Health Equity Impact Evaluation of New Treatments – Evidence Synthesis Methods to Overcome Data Gaps

Jeroen P Jansen (jeroen.jansen@ucsf.edu)
University of California, San Francisco

Research Article

Keywords: distributional cost-effectiveness analysis, health equity impact, evidence gaps, evidence synthesis methods

Posted Date: October 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-951925/v1

HEALTH EQUITY IMPACT EVALUATION OF NEW TREATMENTS – EVIDENCE SYNTHESIS METHODS TO OVERCOME DATA GAPS

Jeroen P Jansen PhD

Department of Clinical Pharmacy, School of Pharmacy
University of California – San Francisco
jeroen.jansen@ucsf.edu

HEALTH EQUITY IMPACT EVALUATION OF NEW TREATMENTS – EVIDENCE SYNTHESIS METHODS TO OVERCOME DATA GAPS

keywords: distributional cost-effectiveness analysis, health equity impact, evidence gaps, evidence synthesis methods.

ABSTRACT

Background: Distributional cost-effectiveness analysis (DCEA) has been introduced as an extension of conventional cost-effectiveness analysis to quantify health equity impacts. Although health disparities are recognized as an important concern, the typical analyses conducted to inform health technology assessment of a new intervention do not include a formal health equity impact evaluation or DCEA. One of the reasons is that the clinical trials for new interventions frequently do not have the power or are not designed to estimate the required treatment effects for sub-populations across which you want to analyze equity. The objective of the paper is to discuss how gaps in evidence regarding equity-relevant subgroup effects for new and existing interventions can potentially be overcome with advanced Bayesian evidence synthesis methods to facilitate a credible model-based DCEA.

Methods: First, the evidence needs and challenges for a model-based DCEA are outlined. Next, alternative evidence synthesis methods will be summarized, followed by an illustrative example of implementing these methods. The paper will conclude with some practical recommendations.

Results: The key evidence challenges for a DCEA relate to estimating relative treatment effects due to lack of inclusion of relevant subgroups in the randomized controlled trials (RCTs), lack of access to individual patient data (IPD) for all trials, small subgroups resulting in uncertain effects, and reporting gaps. Advanced Bayesian evidence synthesis methods can help overcome evidence gaps by considering all relevant direct, indirect, and external evidence simultaneously. Methods discussed include (network) meta-analysis with shrinkage estimation, conventional (network) meta-regression analysis, multi-level (network) meta-regression analysis, and generalized evidence synthesis. For a new intervention for which only RCT evidence is available and no real-world data, estimates can be improved if the assumption of exchangeable subgroup effects or the shared or exchangeable effect-modifier assumption among competing interventions can be defended. Furthermore, formal expert elicitation is worthwhile to improve estimates.

Conclusion: This paper provides an overview of advanced evidence synthesis methods that may help overcome typical gaps in the evidence base to perform model-based DCEA along with some practical recommendations. Future simulation studies are needed to assess the pros and cons of different methods for different data gap scenarios.

INTRODUCTION

Health technology assessment — the systematic assessment of the value of a health technology — can help decision-makers with creating policies to ensure appropriate and efficient use to achieve optimal health outcomes. Health disparities are an important concern. However, the typical analyses conducted to inform value assessment do not include a formal evaluation of the impact a new intervention will have on health equity, despite the availability of a quantitative framework to do so [1-4]. With a distributional cost-effectiveness analysis (DCEA), the impact of the new intervention and standard of care are estimated for the different subgroups of the target patient population across which you want to analyze health equity. For the remainder of this paper we use the term equity relevant subgroups to refer to these subgroups, which can be defined according to sex, gender, race and ethnicity, education, socioeconomic status, or geographic location (urban or remote). The distributions of health outcomes with the new intervention and standard of care are compared in terms of total health (similar to a conventional cost-effectiveness analysis) as well as health inequality, taking into consideration the health opportunity costs.

To date, however, full distributional DCEAs have been performed for a few public health interventions, and not for new drugs [5-10]. The question is whether a DCEA of a new intervention can provide meaningful results given the relative sparseness of the available evidence at the time of market introduction. Clinical trials, which are primarily designed for regulatory approval, frequently do not have the power or are not designed to estimate the required treatment effects by equity relevant patient characteristics that may act as effect-modifiers. For existing competing interventions more information is available, including real-world evidence that may be useful to supplement the trial evidence to obtain required subgroup estimates of treatment effects.

An additional evidence challenge for DCEA, which applies to studies of both new and existing interventions, is that there is frequently no access to the individual level patient data (IPD) of the randomized controlled trials (RCTs); there is only access to aggregate-level information from study reports or publications. This means that even if patient characteristics related to social and biological constructs have been recorded for individual patients in the data set, reported study

results may not have been stratified by these characteristics. This can be considered an evidence reporting gap.

Health economic models, where multiple sources of evidence regarding benefits, risks, and costs for the compared interventions are integrated and extrapolated, are commonly used for conventional cost-effectiveness analysis and are the methodology of choice for DCEAs as well. Supplementing direct evidence with indirect and external evidence to overcome data gaps or data reporting limitations is a key characteristic of model-based health economic evaluations. In the last decade, evidence synthesis techniques have been developed to combine direct, indirect, and external evidence, as well as to combine RCT evidence with observational evidence [11-24]. These are powerful techniques to estimate health economic model parameters. Although these methods may help overcome gaps regarding equity-relevant subgroup effects, they may not be very familiar in the health economics and value assessment community.

The objective of this paper is to discuss how gaps in evidence regarding equity relevant subgroup effects for new and existing interventions can potentially be overcome with advanced Bayesian evidence synthesis methods to facilitate a credible DCEA. The paper will first outline the evidence needs and challenges for such a model-based analysis. Next, the alternative evidence synthesis methods will be summarized, followed by an illustrative example. The paper will conclude with some practical recommendations.

EVIDENCE NEEDS AND CHALLENGES FOR MODEL-BASED DISTRIBUTIONAL COST-EFFECTIVENESS ANALYSIS

DCEA aims to provide decision-makers with information to make trade-offs between improving total health and reducing health inequality [1-4]. For each competing intervention for a given condition, the expected health outcomes (e.g. QALYs), costs, and the net health benefit as a measure of cost-effectiveness factoring in opportunity costs (NHB = QALYs – costs/willingness-to-pay) is estimated as well as the distribution of net health benefits across the different equity relevant subgroups (Figure 1). With information on how adverse society (or a decision-maker) is to health inequality (and expressed with, for example, the Atkinson inequality aversion

parameter), we can quantify the health equity impact (expressed with an index measure) from an increase (decrease) in inequality in the distribution of net health benefit across the equity relevant subgroups due to implementation of the interventions. The equity impact metric can be combined with the total net health benefit in an overall equity-weighted measure, e.g. equally distributed equivalent QALYs, that combines concern for both equity and cost-effectiveness [2,4].

[Insert Figure 1]

For a DCEA we need credible estimates regarding the expected outcomes and costs of interest for the equity relevant subgroups for the alternative interventions being compared. Unfortunately, there is typically no empirical study available that provides all this information. Lacking such a study, we need to integrate multiple sources of evidence to estimate the expected health and economic outcomes of interest for the equity relevant subgroups, with each source providing a piece of the required information. Resulting health economic models combine relevant evidence on the natural course of disease or outcomes with a reference treatment, relative treatment effects for alternative interventions, resource use, costs, and utility estimates for the different disease states. In Figure 2 a simple influence diagram of a fictitious, but representative, health economic model for a DCEA of cancer treatment is presented. It depicts the elements of the model (boxes) and assumptions which elements influence each other directly (arrows). With such a figure in mind, the equity-relevant subgroup-specific evidence needed for a DCEA can be identified more clearly.

[Insert Figure 2]

For a conventional model-based cost-effectiveness analysis, standard practice is to estimate outcomes associated with the natural course of the disease or reference treatment, i.e. the *baseline arm* of the model, as well as *relative treatment effects* of the alternative interventions of interest versus no treatment or the reference intervention. The relative treatment effects are applied to the baseline arm to obtain estimates of the expected outcomes for each of the

alternative interventions. For a DCEA we need these estimates to be representative of the equity relevant subgroups. An intervention will have an impact (positive or negative) on inequality in health outcomes when its relative treatment effects vary between the equity relevant subgroups (i.e. heterogeneous treatment effects) or when there are differences in baseline event rates across the equity relevant subgroups that are affected by treatment. To clarify: absolute differences in the event rates between the subgroups will be affected if a relative treatment effect expressed as a ratio measure (e.g. odds ratio, relative risk, or hazard ratio) is applied to the baseline risk, even if this relative treatment effect is homogenous across these subgroups. Heterogenous baseline and relative treatment effects are represented in Figure 2 with the green-colored arrows. Inequality in health outcomes is further impacted if there is unequal access or uptake across the equity relevant subgroups of the target patient population. Explicitly representing the distribution of uptake of the compared interventions in the economic model allows evaluating the impact of uptake relative to the impact of heterogenous baseline and treatment effects on inequality in health outcomes. For this paper, however, we focus on estimating the equity-relevant baseline and relative treatment effects required for a model-based DCEA.

To perform a DCEA, the expected clinical outcomes by intervention need to be translated into distributions of expected QALYs and costs. The expected costs include treatment acquisition and administration costs that are directly determined by the intervention of choice provided, as well as other costs that depend on the health status or outcomes. The simplest structural model assumption, which is frequently used for a conventional CEA as well, is that the relationship of utilities and other costs are only a function of health status or clinical outcomes. As a result, differences in baseline and relative treatment effects across equity relevant subgroups will result in different expected QALY and cost estimates across the subgroups. In other words, to facilitate the DCEA we only need equity-specific subgroup evidence for the baseline and relative treatment effects and not for its functional parameters: utilities and costs. However, it is required to relax this assumption when we have evidence that utility and other costs as a function of health status or outcomes vary by the equity-relevant subgroups. (This would be depicted with an arrow from the patient characteristics box to the disease management-related cost box and QALY box in Figure 2.)

