

Technical support document (TSD) 20 in action

What to consider when using multivariate meta-analyses or network meta-analyses (NMA) to support health technology assessment (HTA) oncology submissions

Patricia Guyot, PhD, Sanofi, Chilly-Mazarin, France (moderator)*

Miranda Cooper, MSc, Lumanity, Sheffield, UK (panellist)

Oriana Ciani, PhD, SDA Bocconi School of Management, Milan, Italy (panellist)

Sylwia Bujkiewicz, PhD, University of Leicester, Leicester, UK (panellist)

AGENDA

01 Introduction

PATRICIA GUYOT (10 MINUTES)

02 Motivation for using these methods and their potential use for HTA submissions

MIRANDA COOPER (11 MINUTES)

03 Theory of the methods described in TSD 20

SYLWIA BUJKIEWICZ (12 MINUTES)

04 Framework for handling surrogate outcomes with reference to case studies

ORIANA CIANI (12 MINUTES)

05 Q&A

ALL (15 MINUTES)

Surrogate endpoints



- For some endpoints, long follow-up times are required before the treatment effect can be measured. In this case, randomized controlled trials (RCTs) report surrogate endpoints
- Surrogacy is particularly important for first-line (1L) treatments in oncology and/or cardiovascular diseases:
 - Longer trial duration for overall survival (OS)
 - Treatments have most benefit to prevent relapsed/metastatic disease
- Surrogacy includes some potential pitfalls, especially for first-in-class treatments:
 - Surrogates may not consider unmeasured benefits and harms
 - Knowledge about how a drug achieves clinical results may be incomplete

Considerations for surrogate endpoints

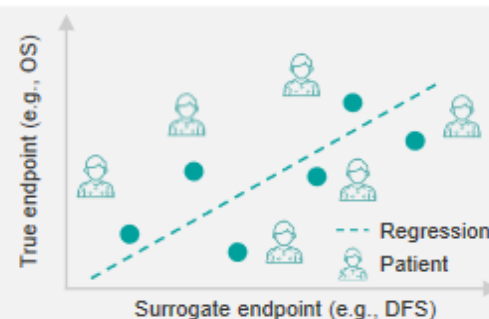
Taylor and Elston proposed a three-step framework for evaluation of a surrogate endpoint

Rationale: why surrogate endpoint is preferred (e.g., *high unmet need, rare events, long follow-up, difficult to measure true endpoint*)

Level 2

Statistical evidence:

Patient-level association: the surrogate and the true endpoint correlate on a patient-level irrespective of treatment. This can be done in a single trial and is of prognostic value (e.g. *useful for patient management*)



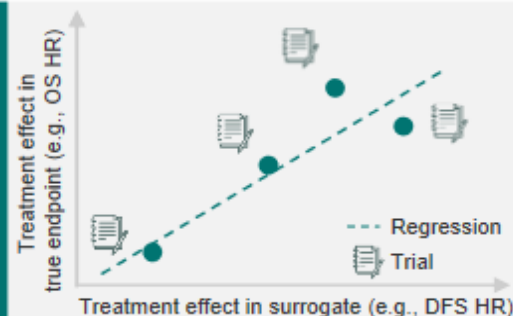
Level 3

Biological plausibility: association between disease mechanism, surrogate and true endpoint (e.g., *epidemiologic evidence, animal models, disease pathogenesis, drug mechanism*)

Level 1

Statistical evidence:

Trial-level association: treatment effect on the surrogate correlates the treatment effect on the true endpoint. Meta-analysis required, ideally with data from similar studies (class of drug, treatment duration etc). It is of predictive value and useful for assessing the benefit of new treatments



Strength of evidence:

Limited

High

Regulatory guidance on surrogate endpoints

- Regulatory agencies such as the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) in the US have introduced flexible licensing pathways, for example by allowing conditional licensing based on treatment effect measured on a surrogate endpoint
- For example, surrogate endpoints are used when clinical outcomes, like strokes, might take a long time to study, or in cases where the clinical benefit of improving the surrogate endpoint, such as controlling blood pressure, is well understood
- Between 2010 and 2012, the FDA approved 45% of new drugs based on a surrogate endpoint
- The effectiveness measured on the final outcome is obtained when data become more mature, and the drug is then re-evaluated

Surrogate endpoints across HTA

Characteristic	Total No. (%) of HTA Reports (N = 124)
Drugs	122 (98)
Medical device	2 (2)
HTA agencies	
NICE	23 (19)
HIS/SMC	20 (16)
HAS	20 (16)
PBAC/MSAC	15 (12)
CADTH	13 (10)
IQWiG/G-BA	13 (10)
ZiN	9 (7)
NIPN	11 (9)
Disease area	
Cancer	65 (52)
Cardiovascular	17 (14)
Pulmonology	8 (6)
Nephrology	8 (6)
Endocrinology	7 (6)
Infectious disease	7 (6)
Ophthalmology	6 (5)
Gastroenterology	6 (5)
Orphan status	8 (6)
Surrogate validation	
Surrogate accepted (yes)	49 (40)
Level of evidence assessed (yes)	61 (49)
Strength of association provided (yes)	27 (22)
Quantification of effect provided (yes)	40 (32)
Final recommendation given	
Approved	32 (26)
Restricted	61 (49)
Rejected	20 (16)
No recommendation	11 (9)

CADTH, Canadian Agency for Drugs and Technologies in Health; HAS, Haute Autorité de Santé; HIS/SMC, Health Improvement Scotland/Scottish Medicines Consortium; HTA, health technology assessment; IQWiG/G-BA, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen/Gemeinsame Bundesausschuss; NICE, National Institute for Health and Care Excellence; NIPN, Országos Gyógyszerészeti és Élelmezés-egészségügyi Intézet; PBAC/MSAC, Pharmaceutical Benefits Advisory Committee/Medical Services Advisory Committee; ZiN, Zorginstituut Nederland.

Ciani et al. Validity of Surrogate Endpoints and Their Impact on Coverage Recommendations: A Retrospective Analysis across International Health Technology Assessment Agencies. MDM 2021 Vol. 41(4) 439–452

Multivariate meta-analyses or network meta-analyses (NMA)

- A Bayesian framework provides flexibility to model complex data structures by allowing multiple parameters to be modelled simultaneously
- Meta-analysis and NMA are the standard tools for synthesizing evidence
- Multivariate meta-analysis or NMA allows to jointly model treatment effects on multiple correlated outcomes
- When multiple outcomes are reported, they often are synthesized separately, using standard univariate meta-analysis for each outcome
- Multivariate meta-analysis or NMA allow more studies to contribute to the meta-analysis for each outcome, which can improve efficiency and decrease bias in evidence synthesis, as well as in cost-effectiveness models

Bayesian advantages of multivariate meta-analysis

- Additional data (external data from observational studies, clinical trials or systematic reviews) or experts' opinions can be incorporated in the form of prior distributions
- This means that direct probabilistic statements can be made about the parameters of interest
- Modelling uncertainty around all relevant parameters can also be flexible

It's Time for a Poll!

What is your current level of understanding of TSD 20 and methods for surrogacy analysis, from a scale of 1 to 5 ?

- ☐ **1 – no understanding**
- ☐ **2**
- ☐ **3**
- ☐ **4**
- ☐ **5 – full understanding**

**Advance to next slide
for the poll**



Motivation for using TSD 20 methods

Potential uses and challenges for application in HTA submissions

Presented by: Miranda Cooper

TSD 20: A framework to simultaneously analyse correlated endpoints

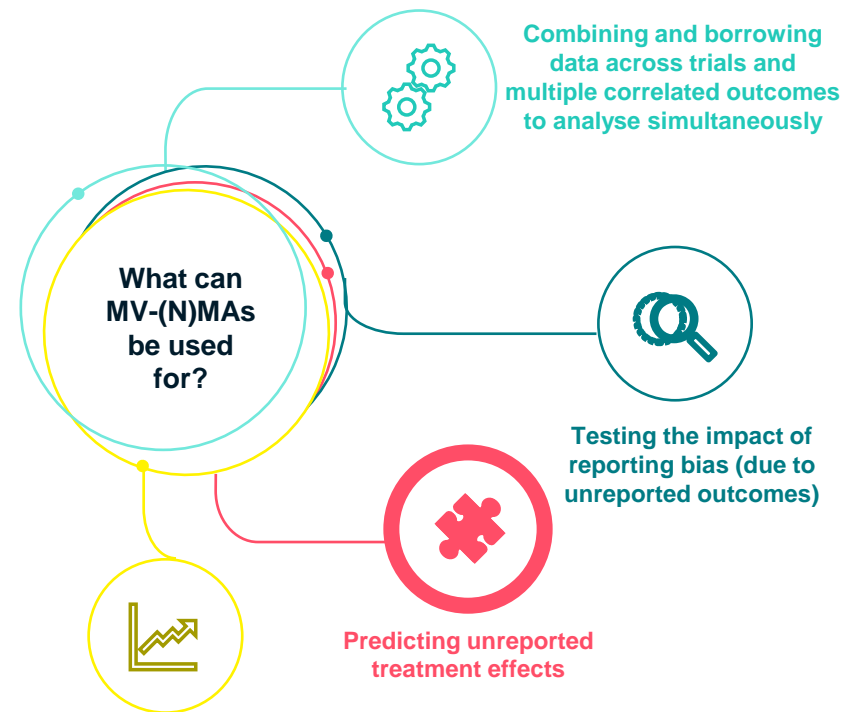
A process that allows cohesive analyses of correlated endpoints is a natural consideration

What?

