

# Borrowing strength from RWE in study-level surrogate endpoint evaluation

*Georgios Nikolaidis, PhD*

*Acknowledgements to: Anastasios Tasoulas, Vagia Daki,  
Julia Oddsdottir, Sylwia Bujkiewicz*

# Outline

1. Introduction to surrogate endpoints. Concept and surrogacy criteria
2. Case-study Research Questions
  - a) Clinical objective
  - b) Methodological objective
3. Leveraging comparative RWE: methods and results
4. Conclusions and key take-aways



Journal of Clinical Epidemiology 187 (2025) 111933

**Journal of  
Clinical  
Epidemiology**

## ORIGINAL RESEARCH

Including comparative observational evidence in trial-level surrogate end point evaluation: assessing relapse-free survival as a surrogate end point for overall survival in patients with acute myeloid leukemia posttransplant

Georgios F. Nikolaidis<sup>a,\*</sup>, Anastasios Tasoulas<sup>a</sup>, Georgia Gourgioti<sup>b</sup>, Marco Groß-Langenhoff<sup>c</sup>, Melinda Hamilton<sup>b</sup>, Christian J.A. Ridley<sup>b</sup>, Sylwia Bujkiewicz<sup>d</sup>

<sup>a</sup>*IQVIA Ltd, Methods and Evidence Generation, Centre of Excellence in Evidence Synthesis, London, UK*

<sup>b</sup>*Astellas Pharma Europe Ltd, Addlestone, UK*

<sup>c</sup>*Astellas Pharma GmbH, Munich, Germany*

<sup>d</sup>*Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK*

Accepted 12 August 2025; Published online 16 August 2025



Abbreviations: AML, Acute myeloid leukaemia; HSCT, Haematopoietic stem cell transplantation; RCT, Randomised clinical trials; RFS, Relapse-free survival; RWE, Real-world evidence; OS, Overall survival.

# Reliable surrogate endpoints can assist & accelerate drug approval in slow progressing diseases

## Motivation:

- Drug efficacy, especially in oncology, typically relies on evidence of **long-term survival** (e.g., OS, PFS). Such outcomes take long to read out, often more than the duration of the average clinical trial.
- In cases of **high unmet need**, regulatory bodies may accept evidence of efficacy for **outcomes that read out earlier**, if those can be reliably assumed to correlate with improvement on the longer-term outcomes.

## Example:



- In this case, we would claim that **tumor response** is a **surrogate** endpoint for **overall survival**, and such evidence could even support accelerated drug approval, as relevant FDA guidance<sup>[1]</sup> states.

[1]: FDA (2022), Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure [\[link\]](#)

# HTA is considered more with trial-level surrogacy rather than patient-level surrogacy

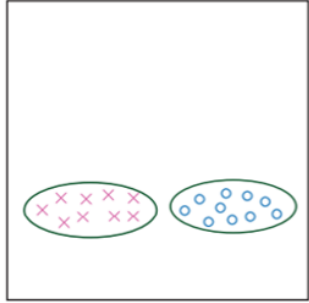
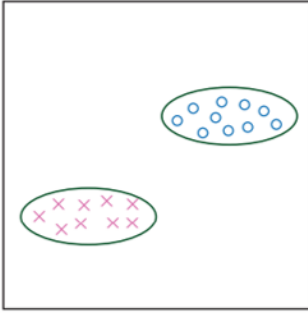

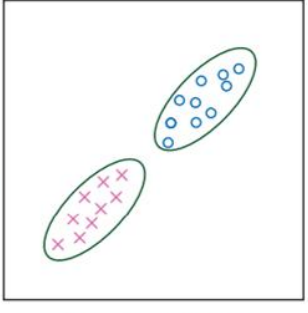
## Types of surrogacy:

- **Patient-level:** Patients with tumor response live longer. Can be established with individual patient data (IPD) from a single trial.
- **Trial-level:** Populations with **higher tumor response rate**, show **better overall survival** consistently across trials. Needs a meta-analysis of several trials.

While patient-level surrogacy is useful, regulators are most interested in **trial-level** surrogacy.

To this end, the UK's **National Institute for Health and Care Excellence (NICE)** has issued relevant guidance on surrogate endpoint evaluation in TSD-20<sup>[1]</sup>, using the Bayesian meta-analytic modelling framework.

[1]: Bujkiewicz, S., et al., NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. 2022. [\[link\]](#)

Surrogacy		Trial-level	
		x	✓
Patient-level	x	Definitive endpoint Y  Candidate surrogate Z	Definitive endpoint Y  Candidate surrogate Z
	✓	Definitive endpoint Y  Candidate surrogate Z	Definitive endpoint Y  Candidate surrogate Z

# What makes a good study-level surrogate outcome?

- Daniels & Hughes (1997)<sup>[1]</sup> introduced 3 surrogacy criteria based on  $\lambda_0$ ,  $\lambda_1$  and  $\psi_2^2$

## Surrogacy criterion

1

Intercept  $\lambda_0$  should be 'near zero'  
This ensures that lack of relative treatment effect on the surrogate outcome implies **lack** of relative treatment effect on the final outcome.