Whether we can estimate the baseline and relative treatment effects to parameterize the health economic model depends on the equity relevant subgroups of interest in relation to the evidence available. As mentioned, relevant population characteristics that relate to health equity concerns to consider when defining the equity relevant subgroups of interest include sex and gender, race and ethnicity, education, socioeconomic status, and geographic location (urban or remote), among others. How we define the subgroups of interest cannot be completely done independently of the data available to estimate the baseline and relative treatment effects. If we have access to IPD, we can use population adjustment methods such as propensity score approaches or multivariable regression analysis to estimate baseline and relative treatment effects for more nuanced target subgroups defined by multiple factors [25]. If we only have access to aggregate-level data extracted from publications or reports, we may only be able to obtain appropriate estimates for subgroups defined by one or two factors. A covariate is more likely to be a prognostic factor of an outcome than a relative treatment effect modifier because the impact of a covariate on outcomes cancels out for (RCT-based) relative treatment effects. As such, when defining the target equity relevant subgroups of interest, it is arguably more important to consider the level of detail in the data available to estimate the baseline arm of the model than the available data used for estimating relative treatment effects. It is important to recognize that certain factors may be (imperfect) surrogates of other underlying factors responsible for disparities in health outcomes. Variables such as socioeconomic status, geographic location, insurance status, access to quality healthcare, health behaviors, or genetic ancestry are more informative than race to define subgroups of interest. Although race is increasingly recognized as a poor surrogate of social and biological constructs, it may be all the information available in the available evidence, especially in RCTs [26]. The US Food and Drug Administration (FDA) released guidance to ensure race and ethnicity subgroup data is collected consistently [27,28].

Preferred sources of evidence for relative treatment effects of the compared interventions are RCTs. Since relative effects remain relatively stable from one study population to next (in contrast to absolute effects of a treatment), the low internal bias (i.e. absence of confounding or selection bias) in RCTs relative to observational studies outweighs any concerns about the external bias (applicability of estimates to different populations) of relative treatment effect estimates. The available RCTs for biopharmaceutical interventions have typically been designed

to detect a relative treatment effect for the overall study population of interest to support regulatory approval. This will pose challenges for a DCEA when subgroup effects have not been reported or subgroup data are not available for levels of effect-modifiers relevant to the equity relevant subgroups of interest. Even if the available RCTs do provide information on relative treatment effects relevant for the subgroups of interest, the studies may not have been powered to detect these subgroup effects and the relative treatment effect estimates may be characterized by substantial uncertainty due to small sample sizes. Another potential evidence gap is that the RCT study population excludes certain equity relevant subgroups of interest. Finally, a challenge in estimating subgroup effects is if there is no access to IPD for the competing interventions. Although we likely do not need to obtain relative treatment effect estimates stratified to the same level of detail as the equity relevant subgroup of interest because not all variables that define the equity relevant subgroups are modifiers of relative treatment effects, there are still challenges when aiming to estimate relative treatment effects for key equity relevant effect-modifiers.

The sources of evidence for the baseline effects should be as specific as possible to the target equity relevant subgroups because, as mentioned, absolute effects are likely to vary with the study population. More specifically, the available evidence needs to match the subgroups of interest regarding prognostic factors if the baseline arm of the model reflects the natural course of disease in absence of treatment, and the available evidence needs to match the subgroups of interest regarding both prognostic factors and effect-modifiers if the baseline arm represents the reference treatment. For the above-mentioned reasons, RCTs may not constitute the best evidence for baseline effects. Preferred evidence for the baseline arguably comes from real-world cohort studies. Again, evidence gaps and challenges relate to a mismatch between the available and required evidence given the equity relevant subgroups of interest, and whether there is access to IPD.

Evidence gaps and challenges to estimate baseline and relative treatment effects for the equity relevant subgroups can be characterized as follows, as depicted in Figure 3: 1) no evidence for some or all of the subgroups of interest due to exclusion of representative individuals from the studies; 2) lack of access to IPD, and aggregate-level information is not stratified by the subgroups of interest (e.g. results are provided for the combination of subgroup A and B, but not for A and B separately); 3) subgroup effects are uncertain due to small sample sizes; and 4) a

combination of any these factors. All this with the notion that relative treatment effects need to be estimated for the different levels of the equity relevant effect modifiers, and the baseline effects need to be estimated for the equity relevant prognostic factors (and effect modifiers if the baseline arm of the model reflects the reference treatment).

[Insert Figure 3]

EVIDENCE SYNTHESIS METHODS TO ESTIMATE TREATMENT EFFECTS ACROSS EQUITY RELEVANT SUBGROUPS

The fundamental premise of evidence synthesis is that each empirical study is a piece of a larger evidence base and its findings are interpreted as such. Each study evaluates a subset of all information of interest and by considering the findings of all relevant studies simultaneously we have a lot more information to estimate the parameters of interest. For example, under the assumption of consistency, study 2 and 3 in Figure 3 in combination informs the estimates for subgroup A. When we add study 4 in the synthesis we also get information for subgroups. Adding study 5 results in more precision for the estimates for subgroup A, B, and C. In this scenario, study 2 provides direct evidence, and studies 3, 4, and 5 provide relevant indirect evidence.

In this section, evidence synthesis methods will be described that can potentially help estimate treatment effects for new and existing alternative interventions for equity relevant subgroups needed to perform DCEA despite gaps in the evidence base. We take the position that, in principle, we want to estimate relative treatment effects for equity relevant subgroups based on RCT evidence and the baseline effects (with the reference treatment) based on real-world evidence. Furthermore, we assume that we face small subgroups in RCTs, limited reporting of subgroup effects, no or limited access to IPD for RCTs, and real-world evidence is only available for established interventions and not for the new health technology.

The focus here is a Bayesian approach to statistical inference. In the Bayesian framework parameters of the statistical evidence synthesis models are viewed as random variables with probability distributions that reflect the belief about their estimates. With a Bayesian approach to statistical inference, we need to define a prior distribution summarizing our belief about the model parameter estimates before considering the data we analyze. This prior distribution is combined with the observed data represented by the likelihood according to using Bayes' theorem to obtain a posterior distribution about the model parameter estimates. In principle, we like the model parameter estimates to reflect the available empirical data and will therefore consider non-informative, or minimally informative prior distributions, wherever possible [13]. However, in the context of limited data, we may want to use informative prior distributions to improve parameter estimation.

Estimation of relative treatment effects

When we want to estimate relative treatment effects based on RCT evidence, it is important to discuss network meta-analysis (NMA). In general, if the relevant evidence base consists of multiple RCTs each comparing a subset of all the competing interventions for a certain condition of interest, and each of these trials has at least one intervention in common with another trial such that the evidence base can be represented with one connected network, an NMA can be performed [13,17,18,29-36]. In the following sections, we will discuss modifications of the standard Bayesian NMA approach that may be relevant to estimate relative treatment effects for equity relevant subgroups. These methods are shrinkage estimation, network meta-regression, IPD-level network meta-regression analysis, and multi-level network meta-regression-based methods [11,19,23,37-41]. However, we will first provide a summary of the standard NMA methodology to provide a foundation for these modified methods. These methods do not only apply to the synthesis of networks of trials, but also to pairwise meta-analysis involving two competing interventions.

Standard network meta-analysis

The purpose of an NMA is to estimate relative treatment effects for a specific target population, which means that the study population in each of the RCTs included in the NMA needs to be representative of the target population of interest. This is the case if there are no systematic differences in patient-related effect modifiers between the trial populations and the target population and implies that there are no systematic differences in patient-related effect modifiers between the different RCTs either [22,32]. As a result, relative treatment effect estimates obtained with the NMA are valid [18,29,32,42]. If, however, the study population in a subset of the RCTs differs from the target population in terms of effect-modifiers then relative treatment effects based on indirect evidence will be biased. If there are no between-trial differences in study populations but there are differences in effect modifiers relative to the target population then the relative treatment effects obtained with the NMA are valid, but not representative of the target population of interest; we have external bias [22,38]. Biased relative treatment effect estimates are not only caused by imbalances in patient-related effect-modifiers but also occur if there are differences in effect-modifiers related to study design or contextual factors between the studies. For a credible and relevant NMA we need a connected network of RCTs where each trial has at least one intervention in common with another trial, without systematic differences in known and unknown effect modifiers between studies, and no differences in patient and context related effect-modifiers relative to the target population and setting of interest. In principle, a standard NMA can be performed by equity relevant subgroups, evidence permitting.

The general random-effects NMA model can be described as follows:

$$g(\gamma_{ik}) = \theta_{ik} = \begin{cases} \mu_i & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} & k \succ b \end{cases}$$
$$\delta_{i,bk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2) \qquad (1)$$

where g is an appropriate link function (e.g. the logit link for binary outcomes) and θ_{ik} is the linear predictor of the expected outcome with intervention k in trial i (e.g. the log odds). μ_i is the study i specific outcome with comparator intervention b. $\delta_{i,bk}$ reflects the study-specific relative treatment effects with intervention k relative to comparator b and are drawn from a normal distribution with the pooled relative treatment effect estimates expressed relative to the overall reference intervention A: $d_{bk} = d_{Ak} - d_{Ab}$ (with $d_{AA} = 0$) . Estimates of d_{Ak} reflect the relative

treatment effect of each intervention *k* relative to overall reference intervention A based on direct and/or indirect evidence. Variance parameter σ^2 reflects the heterogeneity across studies. With a fixed effect NMA, $\delta_{ibk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2)$ is replaced with $\delta_{ibk} = d_{Ak} - d_{Ab}$ because σ^2 is assumed to be 0. The model applies to many types of data, by just specifying an appropriate likelihood describing the data generating process and corresponding link function. (Dias et al. 2018b). With the NMA performed in a Bayesian framework, we need to define prior distributions for the parameters to be estimated, μ_i, d_{Ak} , and σ^2. For example, $\mu_i \sim Normal(0, 100^2)$, $d_{Ak} \sim Normal(0, 100^2)$, and $\mu \sim uniform(0, x)$ with x a reasonable upper bound dependent on the expected range of observed relative treatment effects.

Network meta-analysis with shrinkage estimation

If we have an evidence network of RCTs for which results are reported relevant for the equity relevant subgroups but are uncertain due to a limited number of studies and small sample sizes in each of the subgroups, borrowing strength from other interventions or subgroups by deriving a shrinkage estimate may be useful [16,23,43,44,45]. This approach can be implemented by grouping the multiple interventions in the network into a smaller set of classes with the underlying assumption that the intervention-specific relative effects within a class of interventions are exchangeable. Interventions assigned to the same class, for example, based on mechanism of action, are deemed more alike regarding relative treatment effects for a specific subgroup than interventions from different classes [38]. The model expressed with equation 1 can be modified accordingly by defining that the relative treatment effect parameters d_{Ak} are assumed to come from a distribution with a common mean and variance, if they belong to the same class:

$$d_{Ak} \sim Normal\left(m_{D_k}, \sigma^2_{D_k}\right) \tag{2}$$

where D_k is defined as the class to which intervention *k* belongs. m_{D_k} is the mean class effect in class D_k, and $\sigma^2_{D_k}$ are the within-class variances. The key benefit of the exchangeability assumption is that unstable estimates for d_{Ak} of interventions within a class due to limited subgroup data will be shrunken towards the class mean effect and become more precise than obtained with model 1 where d_{Ak} are assumed to be independent. Informative distributions and

sensitivity analysis may be needed for $\sigma^2_{D_k}$ if the number of interventions per class is limited [38]. With this approach we perform a separate NMA for each of the equity relevant subgroups and highly uncertain relative treatment effects are stabilized by borrowing information from the data from other interventions for the same subgroup.