- One of a series of TSDs from NICE
- Introduces a framework for conducting MV-(N)MAs
- Allows multiple correlated outcomes and/or multiple treatments to be analysed simultaneously
- Brings together surrogacy and (network) meta-analysis ideas
- Can account for within-study and between-study correlations between outcomes

Why?

- Allows more studies to contribute towards each outcome and treatment comparison, so all relevant evidence can be used
- Potential to reduce uncertainty, if assumptions of homogeneity and similarity of the underlying studies are reasonable
- Prevents clinical data from being discarded – reduced research waste
- Potential to reduce bias (due to selective outcome reporting)



Bivariate meta-analysis is a natural extension to a univariate case

A visual representation

Univariate case

- Treatment A versus Treatment B
- One outcome: e.g. PFS



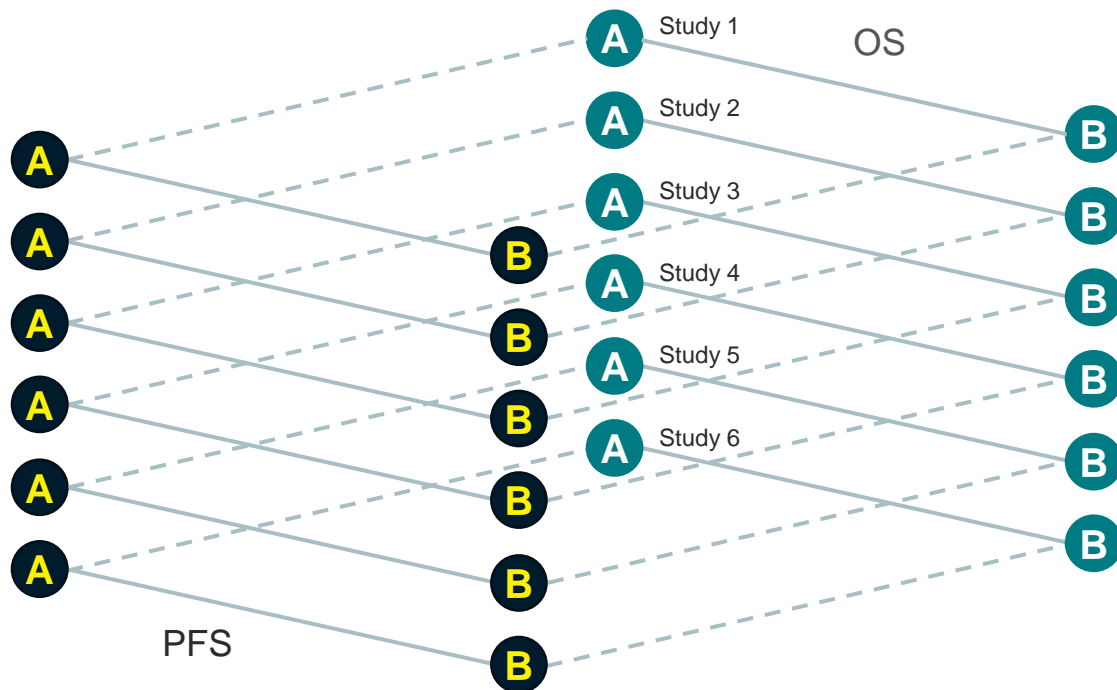
Bivariate meta-analysis is a natural extension to a univariate case

A visual representation

Bivariate meta-analysis is the simplest extension to univariate meta-analysis, where two outcomes are considered on multiple studies of A versus B.

Bivariate case

- Treatment A versus Treatment B
- Two outcomes: PFS and OS
- Outcomes assessed simultaneously



Another why: Trends for more immature data at HTA require new thinking

However, this is driving the need for higher standards

- The use of surrogate endpoints in clinical trials allows **earlier approval** of new drugs to treat serious diseases, such as cancer
 - 59% of treatments were approved by the FDA **before** significant OS data for the indication were published (between 1952 and 2016)¹
- Regulatory submission – and therefore reimbursement – is happening earlier in trial follow-up
 - Between 2016 and 2019, there was a **34.6% increase** in the proportion of HTAs to NICE, G-BA and HAS using immature OS²
 - This only looks set to increase



Reduced availability of data

Despite the data to support HTA becoming increasingly more limited³, expectations for information on long-term outcomes **remain consistent**. This leads to an increased focus on surrogacy and the possibility of predicting outcomes in an HTA setting.



Increased standards for surrogate outcomes

The recent NICE methods update clearly states the preference for validating surrogate outcomes using the methods outlined in TSD 20, which indicates more demanding standards for future surrogate use.



Another why: Trends for more immature data at HTA require new thinking

However, this is driving the need for higher standards

- The use of surrogate endpoints in clinical trials allows **earlier approval** of new drugs to treat serious diseases, such as cancer

- 59% of treatments were approved by the FDA **before** significant OS data for the indication were published (between 1952 and 2016)¹

- Regulatory submission – and therefore reimbursement – is happening earlier in trial follow-up

- Between 2016 and 2019, there was a **34.6% increase** in the proportion of HTAs to NICE, G-BA and HAS using immature OS²

- This only looks set to increase

Reduced availability of data


Despite the data to support HTA becoming increasingly more limited³, expectations for information on long-term outcomes **remain consistent**. This leads to an increased focus on surrogacy and the possibility of predicting outcomes in a setting.



As reliance on surrogates becomes greater, the standards for their use are increasing.

Increased standards for surrogate outcomes

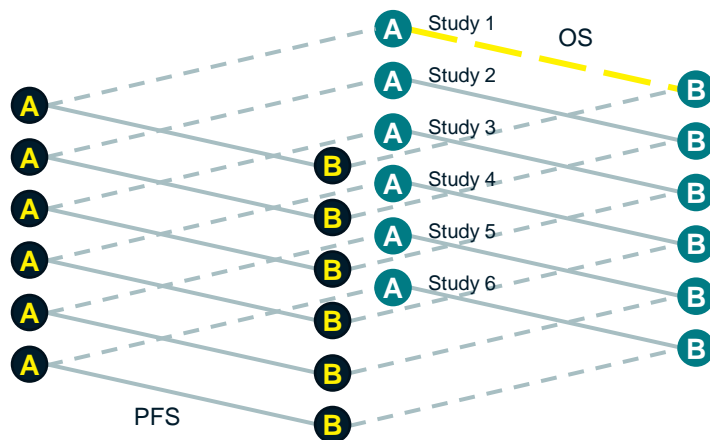
The recent NICE methods update clearly states the preference for validating surrogate outcomes using the methods outlined in TSD 20, which indicates more demanding standards for future surrogate use.



Why TSD 20 may be useful from an HTA standpoint

Trends for more immature data at HTA require new thinking

- Conducting HTA requires estimating the relative efficacy of a treatment versus a comparator
- Lack of mature data on key outcomes → **alternative routes** to estimating relative efficacy of this treatment on this key outcome
- In such scenarios, the methods introduced in TSD 20 can **provide a route to prediction** of outcomes, or **improve estimation and confidence** in the outcome using the available data on an appropriate surrogate endpoint



By using an outcome to predict or support estimation, we are inherently using this outcome as a surrogate.

Ideally this surrogate should first be **validated**.

Validating a surrogate outcome is challenging

For new treatments, it is unlikely that the available evidence base naturally lends itself to TSD 20

**Surrogate relationships
should only be validated for:**



When validating a surrogate for a new treatment:

- Sometimes there are only one or two treatments with data that are of the same class as the new therapy
- Sometimes there are no treatments of the same class – particularly in oncology!

As such, the most appropriate set of studies to form surrogacy analyses may not be large enough for the data requirements.

Validating a surrogate outcome is challenging

For new treatments, it is unlikely that the available evidence base naturally lends itself to TSD 20

**Surrogate relationships
should only be validated for:**



To resolve this issue, **expansions to the evidence base** need to be considered. This could be:

- Expansion to other therapy classes in the same indication
→ Considering a heterogeneous evidence base in terms of treatments and (potentially) study design
- Expansion to other similar indications for data on the target class of therapies
→ Considering a heterogeneous evidence base in terms of indications

The natural route when considering a broader range of treatments is to consider multivariate **network** meta-analyses.

However, the **data requirements for this become substantially more intensive**. The necessary simplification of the validation of a quantitative surrogate relationship in a multivariate **meta**-analysis can lead to interpretation issues...

Validating a surrogate outcome is challenging

For new treatments, it is unlikely that the available evidence base naturally lends itself to TSD 20

**Surrogate relationships
should only be validated for:**



To resolve this issue, **expansions to the evidence base** need to be considered. This could be:

The motivations for considering the methods of TSD 20 are likely to be in scenarios where TSD 20 methods are most challenging to apply.

The natural route when considering a broader range of treatments is to consider multivariate **network** meta-analyses.

However, the **data requirements for this become substantially more intensive**. The necessary simplification of the validation of a quantitative surrogate relationship in a multivariate **meta**-analysis can lead to interpretation issues...