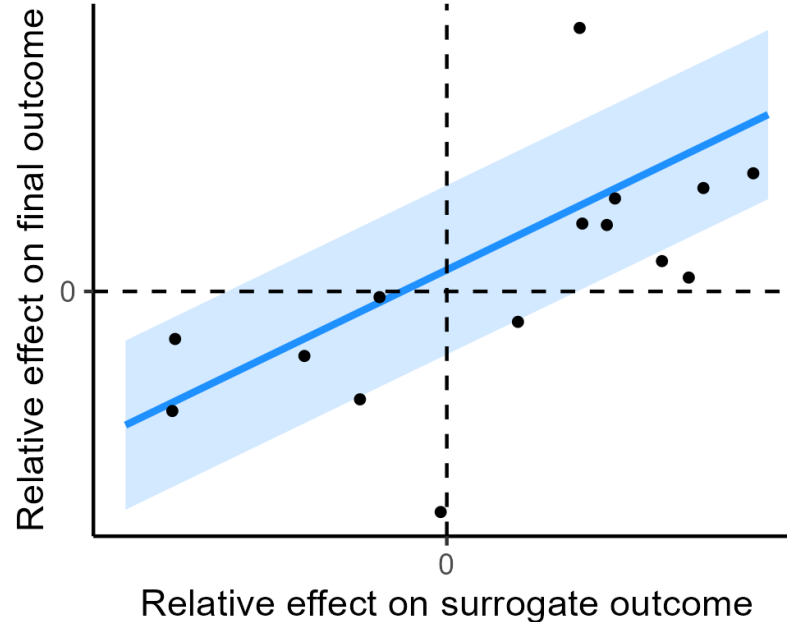
2

Slope  $\lambda_1$  should be 'away from zero'  
This ensures that relative treatment effect on the surrogate outcome translates to a **measurable** relative treatment effect on the final outcome.

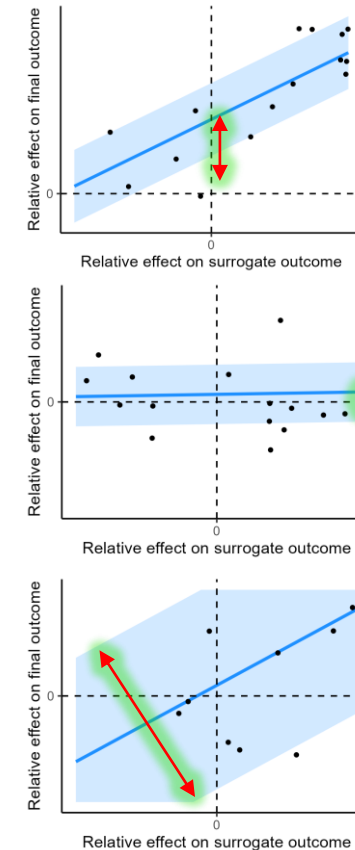
3

Conditional variance  $\psi_2^2$  should be 'small'  
This ensures that **given** the relative treatment effect on the surrogate, the relative treatment effect on the final outcome can be predicted with **precision**.

## A **good** surrogate



## A **bad** surrogate



## Why?

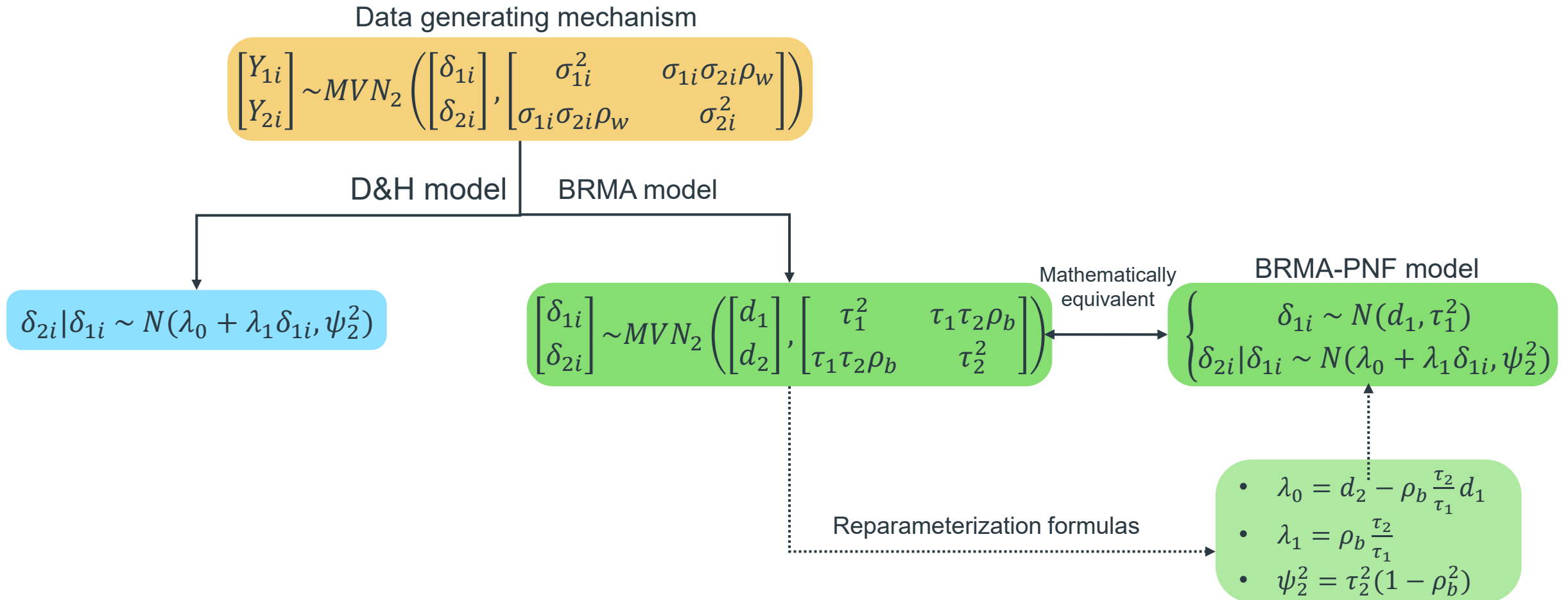
Large final outcome effect even when surrogate effect is zero

No impact on final outcome effect as surrogate effect changes

Too much uncertainty!

