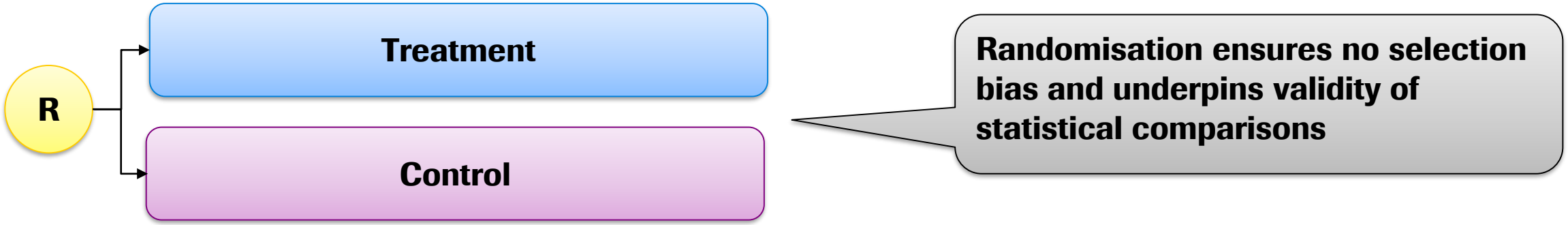
A Decision Making Framework For Utilising External Control Arms