The impact of heterogeneity in the evidence base on interpretations

Simplifying to a multivariate **meta-analysis** framework

If we simplify the validation to a **bivariate meta-analysis**, then:

- Control treatments are a heterogeneous mixture of different treatments and different treatment pathways
- Intervention treatments are a heterogeneous mixture of different treatments and different treatment pathways
- Intervention treatments in one study could be the control treatment in a different study

The relative treatment effect on the final outcome obtained from any multivariate meta-analysis will be difficult to interpret:

- Results would offer an estimate of the relative effect of **one heterogeneous mix of treatments** versus **an alternative heterogeneous mix of treatments**

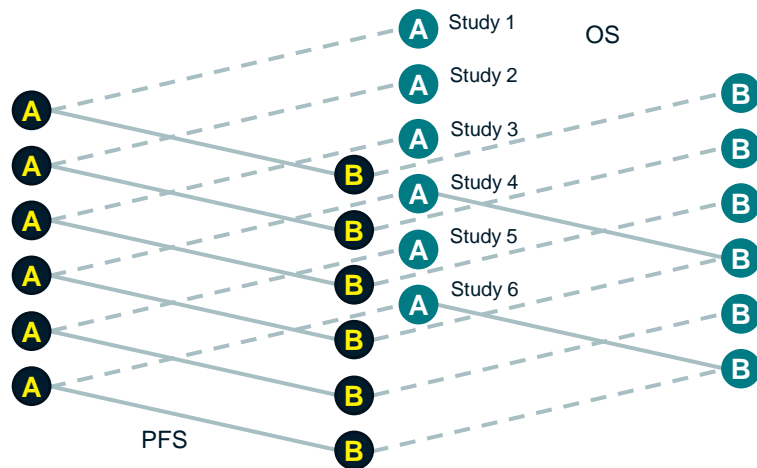
Applying this in the cost-effectiveness model will therefore depend on which treatment the estimated relative effect is applied to.

Data requirements when applying TSD 20 methods

The data requirements for analysis ideally require at least one study with PLD

- Several models are presented in TSD 20. They cover a range of complexities and scenarios dependent on the purpose of your analysis
- Regardless of method, we ideally need sufficient data to provide:

- 1) The association between the surrogate and final outcomes across studies
- 2) An informed estimate of the within-study correlation



- When validating a surrogate endpoint, the focus is on studies where both endpoints are reported, as these are the only studies that can provide evidence on the connection between outcomes
- In reality, this implies that **multiple important parameters may be predicted from very little data**, which leads to a risk of drawing strong conclusions based on minimal data
- This only becomes more difficult with more treatments and/or more outcomes

Data requirements when applying TSD 20 methods

The data requirements for analysis ideally require at least one study with PLD

- Several models are presented in TSD 20. They cover a range of complexities and scenarios dependent on the purpose of your analysis
- Regardless of method, we ideally need sufficient data to provide:
 - 1) The association between the surrogate and final outcomes across studies
 - 2) An informed estimate of the within-study correlation

PLD available from all studies

PLD available for one study (for example, a Phase III RCT of the new treatment)

PLD not available

- A rare occurrence
- The motivation for these methods in such scenarios is likely to be different (publication as opposed to prediction)
- The expected benefit should be considered alongside the expected effort

Data requirements when applying TSD 20 methods

The data requirements for analysis ideally require at least one study with PLD

- Several models are presented in TSD 20. They cover a range of complexities and scenarios dependent on the purpose of your analysis
- Regardless of method, we ideally need sufficient data to provide:
 - 1) The association between the surrogate and final outcomes across studies
 - 2) An informed estimate of the within-study correlation

PLD available from all studies

PLD available for one study (for example, a Phase III RCT of the new treatment)

PLD not available

- A common scenario
- However, it is likely that the only study with PLD available is the one that includes the new treatment. At which point, prediction is less likely to be appropriate
- Instead, TSD 20 methods could be used to improve the confidence in the relative treatment effect by borrowing from other studies to support the observed outcomes
- Consideration of the possible benefits versus effort required

Data requirements when applying TSD 20 methods

The data requirements for analysis ideally require at least one study with PLD

- Several models are presented in TSD 20. They cover a range of complexities and scenarios dependent on the purpose of your analysis
- Regardless of method, we ideally need sufficient data to provide:
 - 1) The association between the surrogate and final outcomes across studies
 - 2) An informed estimate of the within-study correlation

PLD available from all studies

PLD available for one study (for example, a Phase III RCT of the new treatment)

PLD not available

- An increasingly common scenario as the time between study initiation and regulatory/HTA proceedings continues to decrease
- Estimation of the within-study correlation between the two or more outcomes is challenging, proxies or other approaches need to be explored
- We should consider whether we can justify a surrogate relationship given the evidence available, i.e. can we answer the question with the evidence available?

Summary

- The most probable motivation for using TSD 20 is the situation that is likely to make applying these methods most challenging
- The standards for using surrogates are demanding, for good reasons. Better studies with longer follow-up that directly observe the final outcome are always preferable **from a statistical perspective** to predicting outcomes from a statistical model
- The challenges we face here are not solely because of TSD 20 methods, but can also be a product of **highly complex and heterogeneous evidence bases** (which affect our approach to standard ITCs)
- Data requirements are high
- Clinical interpretation of the pooled relative effects can be extremely limited depending on the assumptions made when pooling the data

It's Time for a Poll!

Which aspect of TSD20 are you the most interested in?

- **Multivariate MA for synthesis of treatment effectiveness data**
- **Bivariate MA for surrogate endpoint evaluation**
- **MV-NMA for both multiple outcomes and treatments**

**Advance to next slide
for the poll**



TSD20: methods for multivariate meta-analysis and surrogate endpoint evaluation

Sylwia Bujkiewicz

Professor of Biostatistics

Biostatistics Research Group

University of Leicester

9 November 2022

Content

- Brief introduction
- Random effects meta-analysis: from univariate to bivariate case
- Use of bivariate meta-analysis for surrogate endpoint evaluation
- Model by Daniels and Hughes
- Model comparison, pros and cons and example in multiple sclerosis

Multivariate meta-analysis

- Models jointly treatment effects on **multiple correlated outcomes**
- Advantages of multivariate meta-analysis
 - Potential for **reduced uncertainty**
 - Potential for **reduced outcome reporting bias**
 - Inclusion of **wider relevant evidence base** (efficient use of valuable trial data)
 - **Accounting for the correlation** more appropriate
 - Suitable for **validation of surrogate endpoints** and making **predictions**

Random-effects meta-analysis

In the presence of the between-study heterogeneity, a random effects approach:

- **treatment effects** Y_i (for example log HR on OS) are assumed to estimate **study-specific true treatment effects** δ_i
- different in each study i
- they follow a common distribution

$$Y_i \sim N(\delta_i, \sigma_i^2),$$
$$\delta_i \sim N(d, \tau^2)$$

τ^2 – the between-studies variance (τ - heterogeneity parameter)

Bayesian framework: unknown parameters are given prior distributions:

$$\tau \sim \text{unif}(0, 2)$$
$$d \sim N(0, 10^3).$$

Use of bivariate meta-analysis: pooled effects

Within-study variability:

$$Y_{1i} \sim N(\delta_{1i}, \sigma_{1i}^2),$$
$$Y_{2i} \sim N(\delta_{2i}, \sigma_{2i}^2)$$

Between-study variability:

$$\delta_{1i} \sim N(d_1, \tau_1^2),$$
$$\delta_{2i} \sim N(d_2, \tau_2^2)$$

Hierarchical framework:

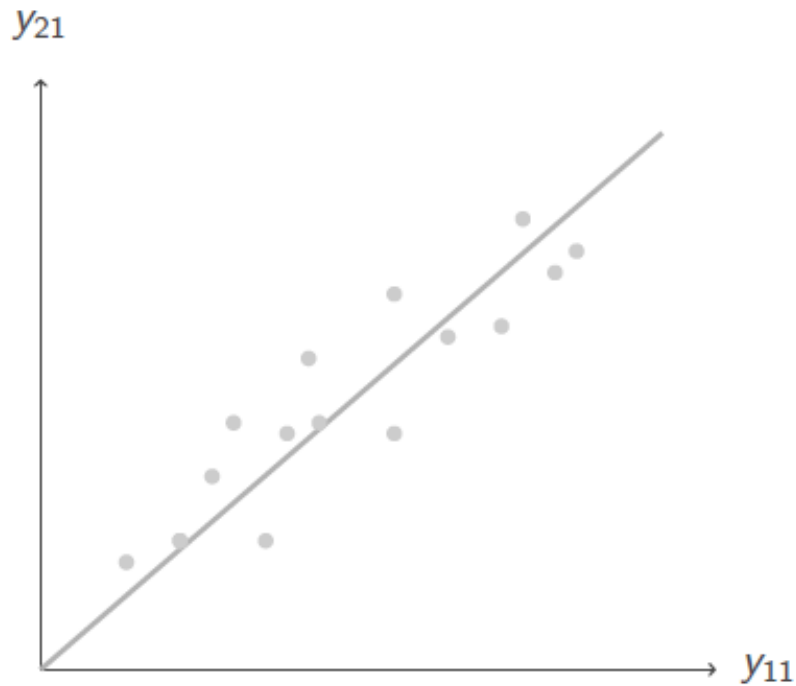
- Y_{1i}, Y_{2i} - estimates of correlated effects δ_{1i}, δ_{2i}
- $\sigma_{1i}^2, \sigma_{2i}^2$ - within-study variances
- δ_{1i}, δ_{2i} - true effects
- d_1, d_2 - pooled effect estimates
- τ_1^2, τ_2^2 - between-study variances.