[1]: Daniels, M. J., & Hughes, M. D. (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in medicine*, 16(17), 1965-1982. [\[link\]](#)

# Surrogacy evaluation relies on multi-variate meta-analysis



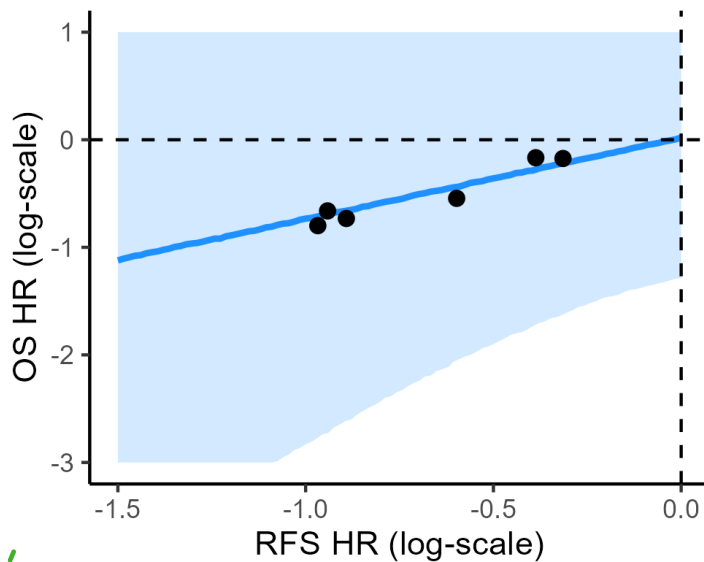
Abbreviations: BRMA: bivariate random-effects meta-analytic; D&H: Daniel & Hughes; PNF: product-Normal formulation.

# Clinical Research question



To assess whether Relapse-free survival (**RFS**) is a good study-level surrogate endpoint for Overall survival (**OS**) in the post-Haematopoietic Stem Cell Transplantation (**post-HSCT**) setting in Acute Myeloid Leukemia (**AML**)

Using only RCT data (N=6) results in high uncertainty and does *not* meet the surrogacy criteria. (But the direction is encouraging)



Model	Parameter	Median	95% CrI Lower	95% CrI Upper
D&H	$\lambda_0$	0.08	-0.74	0.79
	$\lambda_1$	0.85	-0.30	1.84
	$\psi_2^2$	0.01	0.00	0.53



*Expanding the evidence criteria to include comparative observational evidence may reduce uncertainty*

# Methodological Research question: Can RWE complement trial evidence?

- **Risks and benefits** of information-sharing (Benefits: gain precision and reduce uncertainty, Risks: bias, complexity, borrow weakness)
- There is a **misconception** in HTA that indirect evidence have to be completely discarded (*splitting*) or taken at face value (*lumping*)
- However, **several methods** imposing intermediate degrees of information-sharing have been proposed in the biostatistics literature
- These ‘sophisticated’ information-sharing methods have been used to help decision-making in children, in trial-design, and in rare diseases



Nikolaidis et al. *BMC Medical Research Methodology* (2021) 21:107  
<https://doi.org/10.1186/s12874-021-01292-z>