Another approach to implementing shrinkage estimation for NMA is by assuming that the subgroup-specific relative treatment effects are exchangeable within interventions. All the mutually exclusive subgroups are incorporated in the NMA of the competing interventions simultaneously according to:

$$g(\gamma_{is,k}) = \theta_{is,k} = \begin{cases} \mu_{is} & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_{is} + \delta_{is,bk} & k \succ b \end{cases}$$
$$\delta_{is,bk} \sim Normal\left(d_{s,Ak} - d_{s,Ab}, \sigma_s^2\right)$$
$$d_{s,Ak} \sim Normal(D_{Ak}, \sigma_{Ak}^2) \qquad (3)$$

where $\theta_{is,k}$ is the linear predictor for the expected outcome with intervention *k* in subgroup *s* of trial *i*. μ_{is} is the expected outcome with comparator intervention *b* in subgroup *s* of study *i*. $\delta_{is,bk}$ reflects the relative treatment effect with intervention *k* relative to comparator *b* in subgroup *s* of trial *i* and are drawn from a normal distribution with the pooled estimates expressed in terms of the overall relative treatment effects versus intervention A in that subgroup $d_{s,Ak}$. With this model we make the additional assumption that the subgroup-specific relative treatment effects $d_{s,Ak}$ are drawn from a common normal distribution with mean D_{Ak} and intervention-specific variance σ_k^2. As a result, highly uncertain relative treatment effects for each subgroup are stabilized by borrowing information from the data from other subgroups for that intervention [16].

With these shrinkage models we improve estimation for both new and existing interventions by assuming exchangeability between interventions or between subgroups. The first approach may be difficult to defend if a new intervention has a very different mechanism of action and efficacy than its competing interventions. The second approach does not rely on this assumption. The assumption of exchangeable subgroup-specific relative treatment effects for a given intervention

is in line with a long tradition in meta-analysis and epidemiology that relative treatment effects are relatively stable across subgroups.

Network meta-regression

When there are observed differences between the target equity relevant subgroups of interest and the study populations of the individual RCTs regarding effect-modifiers, a meta-regression can potentially be used to adjust for this external bias and provide relevant relative treatment effect estimates [22,37,38]. When the available evidence base only consists of aggregate-level data, the model presented in equation 1 can be extended with a covariate according to [37,38,46-49]:

$$\theta_{ik} = \begin{cases} \mu_i & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + (\beta_{Ak} - \beta_{Ab})(m_i - x_{target}) & k \succ b \end{cases}$$
$$\delta_{i,bk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2) \qquad (4)$$

m_i is the study-level covariate value of the effect-modifier of interest for trial i. β_{Ak} represent the covariate effects with intervention *k* relative to the overall reference intervention A. x_{target} is the centered covariate value representing the target subgroup of interest. d_{Ak} represent the relative effect of the intervention *k* compared to intervention A for the target subgroup of interest. With this model we do not only assume consistency regarding relative treatment effects, but also regarding the parameters reflecting the impact of the covariates ($\beta_{bk} = \beta_{Ak} - \beta_{Ab}$ and $d_{AA} = \beta_{Ak} = 0$). In model 4 the impact of the covariate on the relative treatment effects is assumed to be independent for each intervention *k* relative to A. However, we can also simplify the model by assuming the impact of the covariate is the same for every intervention *k* relative to A, $\beta_{Ak} = B$, or assume these to be exchangeable, $\beta_{Ak} \sim Normal(B, \sigma_B^2)$ [37,38,46]. This shared or exchangeable effect-modifier assumption is useful when the number of studies is limited.

It is important to emphasize the limitation of meta-regression analysis involving patient characteristics based on trial level information extracted from summary reports or publications [37,38]. If the study population of a particular trial is homogeneous regarding a dichotomous patient characteristic but may differ between studies (e.g. only patients of Western decent or only Japanese patients are included in a study) then we have a dichotomous study-level covariate. If the trial population is heterogeneous regarding a dichotomous characteristic (e.g. a mixed study

population of males and females) the study-level covariate is continuous representing the proportion of individuals with the characteristic. For a study level summary measure of a continuously distributed patient characteristic (e.g. mean age), the study-level covariate is continuous as well. If the precision of a relative treatment effect in each trial is relatively large, the number of studies is small, and the contrast in the study-level covariate between studies sufficiently large, a spurious relationship between the relative treatment effect and the covariate may be statistically significant [38]. On the other hand, with continuously distributed patient characteristics, the within-study variation is typically much larger than the variation in aggregated means used for the meta-regression analysis, thereby not having the power to detect a true relationship [37,38]. Using aggregated information regarding patient characteristics in a network meta-regression is vulnerable to ecologic bias. Due to study-level confounding the estimated relationship between a study-level patient characteristic and the relative treatment effect based on between-study comparisons may be very different than the within-study relationship [39,50-54]. Such ecological bias can occur in non-linear models in the absence of study level confounding [53-57].

With meta-regression analyses, we can improve the estimation of subgroup-specific results of the new intervention with the shared or exchangeable effect-modifier assumption and, the trials of the new intervention have the overall reference intervention (i.e. intervention A) included as a control group.

Individual participant data (IPD)-level network meta-regression analysis

The limitations of estimating relative treatment effects for equity relevant subgroups of interest through network meta-regression based on aggregate-level data can be overcome with the use of IPD [38,39,58-60]. If IPD is available for all the RCTs in the evidence network, one evidence synthesis model we can use is the following:

$$\theta_{ijk} = \begin{cases} \mu_i + \beta_{0i} x_{ij} & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + \beta_{0i} x_{ij} + (\beta_{Ak} - \beta_{Ab}) x_{ij} & k \succ b \end{cases}$$

$$\delta_{i,bk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2) \qquad (5)$$

j reflects the individual in study *i*. β_{0i} is the main effect of covariate *x* on the outcome of interest in study *i*. x_{ij} is the value of the covariate for individual *j* in study *i*. Here we assume the interaction effect β_{Ak} is fixed across studies. We can also separate the within and between-study interaction between intervention and covariate and use a model with a covariate for the study level mean value of the patient characteristic and a covariate for the individual patient value of this effect-modifier minus the mean value in that study to describe the within-study variation [54,61]:

$$\theta_{ijk} = \begin{cases} \mu_i + \beta_{0i} x_{ij} & k = b, b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + \beta_{0i} x_{ij} + (\beta^a_{Ak} - \beta^a_{Ab}) m_i + (\beta^w_{Ak} - \beta^w_{Ab})(x_{ij} - m_i) & k \succ b \end{cases}$$

$$\delta_{i,bk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2)$$

(6)

β^a_{Ak} represent the between-study coefficient for the covariate effects with intervention *k* relative to the overall reference intervention A. β^w_{Ak} represent the within-study coefficient for the covariate effects with intervention *k* relative to the overall reference intervention A. If the within-study and between-study interactions are different then ecological bias may be present and inferences regarding relative treatment effects for specific target subgroups should be based on the within-study interactions [38]. Again, models (5) and (6) can be modified by assuming that the impact of the effect modifier is the same for every intervention *k* relative to A.

Network meta-regression with participant-level data and aggregate-level data

In reality, there is hardly ever access to IPD for all trials in an NMA. At best we have IPD only for a subset of studies. We can perform the evidence synthesis based on a combination of IPD studies and aggregate-level studies with the following model [61-64]:

IPD studies:

$$\theta_{ijk} = \begin{cases} \mu_i + \beta_{0i} x_{ij} & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + \beta_{0i} x_{ij} + (\beta^a_{Ak} - \beta^a_{Ab}) m_i + (\beta^w_{Ak} - \beta^w_{Ab})(x_{ij} - m_i) & k \succ b \end{cases}$$

Aggregate-level data studies:

$$\theta_{ik} = \begin{cases} \mu_i & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + (\beta^a_{Ak} - \beta^a_{Ab}) m_i & k \succ b \end{cases}$$

$$\delta_{i,bk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2)$$

(7)

With this model both IPD studies and aggregate-level data studies contribute to the estimation of the treatment-by-covariate interaction effects. If these are believed to be the same for each intervention *k* relative to A then depending on how IPD and aggregate-level data is distributed over the available direct comparisons in the network, we may be able to "transfer" the within-trial interaction estimate for *Ak* comparison for which IPD is available to the *Ak* comparisons for which we only have aggregate-level data. Unfortunately, this only works for specific evidence structures. We can potentially improve the precision of the interaction effects for studies with only aggregate-level data for any network structure based on the available IPD, if we simplify the model with a single treatment-by-covariate interaction parameter for the within- and between-trial comparisons [39,61,65]. However, as mentioned, this will bias the estimates when there is study-level confounding or when we have non-linear models.

It is not uncommon that an analyst has only access to IPD for the trials of the new intervention. Although this facilities subgroup analysis for the new intervention, the question is how much does the aggregate-level information for the competing interventions contribute to estimation of a shared effect-modifier parameter ($\beta_{Ak} = B$), even if we are not worried about ecological bias.