Within-study variability

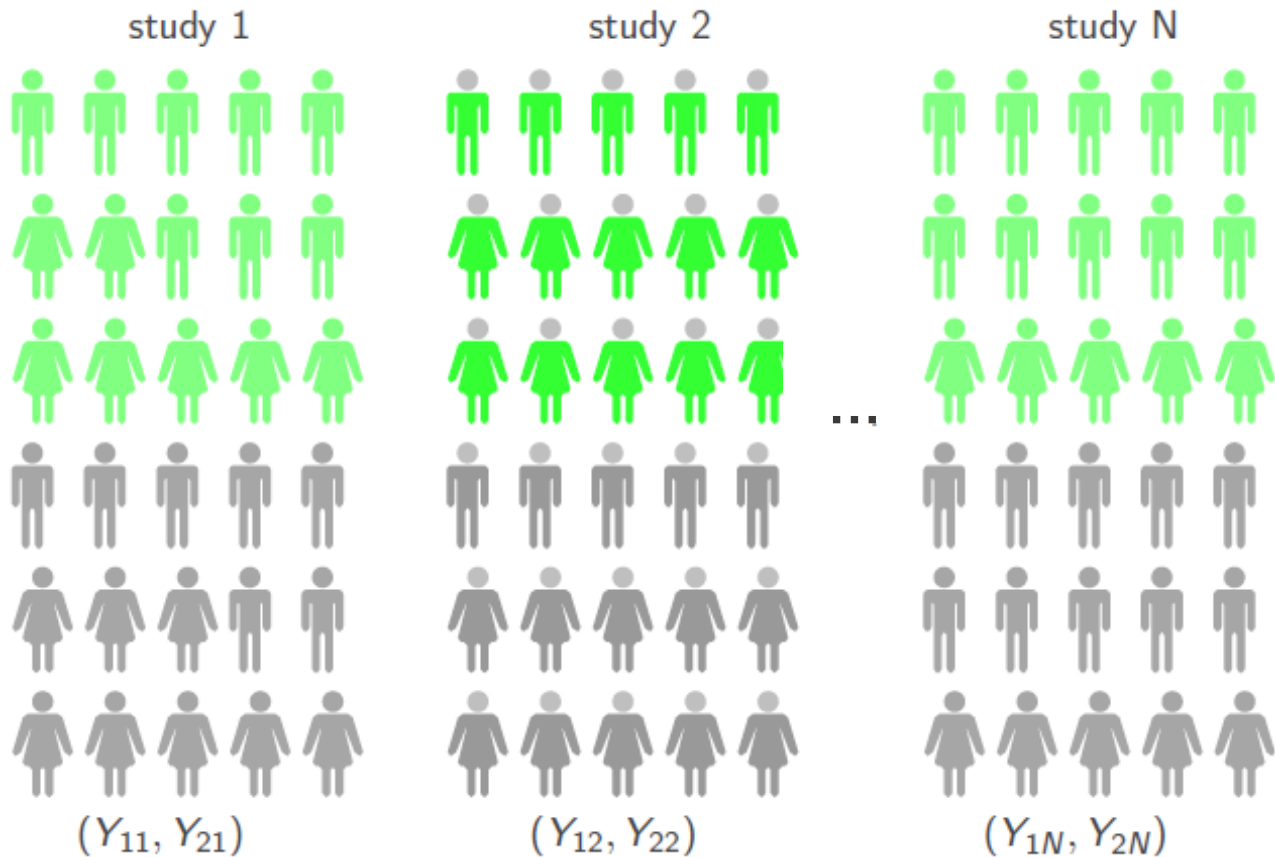


- Data on **two outcomes**, such as systolic blood pressure and diastolic blood pressure, are **collected from the same individuals**: all those randomised to two treatments.
- Patients may differ in their baseline characteristics leading to **variability** in their response to treatment

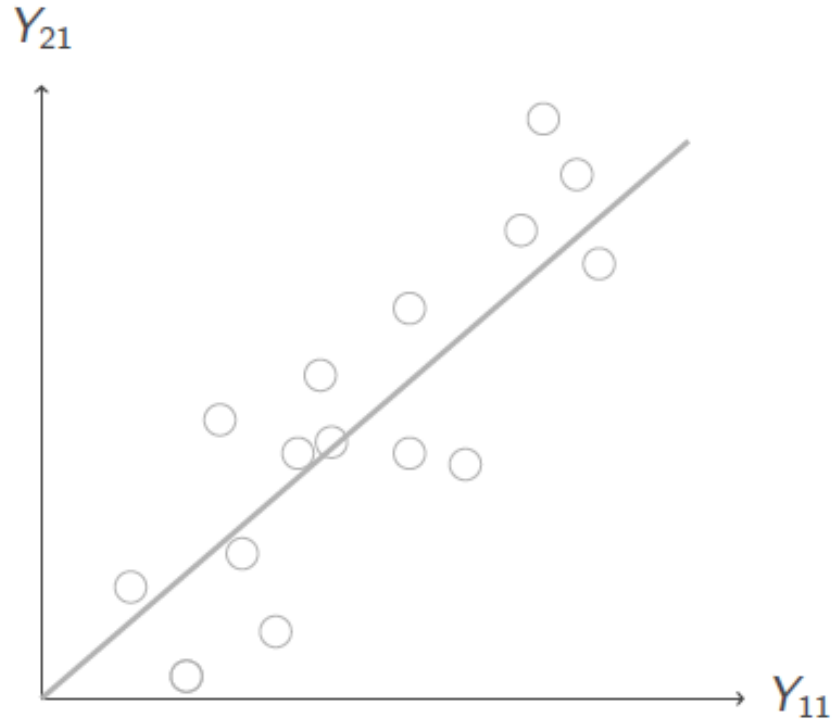
Within-study correlation



Between-study variability



Between-study variability



Summary data on two outcomes, collected from multiple studies.
Patient populations may differ leading to between-studies variability

Use of bivariate meta-analysis: pooled effects

Within-study variability:

$$Y_{1i} \sim N(\delta_{1i}, \sigma_{1i}^2),$$
$$Y_{2i} \sim N(\delta_{2i}, \sigma_{2i}^2)$$

Between-study variability:

$$\delta_{1i} \sim N(d_1, \tau_1^2),$$
$$\delta_{2i} \sim N(d_2, \tau_2^2)$$

Hierarchical framework:

- Y_{1i}, Y_{2i} - estimates of correlated effects δ_{1i}, δ_{2i}
- $\sigma_{1i}^2, \sigma_{2i}^2$ - within-study variances
- δ_{1i}, δ_{2i} - true effects
- d_1, d_2 - pooled effect estimates
- τ_1^2, τ_2^2 - between-study variances.

Bivariate random-effects meta-analysis (BRMA)

Within-study model:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi} \\ \sigma_{1i}\sigma_{2i}\rho_{wi} & \sigma_{2i}^2 \end{pmatrix} \right)$$

Between-study model:

$$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, T = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_b \\ \tau_1\tau_2\rho_b & \tau_2^2 \end{pmatrix} \right)$$

Hierarchical framework:

Y_{1i}, Y_{2i} - estimates of correlated true effects δ_{1i}, δ_{2i}

Σ_i - within-study covariance matrices of the estimates

ρ_{wi} - within-study correlations (assumed known)

d_1, d_2 - pooled effect estimates

T - between-study covariance matrix.

ρ_b - between-study correlation.

Bivariate random-effects meta-analysis (BRMA)

Within-study model:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi} \\ \sigma_{1i}\sigma_{2i}\rho_{wi} & \sigma_{2i}^2 \end{pmatrix} \right)$$

Between-study model:

$$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, T = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_b \\ \tau_1\tau_2\rho_b & \tau_2^2 \end{pmatrix} \right)$$

Bayesian framework: prior distributions placed on all unknown parameters

$$\tau_1, \tau_2 \sim \text{unif}(0, 2)$$

$$d_1, d_2 \sim N(0, 10^3)$$

$$\rho_b \sim \text{unif}(-1, 1)$$

Methods for surrogate endpoint evaluation

- Putative surrogate endpoints are validated by estimating the **pattern of association** between the treatment effects on surrogate and final endpoints across trials.
- A **meta-analytic approach**, based on data from a number of trials to establish the association between the treatment effects on the candidate surrogate endpoint and on the final outcome.
- **Bivariate meta-analysis methods**, that take account of the correlations between the average treatment effects on surrogate and final outcomes, are **suitable tools for modelling surrogate endpoints**

Bujkiewicz et al, NICE DSU Technical Support Document 20; 2019

Daniels and Hughes, Stat Med 1997,

Bujkiewicz et al. Stat Methods in Med Res 2018

BRMA in product normal formulation (PNF)

Within-study model:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi} \\ \sigma_{1i}\sigma_{2i}\rho_{wi} & \sigma_{2i}^2 \end{pmatrix} \right)$$

Between-study model:

$$\begin{aligned} \delta_{1i} &\sim N(d_1, \tau_1^2) \\ \delta_{2i} | \delta_{1i} &\sim N(\lambda_0 + \lambda_1 \delta_{1i}, \psi_2^2) \end{aligned}$$

Prior distributions: $\tau_1, \tau_2 \sim \text{unif}(0,2)$; $\lambda_0 \sim N(0, 10^3)$; $\rho_b \sim \text{unif}(-1,1)$

Implied priors: $\psi_2^2 = \tau_2^2(1 - \rho_b^2)$; $\lambda_1 = \rho\tau_2/\tau_1$

Random effects: assumption true effect follow a common distribution

Model by Daniels and Hughes

Within-study model:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi} \\ \sigma_{1i}\sigma_{2i}\rho_{wi} & \sigma_{2i}^2 \end{pmatrix} \right)$$