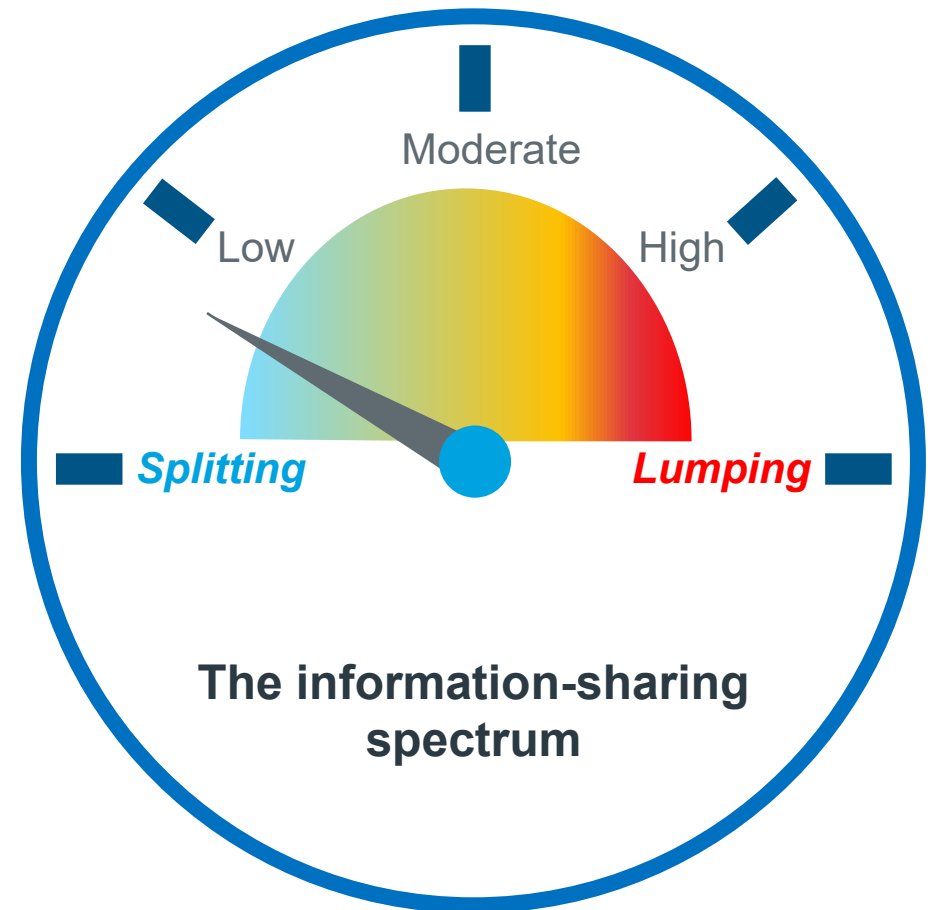
BMC Medical Research  
Methodology

RESEARCH ARTICLE

Open Access

## Classifying information-sharing methods

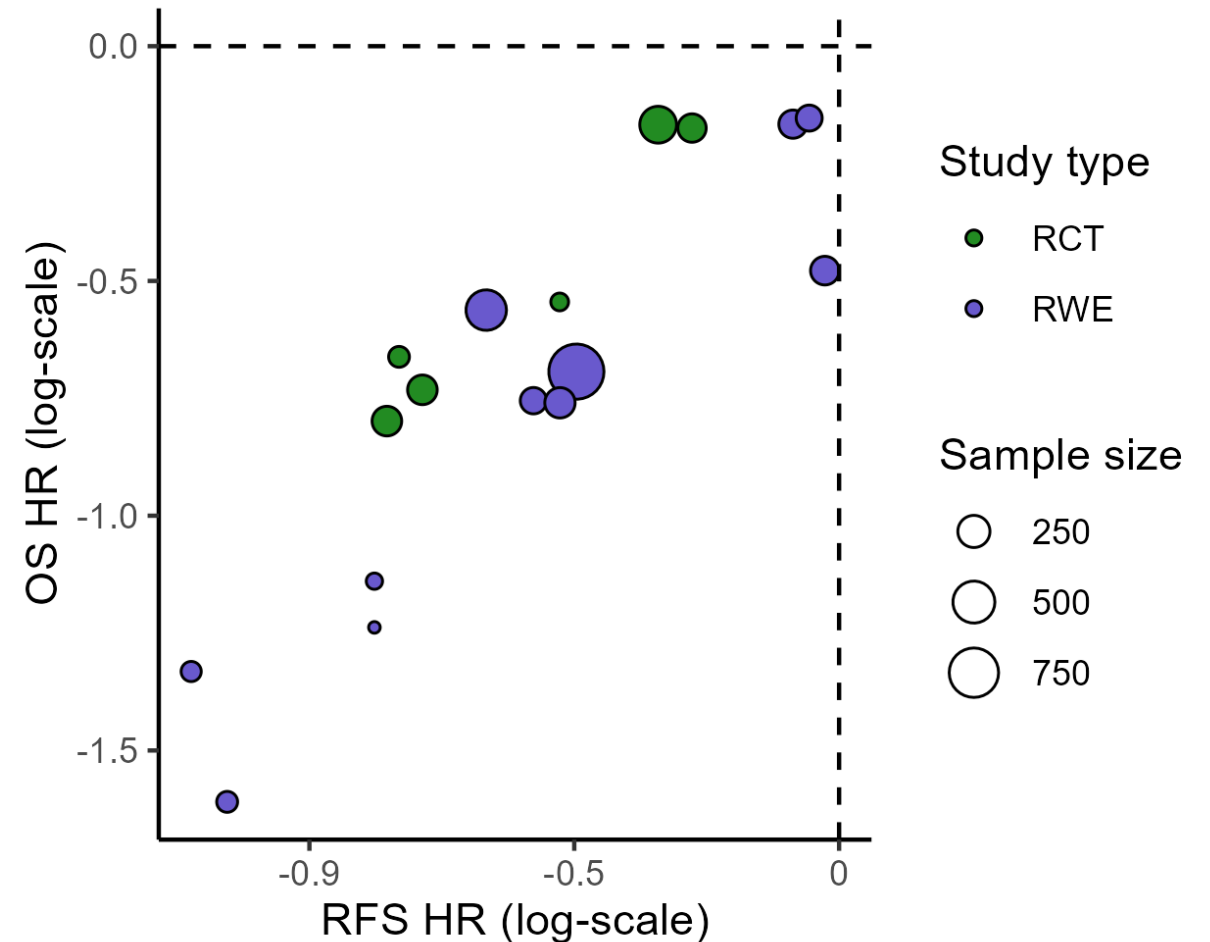
Georgios F. Nikolaidis<sup>1,2\*</sup>, Beth Woods<sup>1</sup>, Stephen Palmer<sup>1</sup> and Marta O. Soares<sup>1</sup>



Abbreviations: HTA, Health technology assessment.

# Expanded Evidence base including RCT & RWE

- Overall, **N=6 RCTs** and **N=11 RWE** studies reflecting a total of **1084** and **2466** patients, respectively
- *Feasibility assessment*
  - There was **some heterogeneity** across studies in terms of baseline characteristics, classes of assessed treatments, posology of treatments, and some outcome definitions
  - Clinical experts were interviewed to evaluate the expected level of bias and potential direction.
  - There was no pattern in the observed relative effects in terms of study type or size
- Overall, **the combination of RCT and RWE was considered plausible** and worth exploring in addition to RCT-only analyses



# The power prior enables user-specified discounting of RWE

## Power prior model

### How it works

- RCT and RWE analyzed together, but the influence of RWE data points is regulated based on a power ( $0 \leq \alpha \leq 1$ ), specified by the user
- $\alpha = 0$  implies RWE are fully discounted, and have zero influence
- $\alpha = 1$  implies RWE have the same influence as RCT evidence

### Pros

- Extent of RWE discounting is **user-specified**
- Discounting allowed to differ across RWE studies