Multilevel network meta-regression with participant-level data and aggregate-level data

A promising new method relevant for the estimation of relative treatment effects for equity relevant subgroups is multilevel network meta-regression (ML-NMR) [39,41]. Unlike the above-mentioned limitation of a network meta-regression model with a shared interaction-effect parameter for the IPD studies and aggregate-level studies required to "transfer" information from

studies with IPD to studies involving other comparisons for which there is only aggregate-level data, ML-NMR avoids such aggregation or ecological bias. A simple ML-NMR model for dichotomous patient-related effect-modifier can be described as follows:

IPD studies:

$$g(\gamma_{ijk}) = \theta_{ijk} = \begin{cases} \mu_i + \beta_0 x_{ij} & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + \beta_0 x_{ij} + (\beta_{Ak} - \beta_{Ab}) x_{ij} & k \succ b \end{cases}$$

Aggregate-level data studies:

$$\gamma_{ik} = \gamma_{ik}^{o}(1 - m_i) + \gamma_{ik}^{1} m_i$$

$$g(\gamma_{ik}^{o}) = \theta_{ik}^{o} = \begin{cases} \mu_i & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} & k \succ b \end{cases}$$

$$g(\gamma_{ik}^{1}) = \theta_{ik}^{1} = \begin{cases} \mu_i + \beta_0 & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i + \delta_{i,bk} + \beta_0 + (\beta_{Ak} - \beta_{Ab}) & k \succ b \end{cases}$$

$$\delta_{i,bk} \sim Normal(d_{Ak} - d_{Ab}, \sigma^2)$$

(8)

The part of the model relevant for IPD studies is the same as used in model 5 with the exception that the coefficient for the prognostic effect of the covariate, β_0, is fixed across studies. For the aggregate-level data part of the model, γ_{ik} is the overall expected outcome in study *i* with intervention *k* and is determined by integrating the individual-level model over the joint within-study distribution of the binary covariate that defines the two subgroups of interest. γ_{ik} equals the sum of the proportion of subjects with covariate *x*=1 in each aggregate-level data study (m_i) multiplied with γ_{ik}^{1} and the proportion of subjects with covariate *x*=0 (1-m_i) multiplied with γ_{ik}^{o}. γ_{ik}^{1}represent the marginal expected outcome with intervention *k* for a subject with the covariate x=1 in study *i*. Similarly, γ_{ik}^{o} is the equivalent for a subject with *x*=0. The key feature of ML-NMR is that an individual model is averaged over the population in study *i* to obtain the aggregate-level model for that study. A generalization of model 8 has been described by Phillippo et al. [41]. ML-NMR addresses several limitations of other proposed evidence synthesis methods when IPD is only available only for a subset of studies, including synthesizing

networks of any size and – important for decision-making – producing estimates in any target population of interest [41].

Although concerns about ecological bias are mitigated with ML-NMR, the question remains whether aggregate-level information from existing interventions can contribute much to the IPD when that is only available for the new intervention.

Network meta-regression with subgroup aggregate-level data integrated over covariate distributions

Using the principles of ML-NMR, we can also imagine a synthesis approach suitable for an evidence base where some studies provide direct evidence for the subgroups of interest and other studies provide indirect evidence via information for the combination of subgroups. (See representative studies 2, 3, 4, and 5 in Figure 3). An appropriate model when we have only aggregate-level data can be described as follows:

Studies providing direct evidence for a subgroup

$$g(\gamma_{ik}^{s}) = \theta_{ik}^{s} = \begin{cases} \mu_i^s & k = b, \quad b \in \{A, B, C, ..\} \\ \mu_i^s + \delta_{i,bk} & k \succ b, \ s = 0 \\ \mu_i^s + \delta_{i,bk} + (\beta_{Ak}^s - \beta_{Ab}^s) & k \succ b, \ s > 0 \end{cases}$$

Studies providing evidence for a combination of subgroups:

$$\gamma_{ik} = \gamma_{ik}^{0}\left(1 - \sum_{s=1}^{S} m_i^s\right) + \sum_{s=1}^{S} \gamma_{ik}^{s} m_i^s$$

$$g(\gamma_{ik}^{s}) = \theta_{ik}^{s} = \begin{cases} \mu_i^0 & k = b, b \in \{A, B, C, ..\}, s = 0 \\ \mu_i^0 + \beta_0^s & k = b, s > 0 \\ \mu_i^0 + \delta_{i,bk} & k \succ b, s = 0 \\ \mu_i^0 + \delta_{i,bk} + \beta_0^s + (\beta_{Ak}^s - \beta_{Ab}^s) & k \succ b, s > 0 \end{cases}$$

(9)

Parameters of the part of the model relevant for studies providing subgroup-specific evidence are defined as follows: γ_{ik}^{s} is the expected outcome in study *i* with intervention *k* for subgroup *s*. μ_i^0 is the study *i* specific outcome with comparator intervention *b* in subgroup *s=0*. β_0^s is the

fixed difference in outcomes with intervention *b* in subgroup *s* relative to the reference subgroup *s*=0. $\delta_{i,bk}$ reflects the study-specific relative treatment effects with intervention *k* relative to comparator *b* for the overall reference subgroup *s*=0. d_{Ak} reflects the pooled relative treatment effect of each intervention *k* relative to overall intervention of reference A for the overall reference subgroup *s=0*. β_{Ak}^{s} represent the difference in the relative treatment effect with *k* versus A in subgroup *s* relative to the reference subgroup *s*=0. For the part of the model describing evidence for a combination of subgroups, the additional parameters are defined as follows: γ_{ik} is the overall expected outcome in study *i* with intervention *k* and is determined by integrating the subgroup-level model over the joint within-study distribution of the categorical covariate that defines the multiple subgroups of interest. γ_{ik} equals the sum of the proportions of subjects in each of the subgroups *s* in study *i,* m_i^s, multiplied with the corresponding expected outcome with intervention *k*, γ_{ik}^{s}.

With this method, we can improve the estimation of equity relevant subgroup effects for the new intervention if we assume the impact of effect-modifiers that define the subgroups of interest is the same for all interventions compared, and the trial(s) of the new intervention have the overall reference intervention (i.e. intervention A) included as a control group.

Generalized evidence synthesis

The evidence synthesis models discussed in the previous sections aim to estimate relative treatment effects for the equity relevant subgroups of interest based on RCT evidence. With specific structural assumptions, the information from existing interventions can help improve subgroup-specific estimates for the new intervention. Depending on the challenges with the evidence base, different evidence synthesis approaches may be more or less relevant, or we may need bespoke models that combine elements of these different models. When the RCT evidence is too limited to obtain relevant and stable estimates for the subgroups of interest, we may want to consider relevant real-world data for the alternative interventions to supplement the RCT evidence [66]. Real-world data sources have likely more information about the effect of the intervention in heterogeneous populations than what is available in RCTs and can therefore be very useful to estimate equity relevant subgroup effects required for our DCEA. Of course, real-world data is not available for a new intervention but will be available for established

interventions that can provide indirect evidence regarding the differences between subgroups or the exchangeability between trial and real-world effects potentially applicable to the new intervention as well. However, relative treatment effect estimates derived from comparative observational studies are typically at greater risk of bias than those obtained from RCTs. As such, we do not want to replace RCT evidence with observational evidence to inform relative treatment effect estimates but use both sources of information wisely in the evidence synthesis [66].

With the Bayesian approach to evidence synthesis, we need to define prior distributions for the model parameters to be estimated. One approach to consider RCT and observational evidence simultaneously is to use the relative treatment effect estimates for the equity relevant subgroups obtained from observational data to define informative prior distributions for the relative treatment effect and interaction effect parameters in models 1-9. For example, a typical non-informative prior distribution for d_{Ak} in model 1 for a specific subgroup analysis is $d_{Ak} \sim Normal(0, 10^4)$, but can be replaced with an informative distribution according to $d_{Ak} \sim Normal\left(d_{Ak_{obs}}, V_{Ak_{obs}}\right)$. $d_{Ak_{obs}}$ is the relative treatment effect estimate obtained from observational data and $V_{Ak_{obs}} = \frac{\sigma^2_{Ak_{obs}}}{\alpha}$ with $\sigma^2_{Ak_{obs}}$ the corresponding variance and α a factor to define the weight the observational evidence will have in the synthesis [66,67]. If $\alpha = 1$, RCTs and observational studies carry the same weight in the overall estimate of d_{Ak}. If $0 < \alpha < 1$ the observational studies have less weight than the RCTs to accommodate concerns of greater bias in the relative treatment effect estimates based on observational data. Sensitivity analyses regarding α are recommended.

If we have some information or informed belief about the extent of bias in observational studies relative to RCTs, we can define the prior for the relative treatment effects according to [66]:

$$d_{Ak} \sim Normal\left(d^*_{Ak_{obs}}, V^*_{Ak_{obs}}\right) \quad (10)$$

with $d^*_{Ak_{obs}} = d_{Ak_{obs}} - \omega$, $\omega \sim Normal(0, \sigma^2_{bias})$ and $V^*_{Ak_{obs}} = V_{Ak_{obs}} + \sigma^2_{bias}$.

ω represents the bias in observational evidence, which can be obtained from external evidence. σ^2_{bias} is the variance of this bias estimate. In line with the approach described by Welton et al

[66], the expected value for the bias is set at zero to indicate that we do not know the direction of the bias, but by incorporating σ^2_{bias} in the prior distribution for the relative treatment effects, the observational evidence is downweighed according to concerns about bias. This approach cannot only be applied to the standard NMA (equation 1) but to models 2-9 as well. For example, with the meta-regression-based synthesis (equation 4) we can define informative prior distributions for d_{Ak} and β_{Ak}. When the impact of effect-modifiers is assumed to be exchangeable or shared between interventions, the observational evidence can contribute to subgroup effects for the new intervention as well, even though no real-world data is (yet) available for it.

Another commonly used approach to combine RCT evidence and observational evidence is with hierarchical models [66]. With such a model we get different estimates for the relative treatment effects for RCTs and observational studies, but these are related given the hierarchical structure of the model (similar to shrinkage estimation). As such, unstable RCT-based estimates will gain precision given the additional information from observational studies due to the assumption of exchangeability across study designs. In such a hierarchical model we can also include factors to downweigh the impact of observational studies given its potential bias.

Relative treatment effects obtained from observational evidence may not only be different from RCT estimates due to internal bias and the difference in the study populations but also due to suboptimal adherence, which may be relevant for the DCEA. Depending on the extent the observational study has been adjusted for internal and external bias relative to the corresponding RCTs, any remaining difference in relative treatment effect estimates may reflect the impact of suboptimal adherence. If this modifying effect is assumed to be the same for all interventions, it can be used to predict how RCT-based relative treatment effects for a new intervention will translate to a routine practice setting.

Expert elicitation

If the new intervention is deemed too different from the competing interventions and the assumption that the impact of equity relevant effect-modifiers is the same for all interventions cannot be defended, then we cannot use external evidence from RCTs or observational studies from the competing interventions to improve the precision of subgroup effects for the new intervention. This is the case if the new intervention is more efficacious than the competing

interventions in the presence of certain biomarkers that are more prevalent in certain equity-relevant subgroups. As an alternative, we can use formal expert elicitation to define informative prior distributions to improve the estimation of subgroup-specific relative treatment effects for the new intervention [68-70]. More specifically, expert judgment is used to describe uncertainties associated with the relative treatment effects of the new interventions in the equity relevant subgroups. If conducted appropriately, credible estimates can be obtained. This requires an explicit process of extracting subjective and implicit knowledge or beliefs of experts about these subgroup effects of interest and representing it with probability distributions. A well-established approach for expert elicitation in the context of health economics is SHELF [71-73]. In essence, expert judgment is combined with empirical treatment effect data from RCTs in a formal and reproducible manner to improve the estimation of equity relevant treatment effects.