Between-study model:

$$\begin{aligned} \delta_{1i} &\sim N(0, 10^3) \\ \delta_{2i} | \delta_{1i} &\sim N(\lambda_0 + \lambda_1 \delta_{1i}, \psi_2^2) \end{aligned}$$

Criteria for **perfect surrogate relationship**:

$\lambda_1 \neq 0$ - for the association between the treatment effects

$\psi_2^2 = 0$ - for perfect association

$\lambda_0 = 0$ - no treatment effect on the surrogate endpoint gives no effect on the final outcome

Fixed effects: no assumption about the distribution of true effects on surrogate endpoint
– effects assumed independent

Daniels and Hughes, Statistics in Medicine 1997

Pros and cons

	Random effect bivariate meta-analysis	Daniels and Hughes fixed effects model
Pros	<ul style="list-style-type: none">• Better borrowing of information• Potentially more precise predictions	<ul style="list-style-type: none">• Better capture heterogeneity of the data• Avoids the risk of over-shrinkage when assumption of normality is not reasonable
Cons	<ul style="list-style-type: none">• Potential for over-shrinkage• May lead to surrogate deemed not valid when there is a good surrogate relationship	<ul style="list-style-type: none">• Missing the opportunity for better predictions when RE assumption valid• Cannot be used for estimation of pooled effects

Example: multiple sclerosis

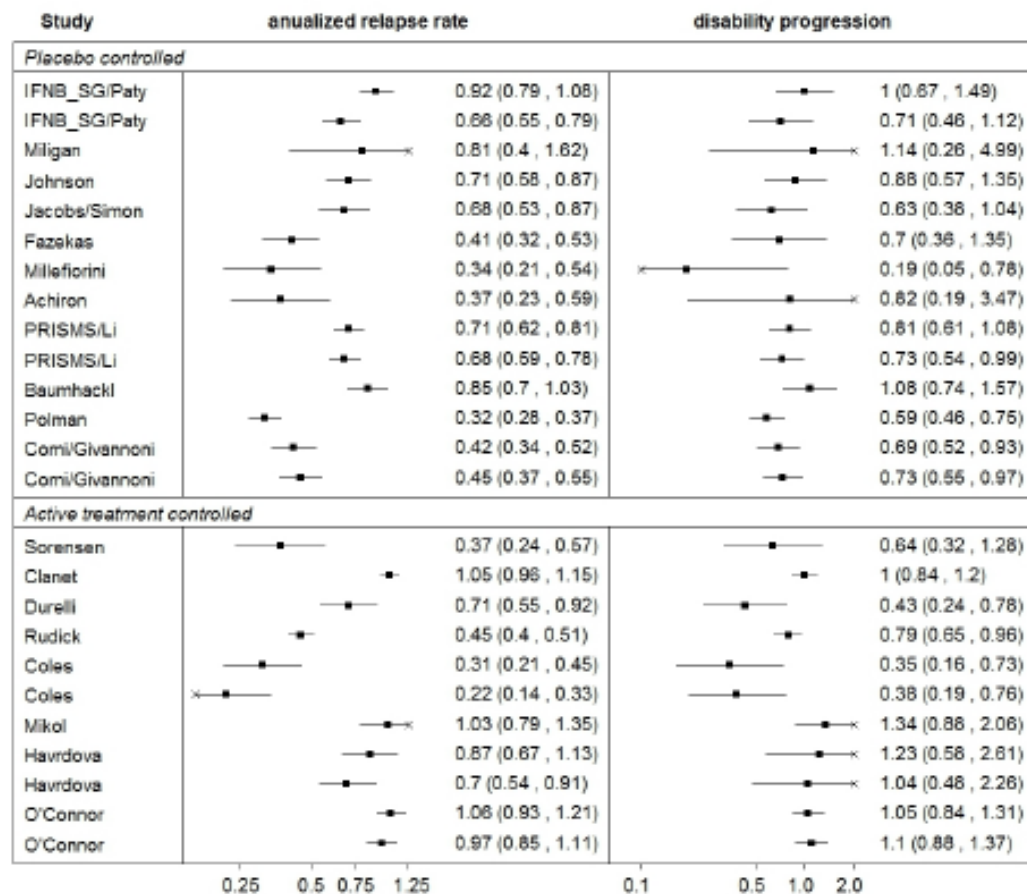
Disease area: multiple sclerosis

Final clinical outcome: Disability progression rate ratio, the ratio between the proportion of patients with a disability progression in the experimental and the control arms.

Putative surrogate endpoint: Annualized relapse rate ratio, the ratio between the relapse rate in the experimental and the control arms.

Sormani et al, Neurology 2010; 75: 302–309

Example: multiple sclerosis



Example: multiple sclerosis

	D&H	BRMA (PNF)	BRMA
d_1	NA	-0.50 (-0.68, -0.33)	-0.50 (-0.68, -0.33)
d_2	NA	-0.22 (-0.35, -0.11)	-0.22 (-0.34, -0.11)
ρ	NA	0.90 (0.63, 1.00)	0.90 (0.62, 1.00)
τ_1	NA	0.42 (0.30, 0.58)	0.41 (0.30, 0.57)
τ_2	NA	0.21 (0.11, 0.34)	0.20 (0.11, 0.33)
λ_0	0.03 (-0.07, 0.14)	0.001 (-0.12, 0.12)	-0.003 (-0.12, 0.11)
λ_1	0.52 (0.35, 0.70)	0.44 (0.24, 0.65)	0.43 (0.24, 0.64)
ψ_2^2	0.01 (0.00, 0.03)	0.01 (0.00, 0.03)	0.01 (0.00, 0.03)
ρ^2 (R^2)	NA	0.82 (0.40, 0.99)	0.81 (0.39, 0.99)
<i>NA – not applicable, D&H – model by Daniels & Hughes, BRMA – bivariate random-effects meta-analysis, PNF – product normal formulation</i>			

Predictions and cross-validation

The bivariate meta-analytic models can be used to make predictions

- For purpose of **cross-validation**:
 - in take-on-out cross-validation procedure
- To **predict treatment effect on the final clinical outcome in a new study**
 - from the treatment effect observed on the surrogate endpoint
 - conditional on the data from previous trials; used to develop the meta-analytic model

Cross-validation

	Disability progression risk ratio, mean (95% CrI)				
	Paty (1)	Paty (2)	Miligan	Johnson	Jacobs
Observed	1.00 (0.67, 1.49)	0.71 (0.45, 1.12)	1.14 (0.26, 5.03)	0.88 (0.57, 1.35)	0.63 (0.38, 1.05)
Daniels & Hughes	0.99 (0.63, 1.55)	0.84 (0.51, 1.37)	0.93 (0.20, 4.35)	0.86 (0.53, 1.40)	0.85 (0.49, 1.47)
BRMA (PNF)	0.95 (0.60, 1.50)	0.83 (0.51, 1.37)	0.86 (0.19, 3.93)	0.85 (0.52, 1.39)	0.85 (0.49, 1.47)
BRMA	0.95 (0.60, 1.51)	0.84 (0.51, 1.37)	0.86 (0.19, 3.93)	0.85 (0.52, 1.40)	0.85 (0.49, 1.47)
	Fazekas	Millefiorini	Achiron	Li (1)	Li (2)
Observed	0.70 (0.36, 1.35)	0.19 (0.05, 0.79)	0.82 (0.19, 3.50)	0.81 (0.61, 1.08)	0.73 (0.54, 0.99)
Daniels & Hughes	0.65 (0.32, 1.32)	0.60 (0.14, 2.59)	0.62 (0.14, 2.73)	0.87 (0.61, 1.22)	0.87 (0.59, 1.22)
BRMA (PNF)	0.68 (0.34, 1.39)	0.66 (0.15, 2.85)	0.68 (0.16, 3.01)	0.86 (0.60, 1.23)	0.85 (0.59, 1.22)
BRMA	0.68 (0.34, 1.40)	0.66 (0.15, 2.85)	0.68 (0.16, 3.01)	0.86 (0.60, 1.23)	0.85 (0.60, 1.22)
	Clanet	Durelli	Baumhackl	Polman	Rudick
Observed	1.00 (0.83, 1.20)	0.43 (0.24, 0.78)	1.08 (0.74, 1.57)	0.59 (0.46, 0.75)	0.79 (0.65, 0.96)
Daniels & Hughes	1.08 (0.81, 1.44)	0.88 (0.47, 1.65)*	0.94 (0.61, 1.46)	0.56 (0.39, 0.81)	0.66 (0.50, 0.87)
BRMA (PNF)	1.03 (0.76, 1.39)	0.86 (0.46, 1.61)*	0.91 (0.59, 1.42)	0.61 (0.42, 0.89)	0.68 (0.51, 0.91)
BRMA	1.02 (0.76, 1.40)	0.86 (0.46, 1.62)*	0.91 (0.59, 1.42)	0.61 (0.42, 0.89)	0.68 (0.51, 0.91)
	Coles (1)	Coles (2)	Mikol	Comi (1)	Comi (2)
Observed	0.35 (0.16, 0.74)	0.38 (0.19, 0.77)	1.34 (0.88, 2.06)	0.69 (0.52, 0.93)	0.73 (0.55, 0.97)
Daniels & Hughes	0.58 (0.26, 1.30)	0.48 (0.22, 1.04)	1.03 (0.63, 1.67)	0.65 (0.44, 0.95)	0.67 (0.47, 0.97)
BRMA (PNF)	0.64 (0.29, 1.43)	0.57 (0.27, 1.23)	0.98 (0.59, 1.58)	0.68 (0.47, 1.01)	0.70 (0.48, 1.02)
BRMA	0.64 (0.29, 1.43)	0.57 (0.27, 1.23)	0.97 (0.59, 1.58)	0.68 (0.47, 1.01)	0.70 (0.48, 1.02)
	Havrdova (1)	Havrdova (2)	Sorensen	O'Connor (1)	O'Connor (2)
Observed	1.23 (0.58, 2.62)	1.04 (0.48, 2.27)	0.64 (0.32, 1.28)	1.05 (0.84, 1.31)	1.10 (0.88, 1.37)
Daniels & Hughes	0.96 (0.44, 2.11)	0.86 (0.38, 1.92)	0.62 (0.29, 1.32)	1.06 (0.78, 1.47)	1.00 (0.74, 1.36)
BRMA (PNF)	0.92 (0.42, 2.03)	0.85 (0.38, 1.91)	0.68 (0.32, 1.43)	1.02 (0.73, 1.42)	0.96 (0.70, 1.31)
BRMA	0.92 (0.42, 2.03)	0.85 (0.38, 1.91)	0.68 (0.32, 1.43)	1.02 (0.73, 1.42)	0.95 (0.70, 1.31)