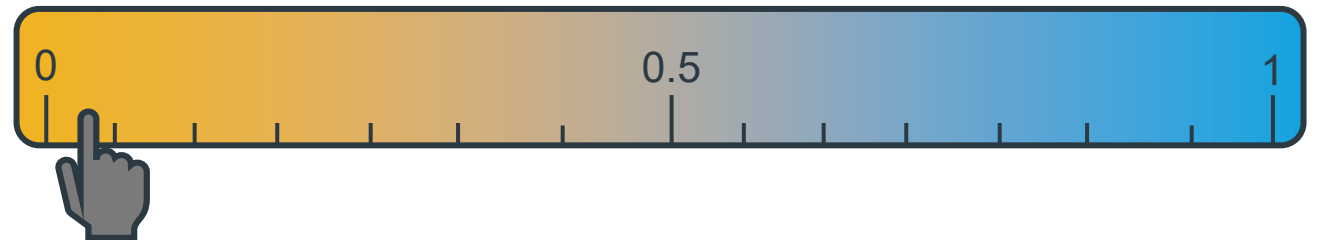
### Cons

- Not very intuitive **interpretation**
- Choice of  $\alpha$  can be arbitrary and hard to elicit

Let  $D = \{D_{RCT}, D_{RWE}\}$  be a dataset comprised of RCT and RWE evidence,  $\alpha \in [0, 1]$  be a scalar, and  $\theta = [\lambda_0, \lambda_1, \psi_2^2]^T$  be the vector of surrogacy parameters with prior  $p(\theta)$  and likelihood  $L(\theta|D)$ . The posterior distribution of  $\theta$  will be **modified** as:

$$p(\theta|D, \alpha) \propto L(\theta|D_{RCT}) \underbrace{L(\theta|D_{RWE})^\alpha}_{\propto p(\theta|D_{RWE}, \alpha)} p(\theta)$$

Power prior weight  $\alpha$



# The mixture prior model lets the data 'decide' how much RWE evidence should be discounted

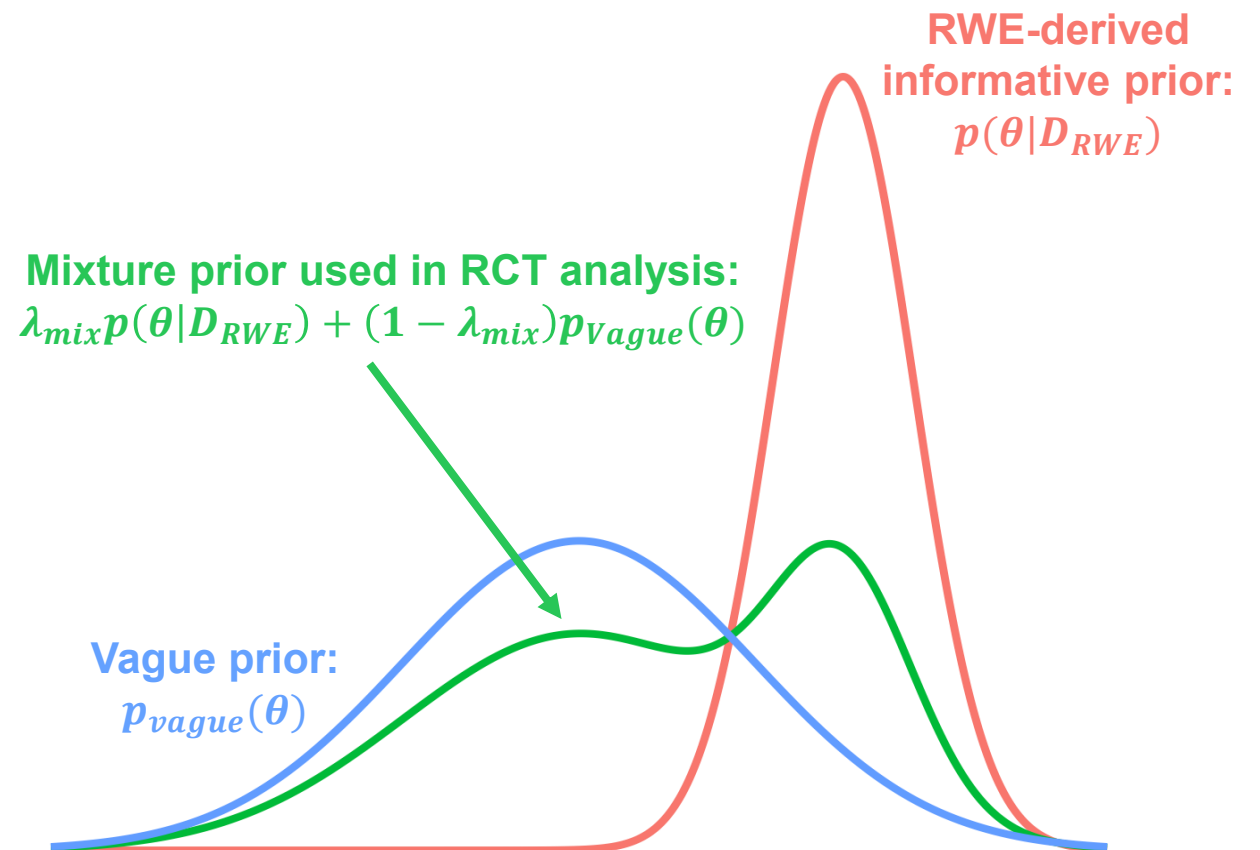
## Mixture prior model

### How it works

- A **three-step** approach
  - ❑ **Step 1:** Analyze **RWE** and obtain posterior estimates of the surrogacy parameters
  - ❑ **Step 2:** Treat RWE posterior as an **informative prior**, and combine it with a **vague prior** to create a **mixture prior**
  - ❑ **Step 3:** Analyse **RCTs alone**, using the **mixture prior**
- Mixing parameter  $\lambda_{mix}$  determines the level of information borrowing, and can be kept fixed or estimated by the model.