Estimation of baseline effects

Evidence for the absolute effects with the reference intervention or outcomes associated with the natural course of the disease in the absence of treatment required for the DCEA is more prone to external bias than evidence for relative treatment effects. Absolute effects do not reflect the target equity relevant subgroups if there are differences in prognostic factors and effect-modifiers between the study populations of the available studies and the target populations. Representative evidence for the outcomes of interest in routine practice can be expected to be available for the reference intervention since it is likely to represent a standard of care. With access to IPD, which is more likely to be accessible for real-world data sources than RCTs, we can estimate baseline effect estimates representative of the subgroups of interest, assuming these data sources have collected the relevant patient characteristics. If multiple IPD data sets are available, we can use generalized linear mixed models to obtain "pooled" estimates for the outcomes by subgroup of interest. If there is no access to IPD and we need to estimate the absolute effects based on aggregate-level data from publications of observational cohort studies or registries, models similar to equation 1, 3, 4, and 9 are modified such that they reflect absolute effects and the impact of prognostic factors, rather than relative treatment effects and effect-modifiers. For

example, we can use a model for the simultaneous analysis of subgroup specific absolute effects assuming exchangeability across subgroups (modified from equation 3):

$$g(\gamma_{is}) = \theta_{is} = \mu_{is}$$
$$\mu_{is} \sim Normal(\mathrm{M}_s, \sigma_s^2)$$
$$\mathrm{M}_s \sim Normal(\mathrm{N}, \sigma^2) \quad (11)$$

where μ_{is} is the expected absolute effect with the reference intervention in subgroup *s* of study *i*. $\mathrm{M}_s$is the overall pooled effect in subgroup *s* and subgroup-specific variance σ_s^2, which are drawn from a common normal distribution with mean N and variance σ^2.

A meta-regression analysis of absolute effects can be performed with the following model (modified from equation 4):

$$g(\gamma_i) = \theta_i = \mu_i + \beta(m_i - x_{target})$$
$$\mu_i \sim Normal(\mathrm{M}, \sigma^2) \quad (12)$$

where m_i is the study-level value of the covariate of interest in study *i*, β represents its effect on the outcome of interest. x_{target} is the centered covariate value representing the target subgroup of interest. M is the pooled absolute effect with the reference intervention for the target subgroup of interest.

EXAMPLE

In this section, the different evidence synthesis methods for aggregate-level data are illustrated with a hypothetical, yet realistic example. The analyses were performed in a Bayesian framework with non-informative prior distributions unless otherwise stated.

Imagine we want to perform a DCEA of four alternative interventions A, B, C, and D indicated for a certain condition with A the overall reference intervention and D the new intervention. These four interventions have been compared in multiple RCTs. Real-world observational evidence is available for A, B, and C. For the DCEA we are interested in three equity relevant

subgroups: population 1, population 2, and population 3. The outcome of interest is treatment response, a dichotomous endpoint. The baseline effects in routine practice with intervention A are about 25%, 20%, and 15% in subgroup 1, 2, and 3 respectively. The efficacy of intervention B relative to A is consistent across the equity relevant subgroups. Intervention C is more efficacious than B, with relative treatment effects greater in subgroup 1 than in subgroup 2 and subgroup 3. New intervention D is the most efficacious with the greatest efficacy in subgroup 3.

The nine available RCTs for this hypothetical example are presented in Table 1. There are three AB trials, two AC trials, two BC trials, an AD trial, and a three-arm ABD trial. Five trials included a heterogeneous population of subgroups 1,2, and 3. Two trials included a heterogeneous population of subgroup 1 and 2. Another two trials included a heterogeneous population of subgroup 2 and 3. Only three out of the nine trials reported subgroup data for each of the equity relevant subgroups of interest. The sample size and number of responders for each study arm stratified by subgroup are listed where available. The proportion of each subgroup in each trial as reported is listed as well. This hypothetical evidence base can be considered representative of the information that is typically available for aggregate-level data evidence synthesis.

[Insert Table 1]

Results of the analyses with different models (1, 2, 3, 4, 9, and 10) are presented in Table 2. In the first row of the table, we see the estimated relative treatment effects with interventions B, C, and D relative to A for each of the three subgroups obtained with an NMA by subgroup if all studies would have reported subgroup results for each of the subgroups: the "benchmark estimates". The efficacy of intervention B is consistent across subgroups. Efficacy of intervention C and D are heterogeneous across subgroups. Given the heterogeneous relative treatment effects, applying the all-comers average relative treatment effects to these subgroups in a DCEA would not be appropriate.

In rows 2-6 of Table 2 we see the relative treatment effects obtained with the alternative methods based on the RCT data that is available as reported in Table 1. With the standard NMA by subgroup (model 1) we get similar results as the benchmark results, but the 95% credible intervals (95%CrI; Bayesian equivalent to 95% confidence intervals) are wider because we can

only use the subset of studies for which subgroup results were reported (See Table 1). For subgroup 3 we do not have an estimate for intervention C because no subgroup results were reported.

If we use an evidence synthesis model with the assumption of exchangeable treatment effects (model 2), we see that the contrast between interventions regarding relative treatment effects for each of the subgroups is reduced. In comparison to results obtained with model 1, the estimates for intervention B have shifted upwards a bit and estimates for intervention D have shifted downwards a bit. The treatment-specific estimates have shrunken towards the average effects across interventions. These changes are not statistically significant but do result in smaller 95%CrIs, closer to the benchmark results.

When the analyses are performed under the assumption of exchangeable subgroup effects (model 3), we see that the contrast in relative treatment effects between interventions is not really reduced, but the differences between subgroups are somewhat. The benefit is that we get more precise estimates (i.e. smaller 95%CrI), closer to the benchmark results. The additional benefit is that we do get an estimate for intervention C for subgroup 3, albeit it is very uncertain.

When we perform a conventional meta-regression analysis (model 4) using specific subgroup data, where available, and otherwise the mixed population data, we get the results presented in row 5 of Table 2. The meta-regression model has two covariates representing the difference in the log-odds ratio between subgroup 2 and subgroup 1 and the difference between subgroup 3 and subgroup 1. Overall, relatively precise estimates are obtained given the data reported. However, the contrast in relative treatment effect estimates between interventions and between subgroups is reduced because we had to assume that the impact of the effect-modifiers associated with the subgroup was the same for intervention B, C, and D relative to A due to the limited number of studies available. The actual benchmark results show that the trend in relative treatment effects for intervention C versus A is the opposite of the trend for intervention B versus A, thereby canceling each other out in the meta-regression analysis resulting in relatively similar estimates across the subgroups.

An analysis using the same aggregate level data as with the conventional meta-regression analysis but according to the structural assumptions of model 9 in line with the principles of ML-

NMR, we see results closer to the benchmark results than with the conventional meta-regression analysis. Again, the analyses are limited by the structural assumption that the difference in relative treatment effects between subgroups (on the log odds ratio scale) is the same for interventions B, C, and D.

Finally, a network meta-regression based on RCT and observational real-world evidence according to the principles expressed in equation 10 was performed. The relative treatment effect estimates of interventions B and C relative to A were strengthened with evidence about the treatment effects of these interventions in routine practice. Observational studies showed odds ratios of response that were 25% smaller than those observed in RCTs. The variance was reduced by about 40% due to the larger sample size in real-world data studies. Given the use of external observational evidence, treatment-specific interaction effects could be used (which was not feasible with the meta-regression based on the nine RCTs). To help improve the estimates for new intervention D across subgroups, it was assumed that the effect-modification for intervention D would not vary more than 2 times the effect-modification seen with intervention B. (Let's say this information was obtained employing a formal expert elicitation exercise.) A study by Ioannidis et al. [74] of 19 meta-analyses comparing RCT with observational study results provided an estimate for the variance of the bias (σ^2_{bias}) in observational studies, which was used to downweigh the observational evidence according to equation 10 [66]. Results of this generalized evidence synthesis are presented in the last row of Table 2 and show estimates closer to the benchmark analysis than any of the other approaches.

[Insert Table 2]

SOME PRACTICAL RECOMMENDATIONS

Unfortunately, detailed recommendations on which evidence synthesis method to use for which data gap scenario cannot be provided at this stage. Comprehensive simulation studies are needed to better understand the performance of the different evidence synthesis methods given the specific challenges associated with subgroup effects according to health equity-related constructs and data availability. That being said, a few practical recommendations are outlined here to

estimate relative treatment effects and baseline effects based on existing data to parameterize a health economic model to perform a DCEA.

Once equity relevant subgroups of interest have been defined, the first step is to determine how this translates into the relevant (levels of the) effect-modifiers and prognostic factors to consider in the evidence synthesis. In general, we want to use IPD as much as possible. With access to IPD, we can define the effect-modifiers and prognostic factors more specifically. With aggregate-level data, we have limited ability to adjust for differences between characteristics of the study populations of the relevant individual studies and the target subgroups of interest for the DCEA. As mentioned, a distinction needs to be made between estimating relative treatment effects and the baseline effects for the equity relevant subgroups. For the former, we only need to worry about differences in effect-modifiers between groups, and for the latter, we need to worry about differences in prognostic factors as well. The final (levels of the) effect modifiers and prognostic factors will be a trade-off between the relevance for the decision problem and data availability.

Given the greater risk of external bias in absolute effects than relative treatment effects, it is recommended to attempt obtaining access to real-world IPD to estimate the baseline effects with the reference intervention or standard of care for the subgroups of interest. If this is not feasible, sufficiently detailed aggregate-level data needs to be obtained from published studies.

Given the potential challenges in estimating subgroup-specific relative treatment effects, it makes sense to first assess whether it is even worthwhile to do so given the baseline effects for the equity relevant subgroups of interest. When the health economic model is parameterized with subgroup specific baseline effect estimates, we can assess whether using relative treatment effects that differ by subgroup results in meaningful different estimates of cost-effectiveness and health equity impact in comparison to using the same relative treatment effect estimate for all subgroups (based on the all-comers trial populations). If the differences in estimates do impact the conclusion which intervention to prefer, it is worthwhile to proceed with estimating equity relevant subgroup specific model input parameters regarding the relative treatment effects.

