschner IC and Schou IM. Underestimation of treatment effects in sequentially monitored clinical trials that did not stop early for benefit. *Stat Method Med Res* 2019; 28: 3027–3041.
- Kimani PK, Todd S and Stallard N. Conditionally unbiased estimation in phase II/III clinical trials with early stopping for futility. *Stat Med* 2013; 32: 2893–2901.
- Schou IM and Marschner IC. Meta-analysis of clinical trials with early stopping: an investigation of potential bias. *Stat Med* 2013; 32: 4859–4874.
- Marschner IC, Askie LM and Schou IM. Sensitivity analyses assessing the impact of early stopping on systematic reviews: recommendations for interpreting guidelines. *Res Synth Methods* 2020; 11(2): 287–300.
- Tan PS, Aguiar P, Haaland B, et al. Addition of abiraterone, docetaxel, bisphosphonate, celecoxib or combinations to androgen-deprivation therapy (ADT) for metastatic hormone-sensitive prostate cancer (mHSPC):

- a network meta-analysis. *Prostat Cancer Prostat Dis* 2018; 21: 516–523.
28. Vale CL, Fisher DJ, White IR, et al. What is the optimal systemic treatment of men with metastatic, hormone-naïve prostate cancer? A STOPCAP systematic review and network meta-analysis. *Ann Oncol* 2018; 29: 1249–1257.
29. Liu N, Zhou Y and Lee JJ. IPDfromKM: reconstruct individual patient data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 2021; 21: 111.