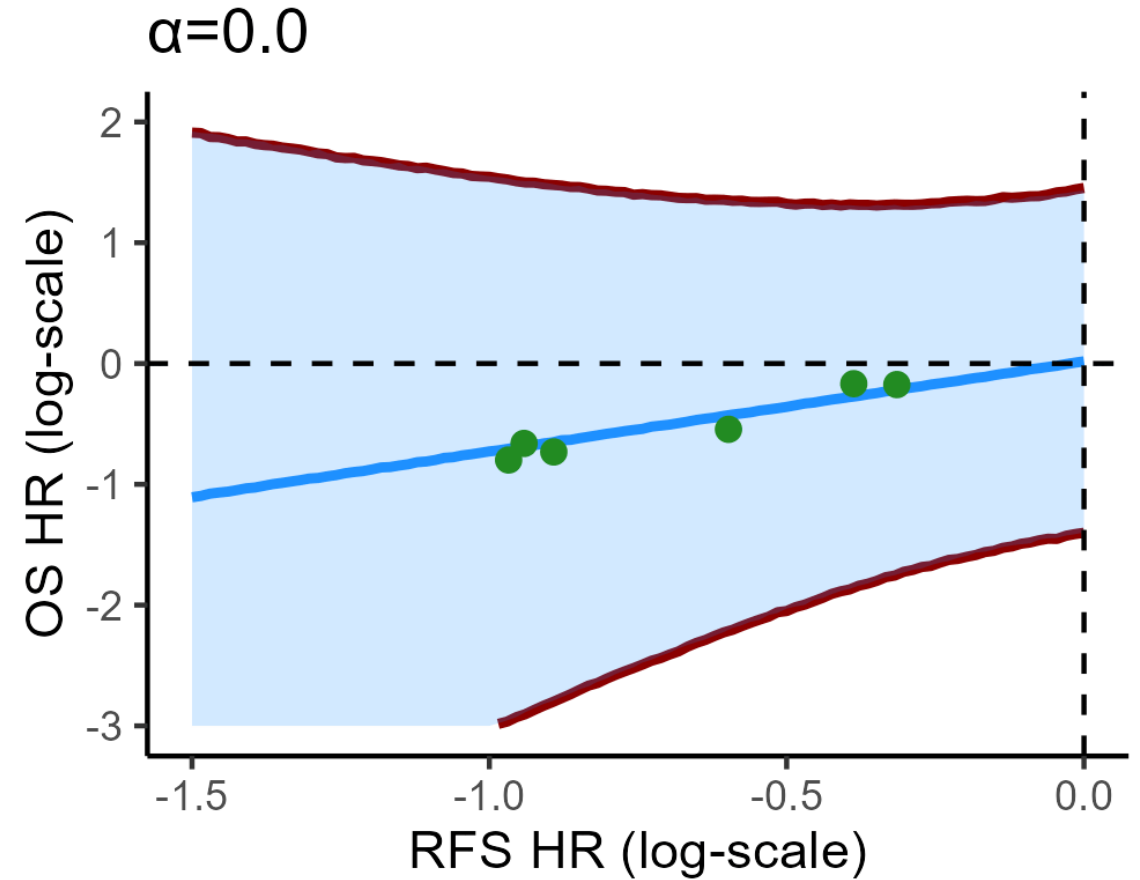
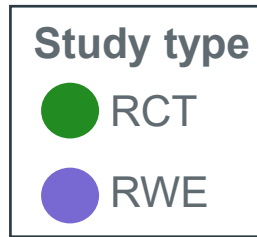
### Pros

- Can be **adaptive** i.e., allowing the 'data to speak' and discourage information-sharing in case of prior data conflict
- Intuitive interpretation



# Results: Incorporating RWE in the D&H model using power priors

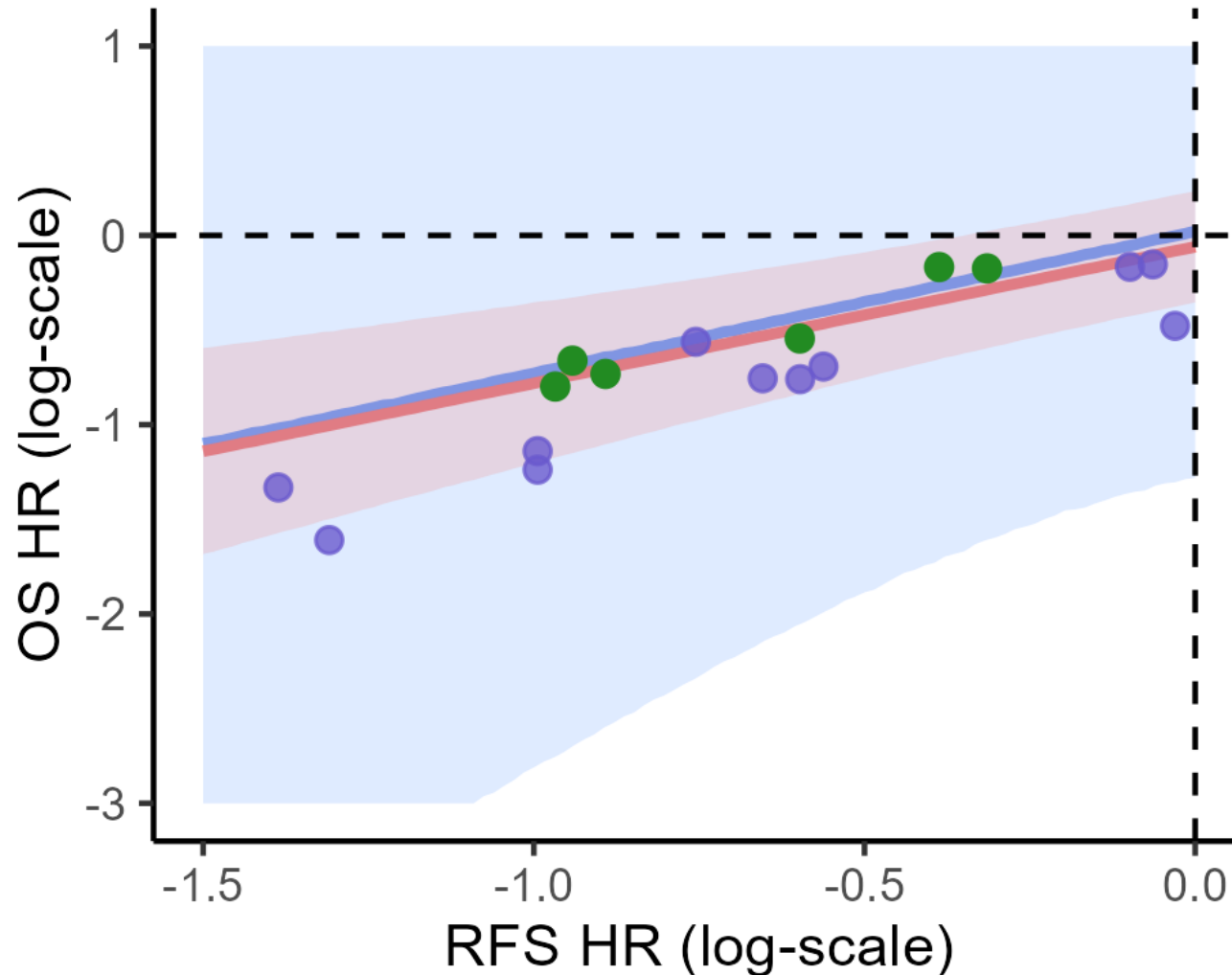
- $\alpha=0$  implies complete discarding of RWE, whilst  $\alpha=1$  implies full inclusion of RWE at face value
- All RWE studies were assigned a common
- Increasing the contribution of RWE **reduces uncertainty** with minor impact on the point estimates of surrogacy parameters
- **Surrogacy is supported** even when RWE is partially used ( $\alpha=0.5$ )
- **High information gains** (i.e., reduction in uncertainty) are quickly realized even starting from very low values of alpha