It is recommended to use multiple evidence synthesis methods to estimate the relative treatment effects of interest given the potential sensitivity of the estimates to the method of choice and the

structure of the statistical model. If IPD is available for a subset of the RCTs, it is recommended to use ML-NMR-based approaches. Simulation studies have shown that estimates for the target populations of interest obtained with ML-NMR are greatly improved over other methods to adjust for external bias [41]. When multiple competing models are considered for the same RCT-based data set, model fit criteria, such as the deviance information criterion can be used to help identify the most parsimonious models for the data at hand.

For the final set of evidence synthesis models considered most appropriate for the information available from RCTs and the obtained subgroup-specific estimates of relative treatment effects, we want to evaluate whether using observational evidence can improve the precision of the estimates while explicitly acknowledging its potential internal bias in the synthesis model.

For a new intervention, there is only RCT evidence available. Its subgroup-specific relative treatment effect estimates can potentially be improved if the assumption of exchangeable subgroup effects is defendable or if there are certain similarities with (some of) the alternative interventions that impact the relative treatment effects. If this is the case, we can use shrinkage estimation assuming exchangeability across (some) interventions or rely on the shared or exchangeable effect-modifiers assumption. With these approaches, any beneficial information from real-world evidence for the competing established interventions will also "transfer" to the new intervention. When considering these approaches, a trade-off needs to be made between potentially biasing the subgroup-specific estimates of the new intervention and precision gains. If the new intervention is deemed too different then these structural assumptions of the evidence synthesis models are inappropriate. As a last resort, formal expert elicitation can be considered to improve subgroup-specific estimates for the new health technology.

Finally, with subgroup-specific relative treatment effect estimates for all the competing interventions of interest available based on what have been deemed the most appropriate methods for the evidence at hand, it is recommended to perform multiple DCEAs to understand their impact on cost-effectiveness and health equity impact estimates.

CONCLUSION

DCEA is an interesting methodology to incorporate health equity concerns into cost-effectiveness analysis to facilitate trade-offs between optimizing total health outcomes and reducing health inequality given the alternative interventions available for a given condition. For a model-based DCEA, it is important to obtain input parameter estimates that are representative of the equity relevant target subgroups of interest. Arguably, it is most important to obtain the baseline effect estimates and relative treatment effect estimates stratified by the subgroups of interest. Equity-relevant subgroups are more likely to vary regarding baseline effects than relative treatment effects between interventions compared. As such, estimation of credible baseline effects requires more detailed information regarding the impact of equity relevant covariates. Luckily, for many conditions, there are real-world data sources with IPD available to estimate the baseline effects for the equity relevant subgroups. On the other hand, estimation of relative treatment effects relevant for these subgroups is likely more challenging despite the fact that less specific information regarding the impact of covariates is needed. We prefer estimating relative treatment effects based on RCT evidence because of its more favorable ratio of lack of internal bias versus external bias. For new health technologies, there is likely only RCT evidence available anyway. However, for RCTs there are evidence challenges due to lack of inclusion of relevant subgroups, lack of access to IPD for all trials, small subgroups resulting in uncertain subgroup effects, and reporting gaps. Advanced Bayesian evidence synthesis methods can help improve equity relevant relative treatment effect estimates given these evidence challenges.

Relevant evidence synthesis methods include shrinkage estimation, conventional (network) meta-regression analysis, multi-level (network) meta-regression analysis, and generalized evidence synthesis. Depending on the available evidence base and the extent IPD is available, bespoke models with relevant elements of different approaches may be required. Furthermore, formal expert elicitation to obtain prior distributions for Bayesian model parameters is worthwhile to improve estimates of treatment effects.

For health disparities to be systematically considered and incorporated in value assessment of new interventions, guidance is needed for how a rigorous model-based DCEA can be performed despite gaps in evidence equity-relevant subgroups. Although this paper provides an overview of advanced evidence synthesis methods to help overcome these gaps, future research (i.e.

simulation studies) is needed to assess pros and cons of different methods for different data gap scenarios.

DECLARATION

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article

Competing interests

The author has no conflicts of interest related to the current manuscript.

Funding

This work was supported by the UCSF Academic Senate Committee on Research (Academic Senate RAP Grant) and grants from the National Cancer Institute of the NIH (R01 CA221870) and from the National Human Genome Research Institute (R01HG011792). The funders had no role in the preparation, review, or approval of the manuscript or the decision to submit the manuscript for publication.

Authors' contributions

The author developed the content, performed the example analyses, and drafted the manuscript.

Acknowledgements

Not applicable

REFERENCES

1. Asaria, M., Griffin, S., & Cookson, R. (2016). Distributional cost-effectiveness analysis: a tutorial. Medical Decision Making, 36(1), 8-19.
2. Asaria, M., Griffin, S., Cookson, R., Whyte, S., & Tappenden, P. (2015). Distributional cost-effectiveness analysis of health programs–a methodological case study of the UK bowel cancer screening program. Health economics, 24(6), 742-754.
3. Cookson, R., & Mirelman, A. J. (2017). Equity in HTA: what doesn't get measured, gets marginalized Israel journal of health policy research, 6(1), 1-4.
4. Cookson, R., Mirelman, A. J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O. F., ... & Culyer, A. J. (2017). Using cost-effectiveness analysis to address health equity concerns. Value in Health, 20(2), 206-212.
5. Arnold, M., Nkhoma, D., & Griffin, S. (2020). The distributional impact of the Malawian Essential Health Package. Health policy and planning, 35(6), 646-656.
6. Collins, B., Kypridemos, C., Cookson, R., Parvulescu, P., McHale, P., Guzman-Castillo, M., ... & O'Flaherty, M. (2020). Universal or targeted cardiovascular screening? Modelling study using a sector-specific distributional cost effectiveness analysis. Preventive medicine, 130, 105879.
7. Dawkins, B.R., Mirelman, A.J., Asaria, M., Johansson, K.A., Cookson, R.A. (2018). Distributional cost-effectiveness analysis in low- and middle-income countries: illustrative example of rotavirus vaccination in Ethiopia. Health Policy Plan, 33(3): 456-463.
8. Griffin, S., Walker, S., & Sculpher, M. (2020). Distributional cost effectiveness analysis of West Yorkshire low emission zone policies. Health economics, 29(5), 567-579..
9. Lee, T. H., Kim, W., Shin, J., Park, E. C., Park, S., & Kim, T. H. (2018). Strategic Distributional Cost-Effectiveness Analysis for Improving National Cancer Screening Uptake in Cervical Cancer: A Focus on Regional Inequality in South Korea. Cancer research and treatment: official journal of Korean Cancer Association, 50(1), 212.
10. Love-Koh, J., Griffin, S., Kataika, E., Revill, P., Sibandze, S., & Walker, S. (2020). Methods to promote equity in health resource allocation in low-and middle-income countries: an overview. Globalization and health, 16(1), 1-12.

11. Ades, A. E., Welton, N. J., Caldwell, D., Price, M., Goubar, A., & Lu, G. (2008). Multiparameter evidence synthesis in epidemiology and medical decision-making. Journal of health services research & policy, 13(3_suppl), 12-22.
12. Dias, S., Ades, A.E., Welton, N.J., Jansen, J.P., Sutton, A.J. (2018a) Introduction to evidence synthesis. In: Network meta-analysis for decision-making, pp. 1-17. John Wiley & Sons
13. Dias, S., Ades, A.E., Welton, N.J., Jansen, J.P., Sutton, A.J. (2018b) The core model. In: Network meta-analysis for decision-making, pp. 19-58. John Wiley & Sons
14. Espinoza, M. A., Manca, A., Claxton, K., & Sculpher, M. J. (2014). The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. Medical Decision Making, 34(8), 951-964.
15. Guyot, P., Ades, A. E., Beasley, M., Lueza, B., Pignon, J. P., & Welton, N. J. (2017). Extrapolation of survival curves from cancer trials using external information. Medical Decision Making, 37(4), 353-366.
16. Henderson, N. C., Louis, T. A., Wang, C., & Varadhan, R. (2016). Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research. Health Services and Outcomes Research Methodology, 16(4), 213-233.
17. Jansen, J. P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., ... & Cappelleri, J. C. (2011). Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value in Health, 14(4), 417-428.
18. Jansen, J. P., Trikalinos, T., Cappelleri, J. C., Daw, J., Andes, S., Eldessouki, R., & Salanti, G. (2014). Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value in Health, 17(2), 157-173.
19. Jansen, J.P., 2020. Heterogeneity and subgroup analysis in network meta-analysis. In: Ting, N., Cappelleri, J., Ho, S., Chen, G. (Eds.), Design and analysis of subgroups with biopharmaceutical applications. Emerging topics in statistics and biostatistics. Springer, pp. 369-385.

20. Sculpher, M. (2008). Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics, 26(9), 799-806.
21. Spiegelhalter, D. J., & Best, N. G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Statistics in medicine, 22(23), 3687-3709.
22. Turner, R. M., Spiegelhalter, D. J., Smith, G. C., & Thompson, S. G. (2009). Bias modelling in evidence synthesis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 172(1), 21-47.
23. Warren, F. C., Abrams, K. R., & Sutton, A. J. (2014). Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. Statistics in medicine, 33(14), 2449-2466.
24. Welton, N. J., & Ades, A. E. (2005). Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Medical Decision Making, 25(6), 633-645.
25. Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., & Welton, N. J. (2018). Methods for population-adjusted indirect comparisons in health technology appraisal. Medical Decision Making, 38(2), 200-211.
26. Ioannidis, J. P., Powe, N. R., & Yancy, C. (2021). Recalibrating the Use of Race in Medical Research. JAMA.
27. FDA, Guidance for Industry and Food and Drug Administration Staff, Collection of Race and Ethnicity Data in Clinical Trials (Oct. 2016). https://www.fda.gov/media/75453/download
28. FDA, Guidance for Industry, Enhancing the Diversity of Clinical Trial Populations - Eligibility Criteria, Enrollment Practices, and Trial Designs (Nov. 2020) https://www.fda.gov/media/127712/download
29. Ades, A. E. (2003). A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. Statistics in medicine, 22(19), 2995-3016.
30. Bucher, H. C., Guyatt, G. H., Griffith, L. E., & Walter, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of clinical epidemiology, 50(6), 683-691.