Discussion

- Use of putative surrogate endpoint – careful **evaluation** is needed.
- Rather than using pre-specified criteria, considering **a balance** between the **strength of the surrogate relationship** and the need for the decision to be made about the effectiveness of the new treatment may be important, in particular in the **high priority disease areas**.
- **Uncertainty** is key. The **strength of the surrogate relationship** will be reflected in the level of **uncertainty** around the predicted treatment effect on the final outcome.
- **Predicted** treatment effect (along with the uncertainty) may be used in HTA decision making framework.
- Novel methods may help improve predictions.

References (and further reading)

- M. J. Daniels and M. D. Hughes. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine*, 16(17):1965-1982, 1997.
- Bujkiewicz S, Achana FA, Papanikos T, Riley RD, Abrams KR, NICE DSU **Technical Support Document 20**: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints; 22 Oct 2019.
- Bujkiewicz S, Thompson JR, Spata E, Abrams KR, Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints, *Statistical Methods in Medical Research*, 2017; 26 (5): 2287–2318.
- Bujkiewicz S, Jackson D, Thompson JR, Turner R, Städler N, Abrams KR, White IR, *Bivariate network meta-analysis for surrogate endpoint evaluation, Statistics in Medicine* 2019;38:3322–3341.
- Papanikos T, Thompson JT, Abrams KR, Städler N, Ciani O, Taylor RS, Bujkiewicz S, *A Bayesian hierarchical meta-analytic method for modelling surrogate relationships that vary across treatment classes, Statistics in Medicine* 2020;39:1103–1124.

It's Time for a Poll!

For which purpose would you use the methods for surrogate endpoints?

- **Validation of a surrogate relationship**
- **Prediction of the clinical outcome**
- **Precision improvement of the clinical outcome estimate**

**Advance to next slide
for the poll**

FRAMEWORK FOR HANDLING SURROGATE OUTCOMES WITH REFERENCE TO CASE STUDIES

TECHNICAL SUPPORT DOCUMENT (TSD) 20 IN ACTION – WHAT TO CONSIDER WHEN USING MULTIVARIATE META- OR
NETWORK META-ANALYSES (NMA) TO SUPPORT HTA ONCOLOGY SUBMISSIONS

7th November 2022
ISPOR Europe

Oriana Ciani, PhD

TSD 20 IN ACTION: A CASE-STUDY

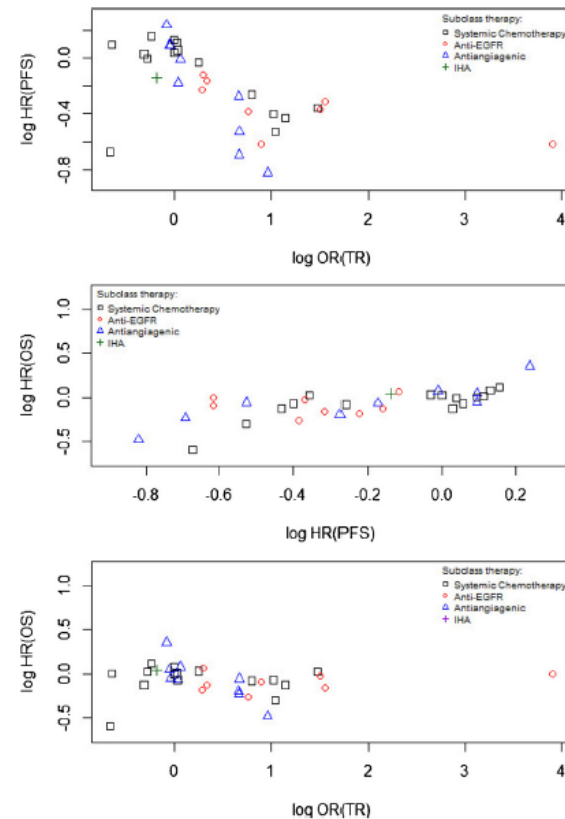
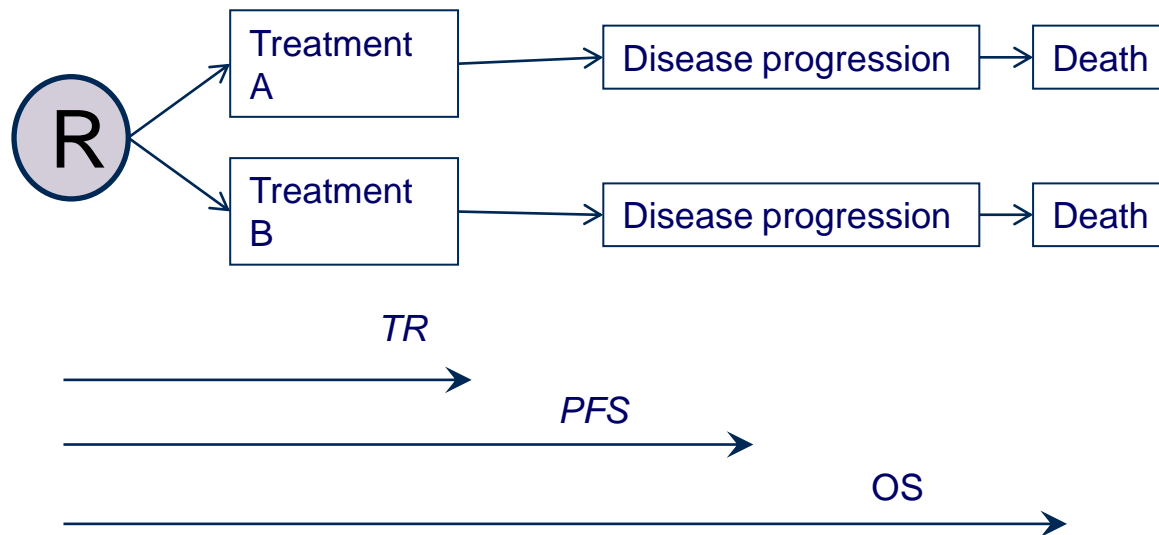


Fig. 1. Scatter plot of treatment effect estimates on PFS vs TR (top), OS vs PFS (middle) and OS vs TR (bottom); for each subclass of therapy for the main data set including trials reporting all three outcomes.

Table 1

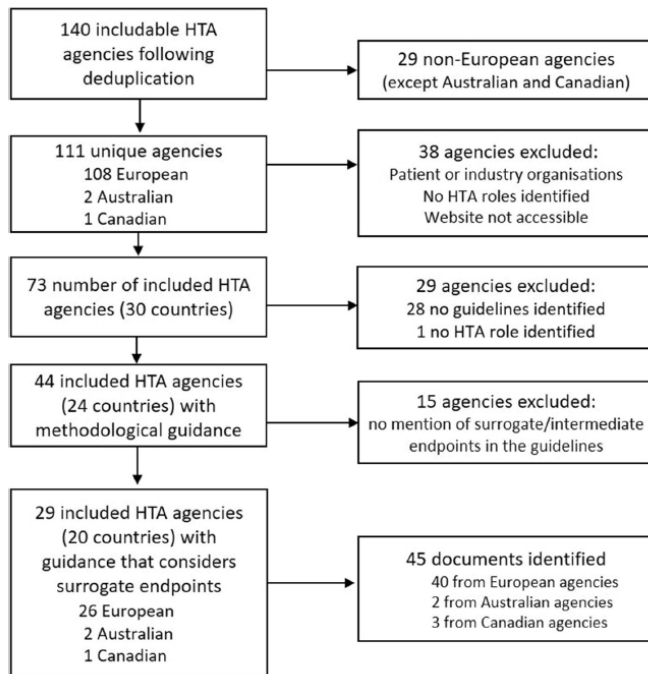
Surrogacy criteria obtained from three bivariate models and a trivariate model for the association between treatment effects on the surrogate (TR or PFS) and the final outcome (OS or PFS in one of the bivariate analyses). The results are posterior means and 95% credible intervals.