Analysis	$\alpha = 0.1$			$\alpha = 0.3$			$\alpha = 0.5$			$\alpha = 0.7$			$\alpha = 0.9$		
Parameter	Median	95% CrI LB	95% CrI UB	Median	95% CrI LB	95% CrI UB	Median	95% CrI LB	95% CrI UB	Median	95% CrI LB	95% CrI UB	Median	95% CrI LB	95% CrI UB
$\lambda_0$	-0.10	-0.65	0.26	-0.10	-0.50	0.20	-0.10	-0.44	0.16	-0.10	-0.40	0.14	-0.10	-0.35	0.12
$\lambda_1$	0.64	-0.10	1.11	0.71	0.14	1.12	0.76	0.27	1.13	0.79	0.36	1.13	0.82	0.44	1.14
$\psi_2^2$	0.03	0.00	0.27	0.02	0.00	0.15	0.02	0.00	0.12	0.02	0.00	0.10	0.02	0.00	0.09

Abbreviations: CrI, Credible interval; D&H, Daniel & Hughes model; LB, Lower bound; RCT, Randomised clinical trials; RFS, Relapse-free survival; RWE, Real-world evidence ;OS, Overall survival, UB, Upper bound.

# Results: Incorporating RWE in the D&H model using mixture priors



Only RCT (i.e., No RWE)			
Parameter	Median	95% CrI Lower	95% CrI Upper
$\lambda_0$	0.08	-0.74	0.79
$\lambda_1$	0.85	-0.30	1.84
$\psi_2^2$	0.01	0.00	0.53