31. Dias, S., Sutton, A. J., Ades, A. E., & Welton, N. J. (2013a). Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making, 33(5), 607-617.
32. Dias, S., Ades, A.E., Welton, N.J., Jansen, J.P., Sutton, A.J. (2018c) Validity of network meta-analysis. In: Network meta-analysis for decision-making, pp. 351-374. John Wiley & Sons
33. Hutton, B., Salanti, G., Caldwell, D. M., Chaimani, A., Schmid, C. H., Cameron, C., ... & Mulrow, C. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Annals of internal medicine, 162(11), 777-784.
34. Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. Statistics in medicine, 21(16), 2313-2324.
35. Lu, G., & Ades, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in medicine, 23(20), 3105-3124.
36. Salanti, G., Higgins, J. P., Ades, A. E., & Ioannidis, J. P. (2008). Evaluation of networks of randomized trials. Statistical methods in medical research, 17(3), 279-301.
37. Dias, S., Sutton, A. J., Welton, N. J., & Ades, A. E. (2013b). Evidence synthesis for decision making 3: heterogeneity—subgroups, meta-regression, bias, and bias-adjustment. Medical Decision Making, 33(5), 618-640.
38. Dias, S., Ades, A.E., Welton, N.J., Jansen, J.P., Sutton, A.J. (2018d) Meta-regression for relative treatment effects. In: Network meta-analysis for decision-making, pp. 227-271. John Wiley & Sons
39. Jansen, J. P. (2012). Network meta-analysis of individual and aggregate-level data. Research Synthesis Methods, 3(2), 177-190.
40. Jansen, J. P., & Cope, S. (2012). Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. BMC medical research methodology, 12(1), 1-16.
41. Phillippo, D. M., Dias, S., Ades, A. E., Belger, M., Brnabic, A., Schacht, A., ... & Welton, N. J. (2020). Multilevel network meta-regression for population-adjusted treatment comparisons. Journal of the Royal Statistical Society: Series A (Statistics in Society), 183(3), 1189-1210.

42. Jansen, J. P., & Naci, H. (2013). Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC medicine, 11(1), 159.
43. Kew, K. M., Dias, S., & Cates, C. J. (2014). Long-acting inhaled therapy (beta-agonists, anticholinergics and steroids) for COPD: a network meta-analysis. Cochrane Database of Systematic Reviews, (3).
44. Lipsky, A. M., Gausche-Hill, M., Vienna, M., & Lewis, R. J. (2010). The importance of "shrinkage" in subgroup analyses. Annals of emergency medicine, 55(6), 544-552.
45. Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. The Lancet Psychiatry, 1(5), 368-376.
46. Cooper, N. J., Sutton, A. J., Morris, D., Ades, A. E., Welton, N. J. (2009). Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Statistics in medicine, 28(14), 1861-1881.
47. Donegan, S., Welton, N. J., Tudur Smith, C., D'Alessandro, U., & Dias, S. (2017). Network meta-analysis including treatment by covariate interactions: Consistency can vary across covariate values. Research synthesis methods, 8(4), 485-495.
48. Donegan, S., Dias, S., & Welton, N. J. (2018). Assessing the consistency assumptions underlying network meta-regression using aggregate data. Research synthesis methods. doi: 10.1002/jrsm.1327
49. Nixon, R. M., Bansback, N., & Brennan, A. (2007). Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Statistics in medicine, 26(6), 1237-1254.
50. Berlin, J. A., Santanna, J., Schmid, C. H., Szczech, L. A., & Feldman, H. I. (2002). Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Statistics in medicine, 21(3), 371-387.
51. Higgins, J. P., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. Statistics in medicine, 23(11), 1663-1682.

52. Lambert, P. C., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of clinical epidemiology, 55(1), 86-94.
53. Schmid, C. H., Stark, P. C., Berlin, J. A., Landais, P., & Lau, J. (2004). Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. Journal of clinical epidemiology, 57(7), 683-697.
54. Riley, R. D., & Steyerberg, E. W. (2010). Meta-analysis of a binary outcome using individual participant data and aggregate data. Research Synthesis Methods, 1(1), 2-19.
55. Greenland, S. (2002). A review of multilevel theory for ecologic analyses. Statistics in medicine, 21(3), 389-395.
56. Jackson, C., Best, N., & Richardson, S. (2006). Improving ecological inference using individual-level data. Statistics in medicine, 25(12), 2136-2159.
57. Jackson, C., Best, A. N., & Richardson, S. (2008). Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. Journal of the Royal Statistical Society: Series A (Statistics in Society), 171(1), 159-178.
58. Cope, S., Capkun-Niggli, G., Gale, R., Lassen, C., Owen, R., Ouwens, M. J., ... & Jansen, J. P. (2012). Efficacy of once-daily indacaterol relative to alternative bronchodilators in COPD: a patient-level mixed treatment comparison. Value in Health, 15(3), 524-533.
59. Debray, T. P., Schuit, E., Efthimiou, O., Reitsma, J. B., Ioannidis, J. P., Salanti, G., ... & GetReal Workpackage. (2018). An overview of methods for network meta-analysis using individual participant data: when do benefits arise?. Statistical methods in medical research, 27(5), 1351-1364.
60. Leahy, J., O'Leary, A., Afdhal, N., Gray, E., Milligan, S., Wehmeyer, M. H., & Walsh, C. (2018). The impact of individual patient data in a network meta-analysis: An investigation into parameter estimation and model selection. Research synthesis methods, 9(3), 441-469.
61. Donegan, S., Williamson, P., D'Alessandro, U., Garner, P., & Smith, C. T. (2013). Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: individual patient data may be beneficial if only for a subset of trials. Statistics in medicine, 32(6), 914-930.

62. Sutton, A. J., Kendrick, D., & Coupland, C. A. (2008). Meta-analysis of individual-and aggregate-level data. Statistics in medicine, 27(5), 651-669.
63. Riley, R. D., Simmonds, M. C., & Look, M. P. (2007). Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. Journal of clinical epidemiology, 60(5), 431-e1.
64. Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L., & Boutitie, F. (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Statistics in medicine, 27(11), 1870-1893.
65. Saramago, P., Sutton, A. J., Cooper, N. J., & Manca, A. (2012). Mixed treatment comparisons using aggregate and individual participant level data. Statistics in medicine, 31(28), 3516-3536.
66. Welton, N. J., Sutton, A. J., Cooper, N., Abrams, K. R., & Ades, A. E. (2012). Evidence synthesis for decision making in healthcare (Vol. 132). John Wiley & Sons.
67. Ibrahim, J. G., & Chen, M. H. (2000). Power prior distributions for regression models. Statistical Science, 15(1), 46-60.
68. Ren, S., Oakley, J. E. and Stevens, J. W. (2018). Incorporating genuine prior information about between-study heterogeneity in random effects pairwise and network meta-analyses. Medical Decision Making, 38 (4), 531-542
69. Soares, M. O., Sharples, L., Morton, A., Claxton, K., & Bojke, L. (2018). Experiences of structured elicitation for model-based cost-effectiveness analyses. Value in health, 21(6), 715-723.
70. Iglesias, C. P., Thompson, A., Rogowski, W. H., & Payne, K. (2016). Reporting guidelines for the use of expert judgement in model-based economic evaluations. Pharmacoeconomics, 34(11), 1161-1172.
71. O'Hagan, A. (2014) SHELF: the Sheffield elicitation framework (version 4.0). http://www.tonyohagan.co.uk/shelf/SHELF4.html
72. Gosling, J. P. (2018). SHELF: the Sheffield elicitation framework. In Elicitation (pp. 61-93). Springer, Cham.
73. O'Hagan, A. (2019). Expert knowledge elicitation: subjective but scientific. The American Statistician, 73(sup1), 69-81.

74. Ioannidis, J. P., Haidich, A. B., Pappa, M., Pantazis, N., Kokori, S. I., Tektonidou, M. G., ... & Lau, J. (2001). Comparison of evidence of treatment effects in randomized and nonrandomized studies. Jama, 286(7), 821-830.

FIGURES

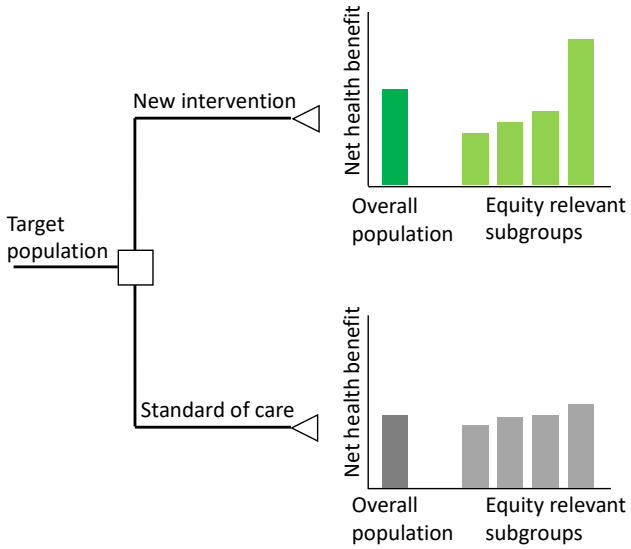

Figure 1: Foundation of DCEA: estimation of the net health benefits for the equity relevant subgroups

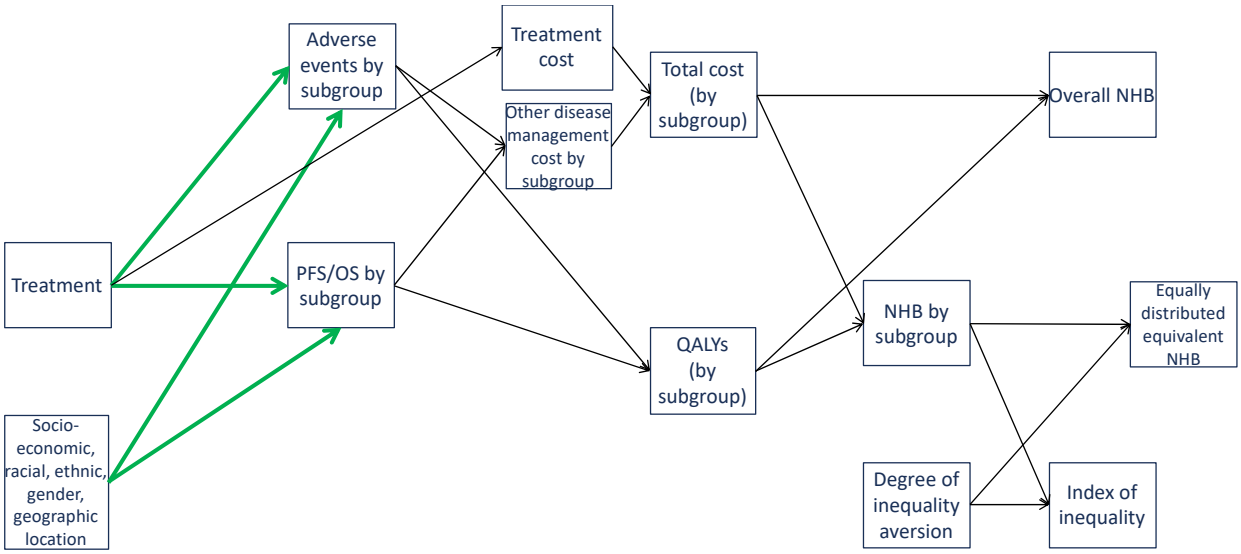

Figure 2: Influence diagram for a fictitious health economic model to perform a distributional cost-effectiveness analysis. Arrows in green represent relative treatment effects and outcomes in the baseline arm of the model representative of the equity relevant subgroups of interest.