	Bivariate analyses			Trivariate analysis	
	TR OS	TR PFS	PFS OS	TR PFS	PFS OS
<i>All treatments (N = 33)</i>					
Intercept	-0.03(-0.07, 0.02)	-0.05(-0.14, 0.02)	-0.02(-0.06, 0.03)	-0.05(-0.13, 0.02)	-0.02(-0.06, 0.03)
Slope	-0.05(-0.13, 0)	-0.32(-0.45, -0.2)	0.22(0.03, 0.41)	-0.31(-0.43, -0.19)	0.19(0.02, 0.4)
Variance	0(0, 0.01)	0.02(0.01, 0.05)	0(0, 0.01)	0.02(0.01, 0.04)	0(0, 0.01)
$R^2_{adjusted}$	0.33(0, 0.91)	0.61(0.27, 0.87)	0.58(0.06, 0.97)	0.64(0.3, 0.89)	0.5(0.02, 0.95)
<i>Systemic chemotherapy (N = 15)</i>					
Intercept	-0.02(-0.08, 0.04)	-0.04(-0.16, 0.06)	-0.02(-0.08, 0.04)	-0.04(-0.14, 0.05)	-0.02(-0.08, 0.04)
Slope	-0.03(-0.11, 0)	-0.26(-0.42, -0.08)	0.17(0, 0.45)	-0.25(-0.4, -0.09)	0.14(0, 0.4)
Variance	0(0, 0.01)	0.02(0, 0.08)	0(0, 0.01)	0.02(0, 0.06)	0(0, 0.01)
$R^2_{adjusted}$	0.39(0, 0.96)	0.58(0.07, 0.96)	0.52(0.01, 0.98)	0.66(0.11, 0.98)	0.47(0, 0.97)
<i>Anti-EGFR therapies (N = 8)</i>					
Intercept	-0.06(-0.16, 0.09)	-0.21(-0.37, 0.01)	-0.06(-0.16, 0.14)	-0.18(-0.37, 0.05)	-0.04(-0.16, 0.19)
Slope	-0.04(-0.18, 0)	-0.14(-0.36, -0.01)	0.14(0, 0.63)	-0.16(-0.39, -0.02)	0.17(0, 0.78)
Variance	0.01(0, 0.05)	0.02(0, 0.08)	0.01(0, 0.04)	0.02(0, 0.1)	0.01(0, 0.05)
$R^2_{adjusted}$	0.19(0, 0.82)	0.45(0, 0.96)	0.22(0, 0.85)	0.5(0.01, 0.97)	0.19(0, 0.83)
<i>Anti-angiogenic agents (N = 9)</i>					
Intercept	0.04(-0.09, 0.2)	0.03(-0.18, 0.25)	0.02(-0.1, 0.15)	0.03(-0.18, 0.25)	0.02(-0.09, 0.14)
Slope	-0.35(-0.89, -0.03)	-0.87(-1.64, -0.3)	0.38(0.05, 0.79)	-0.85(-1.64, -0.27)	0.37(0.04, 0.77)
Variance	0.02(0, 0.07)	0.04(0, 0.15)	0.02(0, 0.06)	0.03(0, 0.14)	0.01(0, 0.06)
$R^2_{adjusted}$	0.52(0.01, 0.97)	0.74(0.13, 0.99)	0.59(0.03, 0.97)	0.72(0.1, 0.99)	0.56(0.02, 0.97)

TR – tumour response, PFS – progression free survival, OS – overall survival.

WHAT DO HTA METHODS GUIDELINES CURRENTLY RECOMMEND?

Fig. 1 Summary of agencies and documents selection. HTA health technology assessment

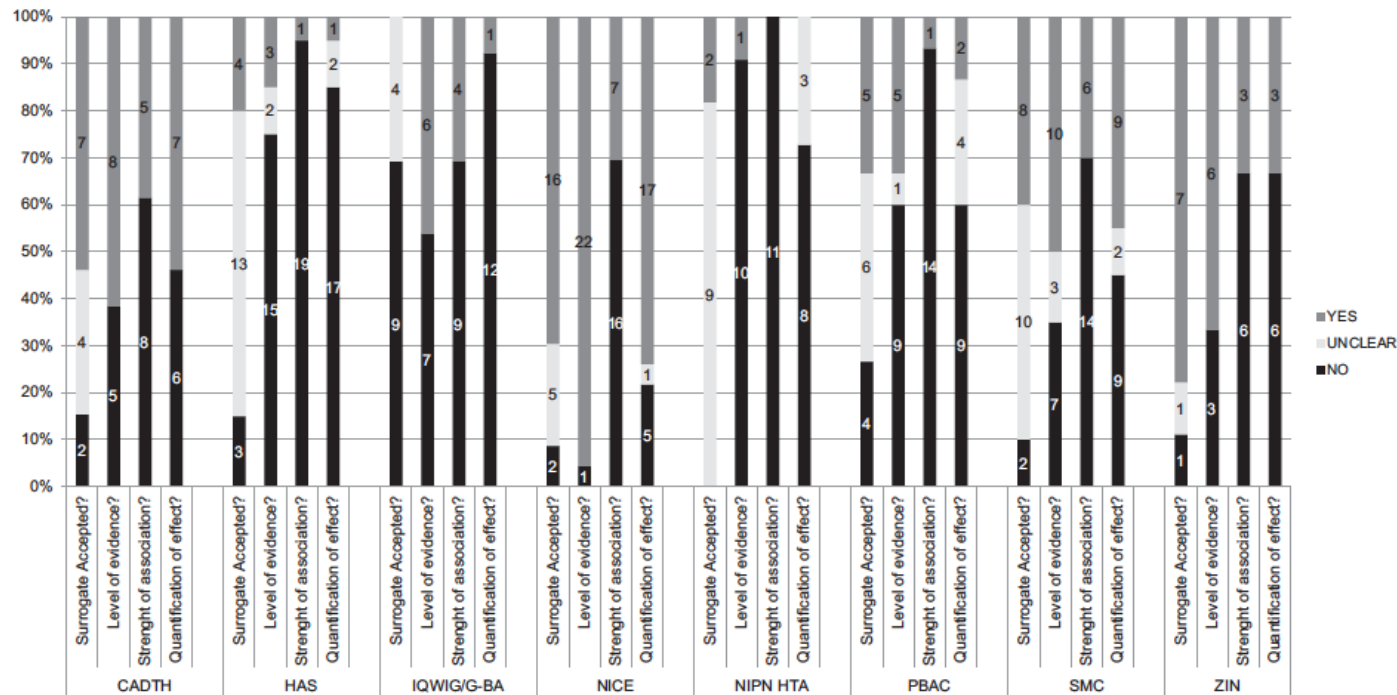


44 (98%)	Argument around use of surrogates in the analysis
18 (40%)	Provide specific examples
13 (29%)	Give a definition for surrogate endpoint
10 (22%)	Report more detailed methods for the handling of surrogate endpoints
2 (4%)	Refer to thresholds for validation
3 (7%)	Specific guidance for disease areas

Table 1 Methods for the Validation of Surrogate Endpoints: 3-Stage Framework

Level of Evidence		Strength of the Association	Quantification of the Expected Effect on the Patient-Centered Outcome
Level 1: Randomized controlled trials showing that treatment changes in the surrogate are associated with treatment changes in the final outcome	<p><i>For trial-level surrogacy</i></p> <p>Meta-analysis of individual patient data/aggregate data from randomized controlled trials that have assessed both the surrogate and patient-centered endpoints</p> <p>With trial/country/center as the analysis unit</p> <p>Preferably within the same indication and treatment class</p>	<p><i>For trial-level surrogacy</i></p> <p>Coefficient of correlation (Kendall's τ, Spearman's ρ, Pearson within-study correlations from multivariate meta-analyses)</p> <p>Coefficient of determination from weighted/unweighted adjusted/unadjusted linear regression of treatment effects on endpoints/copula models</p>	<p><i>For trial-level surrogacy</i></p> <p>Prediction based on the estimated regression equation for the trial-level surrogacy and observed effect on the surrogate endpoint</p> <p>Intercept, slope, and conditional variance of the linear model of the relationship between the treatment effects on the surrogate endpoint and the effects on the final outcome based on aggregate data</p> <p>Bayesian multivariate meta-analyses</p> <p>Surrogate threshold effect, the minimum treatment effect on the surrogate necessary to predict a nonzero effect on the patient-centered outcomes using the 95% prediction limits of the regression line</p>
Level 2: Epidemiological/observational studies showing consistent association between surrogate and final outcome	<p><i>For individual-level surrogacy</i></p> <p>As above or even single large randomized controlled trials/observational studies that have assessed both the surrogate and patient-centered endpoints</p>	<p><i>For individual-level surrogacy</i></p> <p>Coefficient of correlation (Kendall's τ, Spearman's ρ, Pearson)</p> <p>Coefficient of determination from weighted/unweighted adjusted/unadjusted linear regression of treatment effects on endpoints/copula models</p>	
Level 3: Pathophysiological studies and understanding of the disease process demonstrating the biological plausibility of relation between surrogate and final outcome		<p>Hazard ratio from Cox regressions/Bayesian hierarchical analysis</p>	

HOW IS VALIDATION OF SURROGATE ENDPOINTS EMPIRICALLY ADDRESSED IN HTA REPORTS?