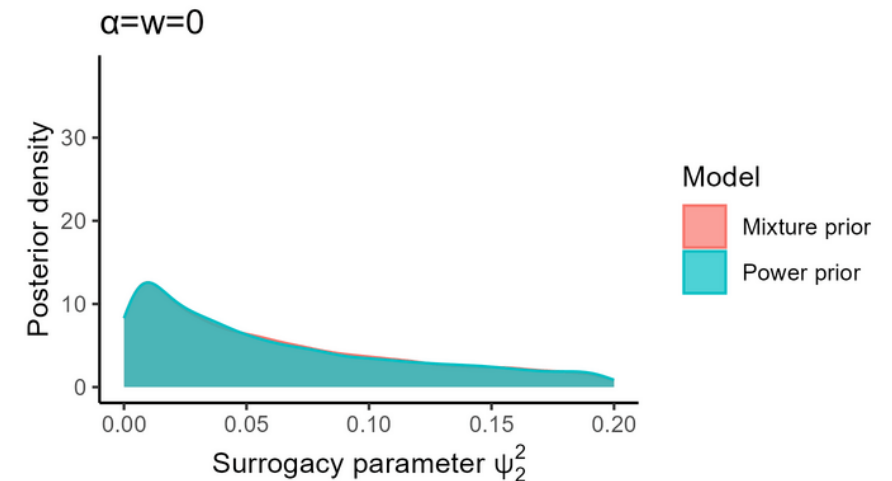
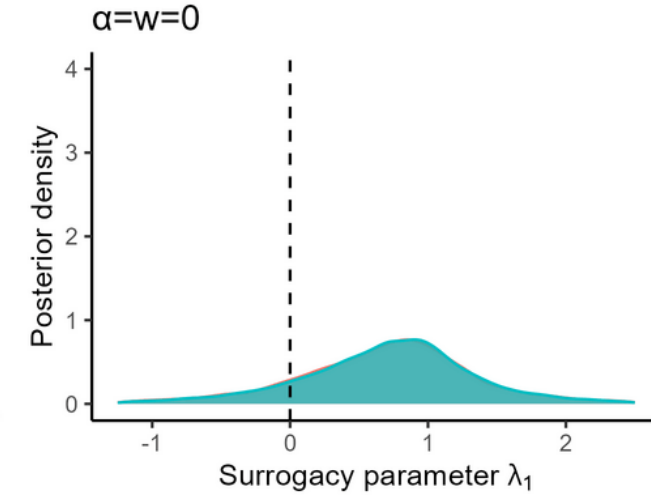
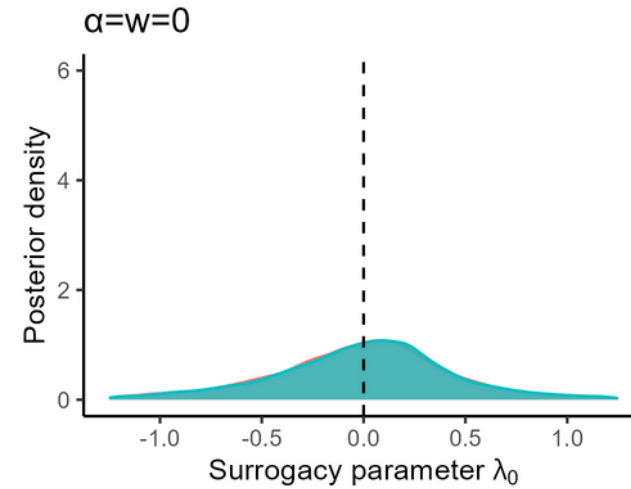
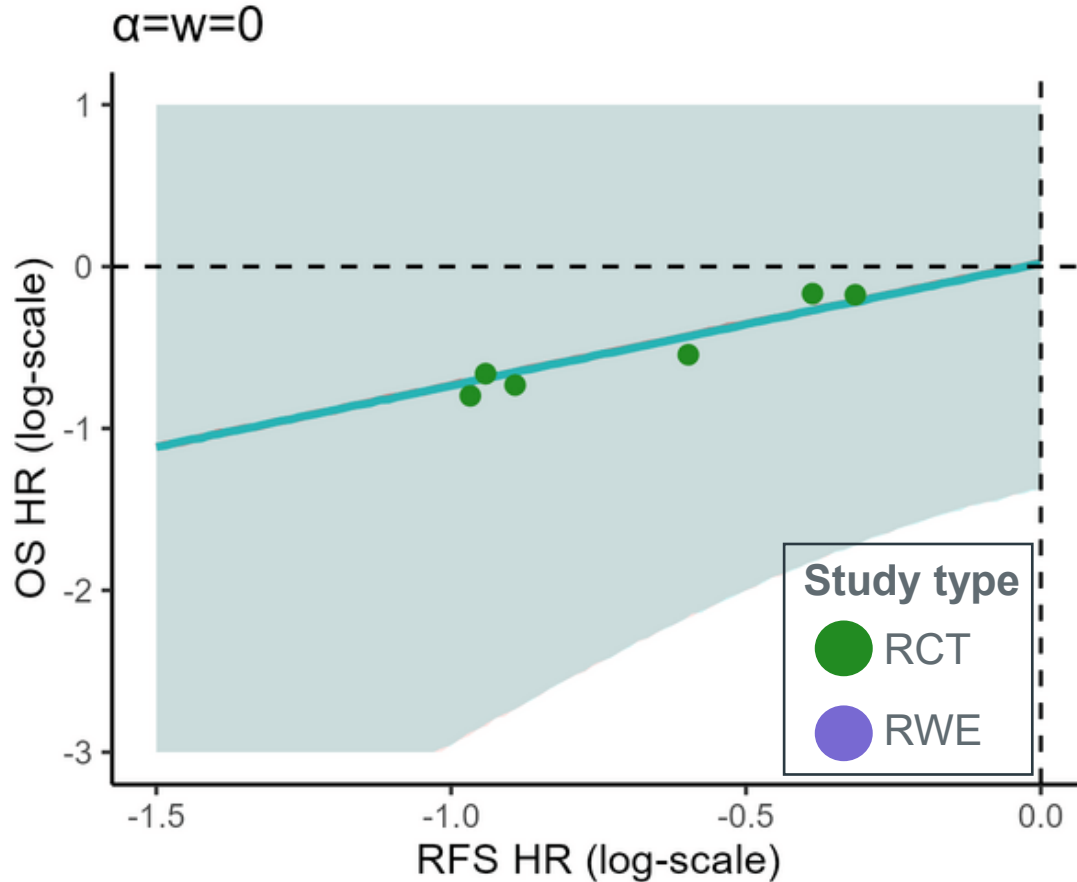
**Study type** ● RCT ● RWE

**Model type**

D&H on RCT only

Abbreviations: CrI, Credible interval; D&H, Daniel & Hughes model; LB, Lower bound; RCT, Randomised clinical trials; RFS, Relapse-free survival; RWE, Real-world evidence ;OS, Overall survival, UB, Upper bound.

# D&H with Mixture priors and Power priors map well to each other in terms of the degree of information borrowing they impose



Abbreviations: CrI, Credible interval; D&H, Daniel & Hughes model; LB, Lower bound; RCT, Randomised clinical trials; RFS, Relapse-free survival; RWE, Real-world evidence ;OS, Overall survival, UB, Upper bound.

# Conclusions

**1** A small pool of RCTs is, generally, unlikely to produce credible conclusions about surrogacy

**2** Including RWE, when appropriate, can reduce uncertainty and assist inference on surrogate endpoint evaluation

**3** Information-sharing methods can balance the RWE contribution to achieve the best of both worlds. Mixture and power priors map well in terms of the degrees of information sharing they impose

**4**

- There is a relatively strong evidence for surrogacy between RFS and OS in post-HSCT AML
- The inclusion of comparative RWE is not only defensible, but also vital for obtaining reliable inference on study-level surrogacy in the context of this case-study