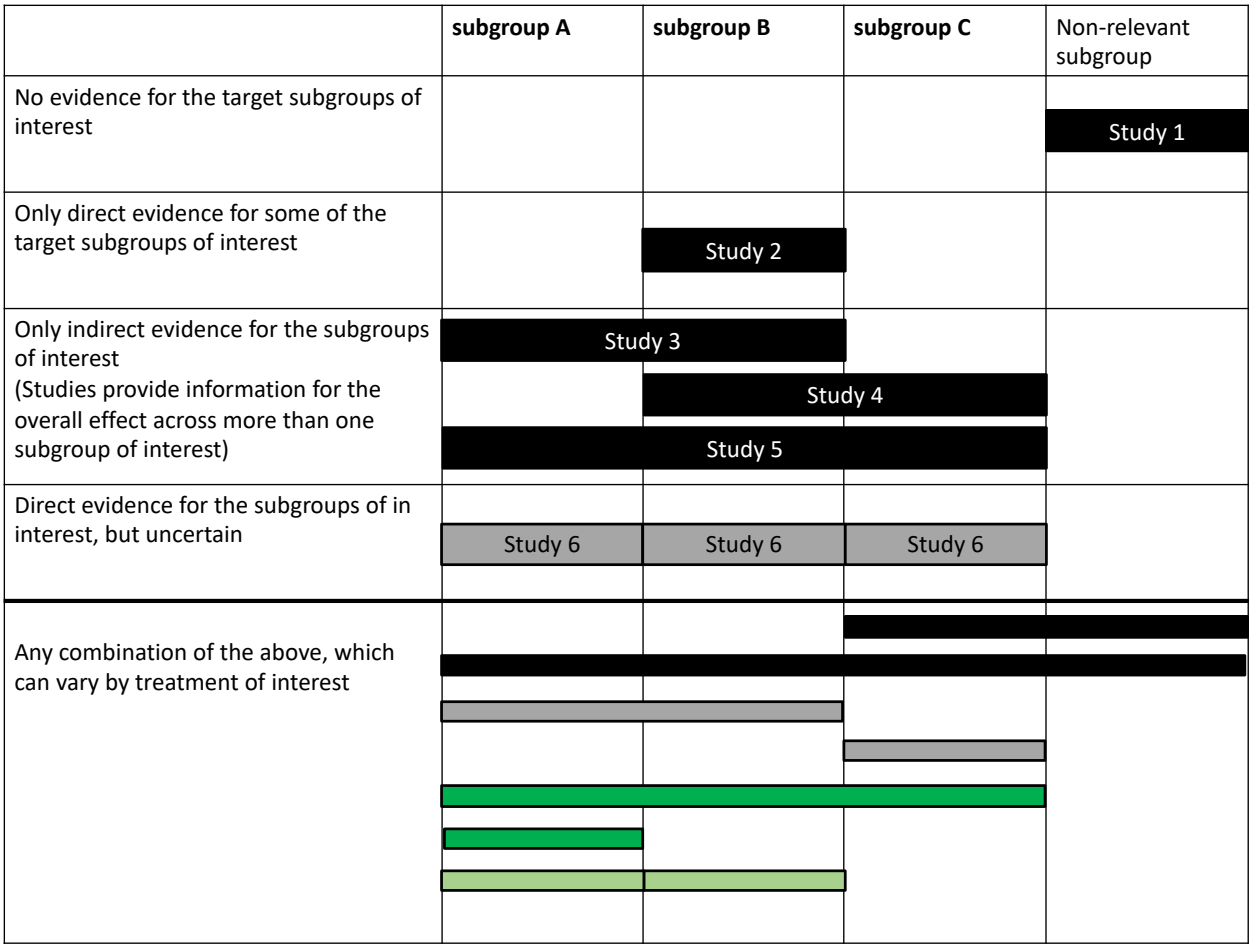

Figure 3: Evidence challenges for baseline and relative treatment effects for the equity relevant subgroups. Lighter color bars reflect uncertain evidence due to small sample sizes. Black/grey and green represent evidence for two different treatments. Bars crossing multiple subgroup columns (e.g. study 3) reflect studies that only report results for a combined population and not for the specific subgroups of interest. (In contrast, study 6 provides results for all subgroups of interest.

TABLES

Table 1: Reported information from nine fictitious randomized clinical trials (by row) regarding sample size (n) and response (r) with interventions A, B, C, D for each of the study-arms. Intervention A is the overall reference intervention and intervention D is the new intervention.

Interventions in arm 1, 2, or 3 of each trial			All (Subgroups 1,2,3 combined)						Subgroup 1+2 combined						Subgroup 2+3 combined						Subgroup 1						Subgroup 2						Subgroup 3						Proportion subgroup in trial		
1	2	3	n1	n2	n3	r1	r2	r3	n1	n2	n3	r1	r2	r3	n1	n2	n3	r1	r2	r3	n1	n2	n3	r1	r2	r3	n1	n2	n3	r1	r2	r3	n1	n2	n3	r1	r2	r3	1	2	3
A	B		70	70		15	22																																0.4	0.3	0.3
A	B		200	200		43	62														120	120		29	41		50	50		9	14		30	30		5	7		0.6	0.25	0.15
A	B														200	200		36	52																			0	0.5	0.5
A	C		150	150		33	67																															0.6	0.3	0.1
A	C								150	150		34	64								75	75		19	40		75	75		15	24							0.5	0.5	0
B	C														100	100		26	28																			0	0.5	0.5
B	C		100	100		34	46																															0.6	0.3	0.1
A	D								100	100		23	49																									0.7	0.3	0
A	B	D	150	150	150	30	47	66													60	60	60	15	23	29	45	45	45	8	14	20	45	45	45	7	10	17	0.4	0.3	0.3

Table 2: Relative treatment effect estimates regarding response with interventions B, C, and D versus A expressed as odds ratios for equity relevant subgroups as obtained with alternative evidence synthesis methods based on the available information. For comparative purposes, "benchmark results" are provided that could have been obtained if all studies would have reported information for all subgroups of interest.

		Overall population		**Subgroup 1**		**Subgroup 2**		**Subgroup 3**	
		Odds Ratio	95%CrI	Odds Ratio	95%CrI	Odds Ratio	95%CrI	Odds Ratio	95%CrI
Randomized controlled trials									
If all information by subgroup would have been reported/available									
Standard NMA (model 1)	Tx A			reference		ref		ref	
	Tx B			1.7	(1.2, 2.5)	1.7	(1.1, 2.5)	1.7	(1.1, 2.9)
	Tx C			3.4	(2.3, 5.2)	2.2	(1.3, 3.6)	1.6	(0.6, 3.8)
	Tx D			2.8	(1.7, 4.5)	3.7	(1.9, 7.1)	3.6	(1.5, 8.6)
Based on available information									
Standard NMA (model 1)	Tx A	ref		ref		ref		ref	
	Tx B	1.7	(1.3, 2.3)	1.7	(1.1, 2.7)	1.9	(1.0, 3.9)	1.6	(0.7, 3.7)
	Tx C	2.9	(1.9, 4.3)	3.4	(1.7, 6.9)	1.9	(0.9, 4.1)	-	
	Tx D	3.0	(1.9, 4.7)	2.7	(1.4, 5.5)	3.6	(1.5, 8.6)	3.4	(1.3, 8.7)
Shrinkage estimation, exchangeability of treatments (model 2)	Tx A			ref		ref		ref	
	Tx B			1.9	(1.2, 2.9)	2.0	(1.1, 3.7)	1.6	(0.7, 3.8)
	Tx C			2.9	(1.6, 6.0)	2.0	(1.0, 3.9)	-	
	Tx D			2.6	(1.5, 4.9)	2.9	(1.4, 7.1)	3.1	(1.3, 8.2)
Shrinkage estimation, exchangeability of subgroups (model 3)	Tx A			ref		ref		ref	
	Tx B			1.7	(1.2, 2.6)	1.8	(1.1, 3.1)	1.7	(0.9, 2.9)
	Tx C			3.1	(1.7, 6.2)	2.2	(1.0, 4.3)	2.6	(0.2, 34.7)
	Tx D			2.9	(1.6, 5.2)	3.3	(1.7, 6.6)	3.2	(1.7, 6.7)
Conventional network meta-regression; shared effect-modifier (model 4)	Tx A			ref		ref		ref	
	Tx B			2.0	(1.4, 2.8)	1.7	(1.1, 2.8)	1.6	(0.9, 3.0)
	Tx C			2.5	(1.7, 3.6)	2.2	(1.4, 3.6)	2.0	(1.0, 4.0)
	Tx D			3.4	(2.2, 5.3)	3.1	(1.8, 5.3)	2.8	(1.4, 5.7)
Network meta-regression with subgroup aggregate-level data integration (model 9)	Tx A			ref		ref		ref	
	Tx B			2.1	(1.5, 3.1)	1.7	(1.0, 2.8)	1.8	(1.0, 3.2)
	Tx C			3.1	(2.0, 4.8)	2.4	(1.4, 4.3)	2.7	(1.3, 5.4)
	Tx D			3.9	(2.5, 6.2)	3.1	(1.7, 5.5)	3.3	(1.7, 6.4)
Randomized controlled trials supplemented with real-world observational evidence									
Based on available information									
Generalized evidence synthesis (model 4 + eq. 10)	Tx A			ref		ref		ref	
	Tx B			1.7	(1.3, 2.4)	1.8	(1.2, 3)	1.5	(0.9, 2.6)
	Tx C			2.8	(1.9, 4.1)	1.9	(1.1, 3.3)	1.6	(0.7, 3.7)
	Tx D			3.1	(2.1, 4.8)	3.4	(2.0, 5.8)	2.8	(1.5, 5.2)