HOW IS VALIDATION OF SURROGATE ENDPOINTS EMPIRICALLY ADDRESSED IN HTA

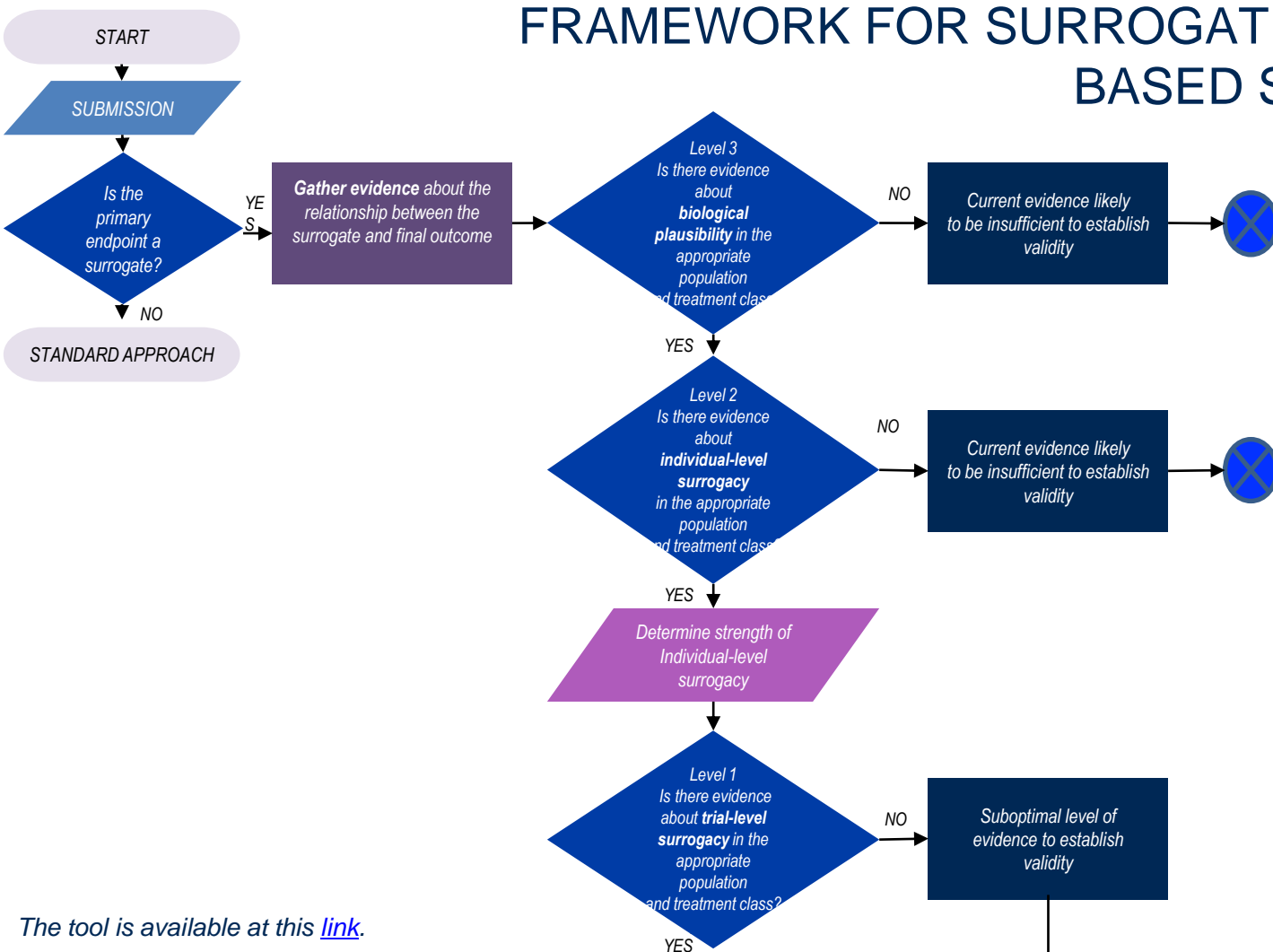
Characteristics	Total number of HTA reports (N = 124)	
Drugs	122	98%
Medical device	2	2%
HTA Agencies		
NICE	23	19%
HIS/ SMC	20	16%
HAS	20	16%
PBAC/ MSAC	15	12%
CADTH	13	10%
IQWiG / G-BA	13	10%
ZiN	9	7%
NIPN	11	9%
Disease area		
Cancer	65	52%
Cardiovascular	17	14%
Pulmonology	8	6%
Nephrology	8	6%
Endocrinology	7	6%
Infectious Disease	7	6%
Ophthalmology	6	5%
Gastroenterology	6	5%
Orphan status	8	6%
Surrogate validation		
Surrogate accepted (YES)	49	40%
Level of evidence assessed (YES)	61	49%
Strength of association provided (YES)	27	22%
Quantification of effect provided (YES)	40	32%
Final recommendation given		
Approved	32	26%
Restricted	61	49%
Rejected	20	16%
No recommendation	11	9%

- The different level of scrutiny applied across agencies translates into different declared level of acceptability for the same surrogate endpoint, in mostly the same indication, and based on what is theoretically the same evidence available to each appraisal committee. Overall, the level of agreement across the agencies is 0.10 ($p = 0.04$)
- In terms of recommendation given, 61(49%) had a restricted approval (e.g. PAS, risk-sharing). Overall, the level of agreement across the agencies is 0.18 ($p = 0.004$). Positive recommendations not formally associated with acceptable surrogate endpoints.

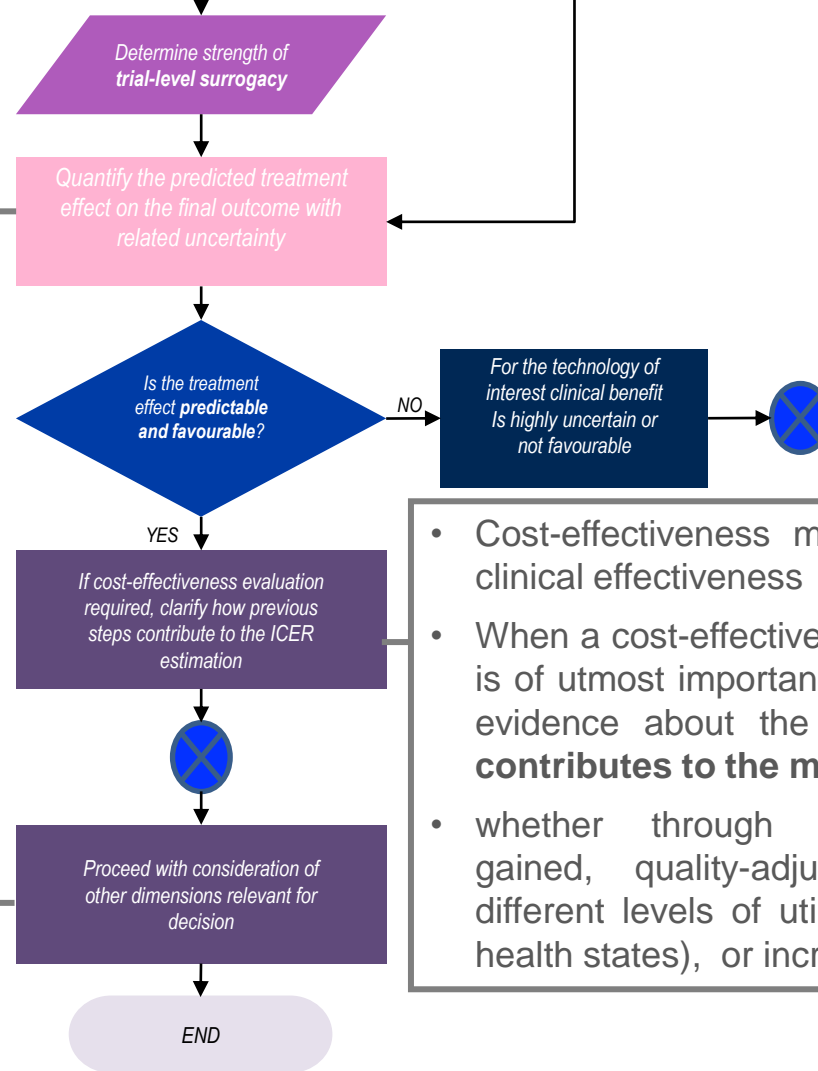
Factors associated with positive recommendation	Multivariate regression analysis*
Acceptability of surrogate endpoint	0.71 (0.23 - 2.20) [$p = 0.55$]
Level of evidence assessed	0.32 (0.07 - 1.37) [$p = 0.12$]
Strength of association provided	2.30 (0.51 - 10.45) [$p = 0.28$]
Quantification of effect provided	1.12 (0.27 - 4.74) [$p = 0.87$]
Orphan status	8.61 (1.03 - 72.94) [$p = 0.047$]

*from mixed-effect logistic regression with clustering at the level of the health technology. OR>1 indicates higher odds of technology receiving positive recommendation

FRAMEWORK FOR SURROGATE OUTCOMES BASED SUBMISSIONS



- How depends on the type of surrogate endpoint and available observational data
- ✓ survival curves estimated separately for responders and non-responders (with or without a landmark time)
- ✓ use of prognostic model research, prediction scores or risk equation...
- Indirectness -> additional decision uncertainty!



- Cost-effectiveness models cannot prescind from clinical effectiveness
- When a cost-effectiveness evaluation is required, it is of utmost importance to be **transparent** on how evidence about the primary surrogate endpoint **contributes to the model**
- whether through estimation of life years gained, quality-adjusted life years gained (e.g., different levels of utility value to use for different health states), or incremental costs.

- Contextual factors may be acknowledged (e.g., rarity, end of life, unmet need, innovative mechanism) in formulating the final recommendations



Università
Bocconi

CERGAS

Center for Research on Health
and Social Care Management

SDA Bocconi
SCHOOL OF MANAGEMENT

THANK YOU

oriana.ciani@unibocconi.it

@OrianaCiani

Thank you! Any questions?