Basel Biometric Section – May 2019
Chris Harbron, MCO, Roche

A Randomised Clinical Trial

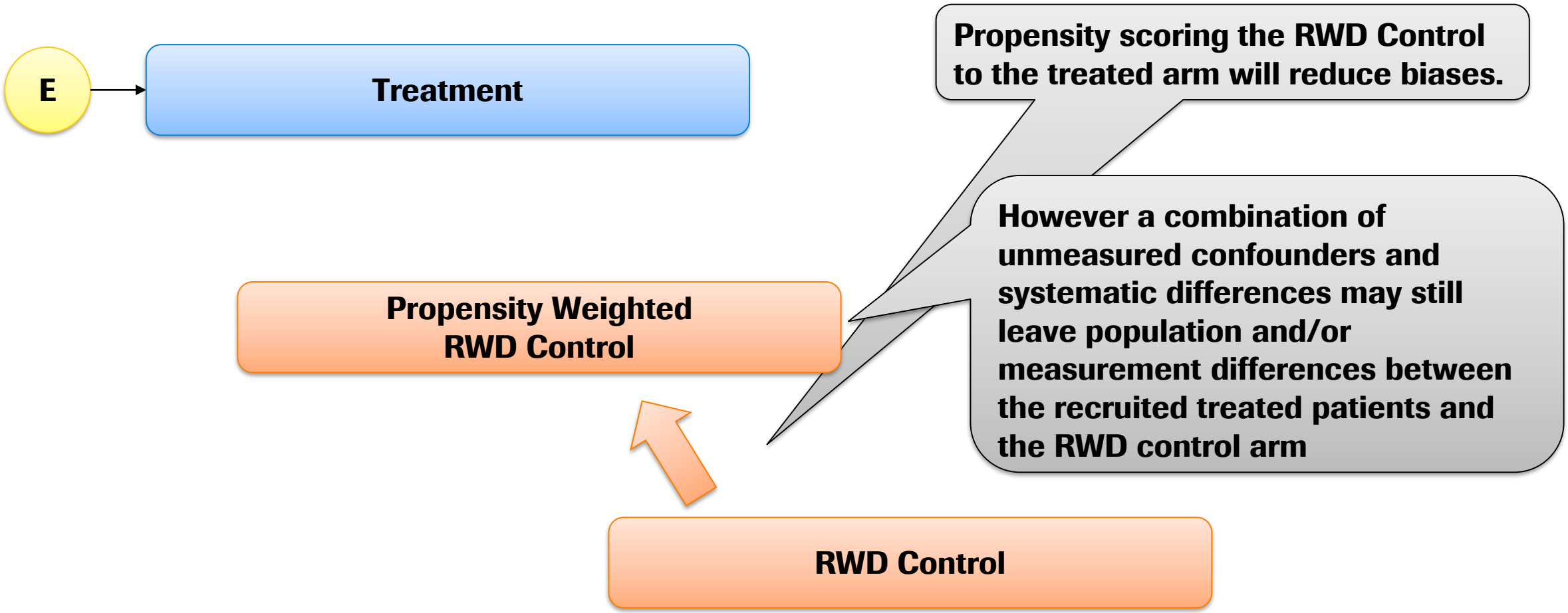


A Naïve Real World Data Control Arm

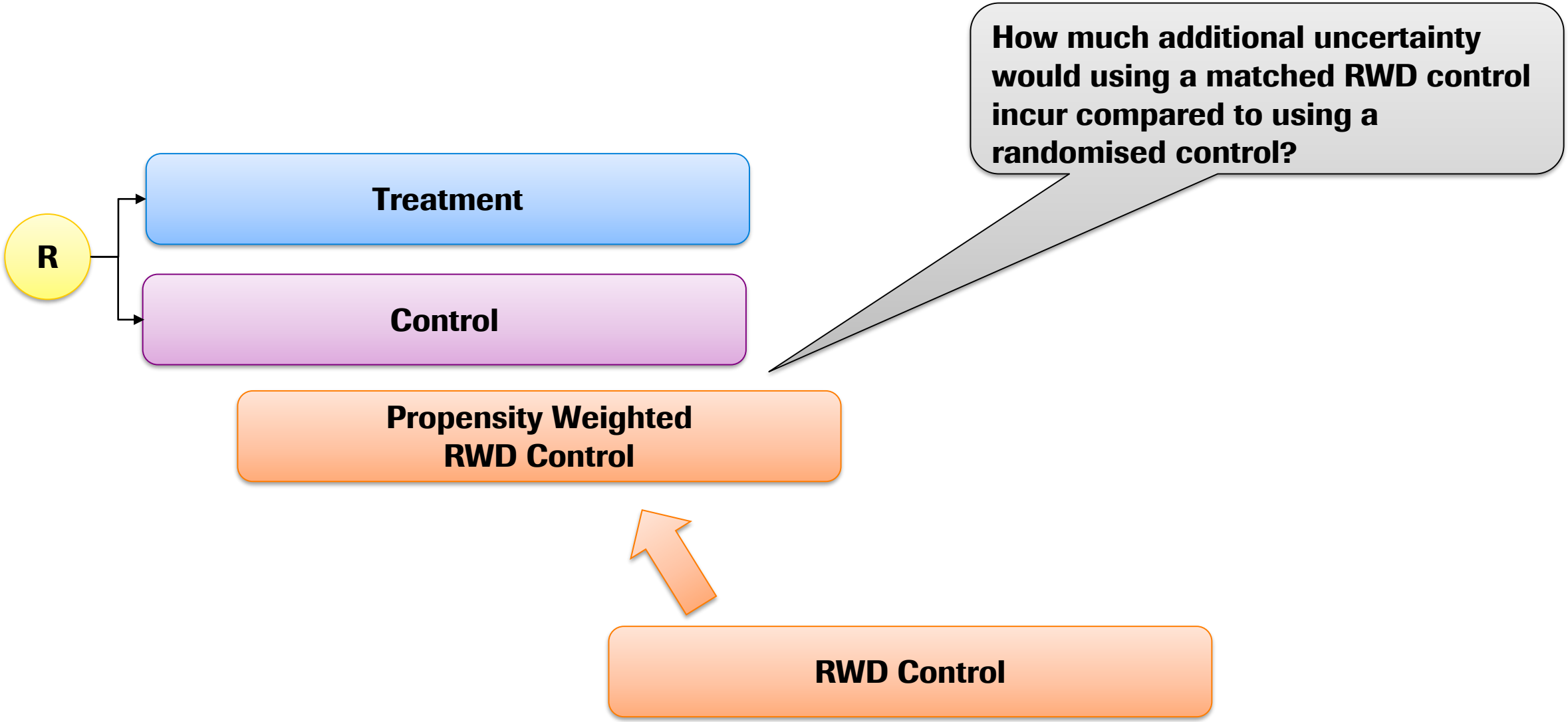
In the absence of a randomised control arm, naïve comparison of recruited treated patients with a RWD control is likely to be biased due to differences between populations

RWD Control

A Real World Data Control Arm With Propensity Scoring



Real World Data Control Arms



Why Might a Randomised and a RWD Control Arm Differ?

STUDY SPECIFIC BIASES

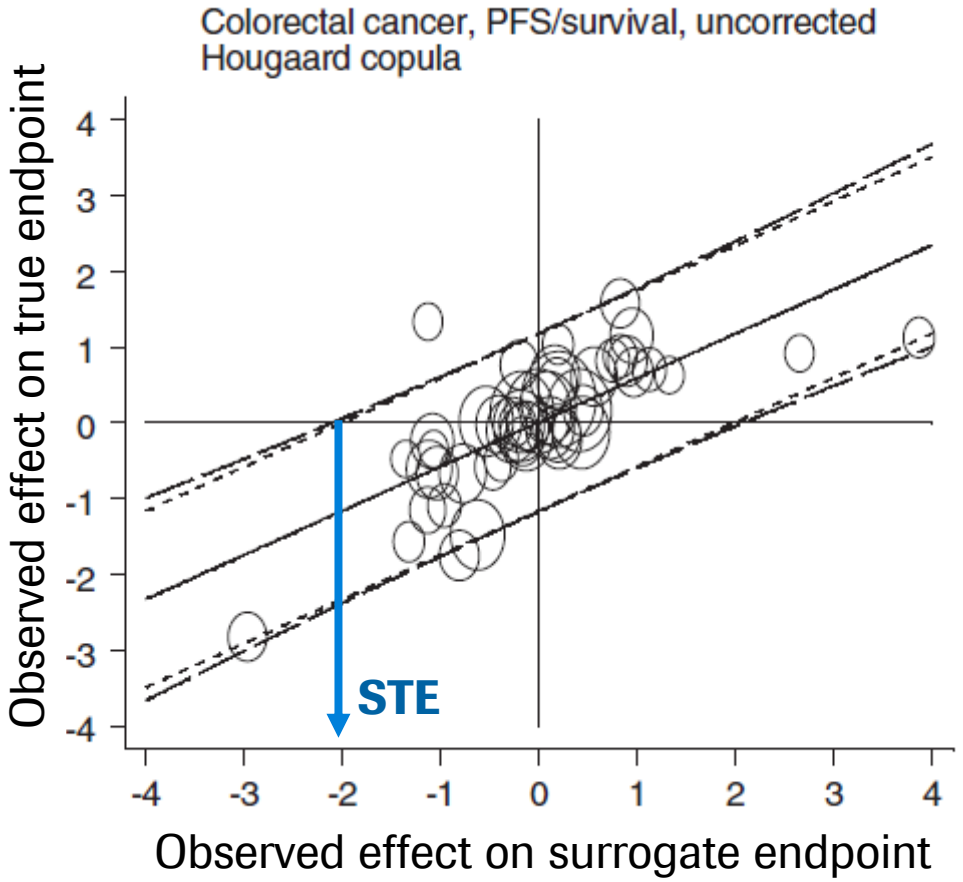
- **The process of being in a clinical trial**
  - Patient selection
  - Site selection
  - Higher levels of attention

SYSTEMATIC BIASES

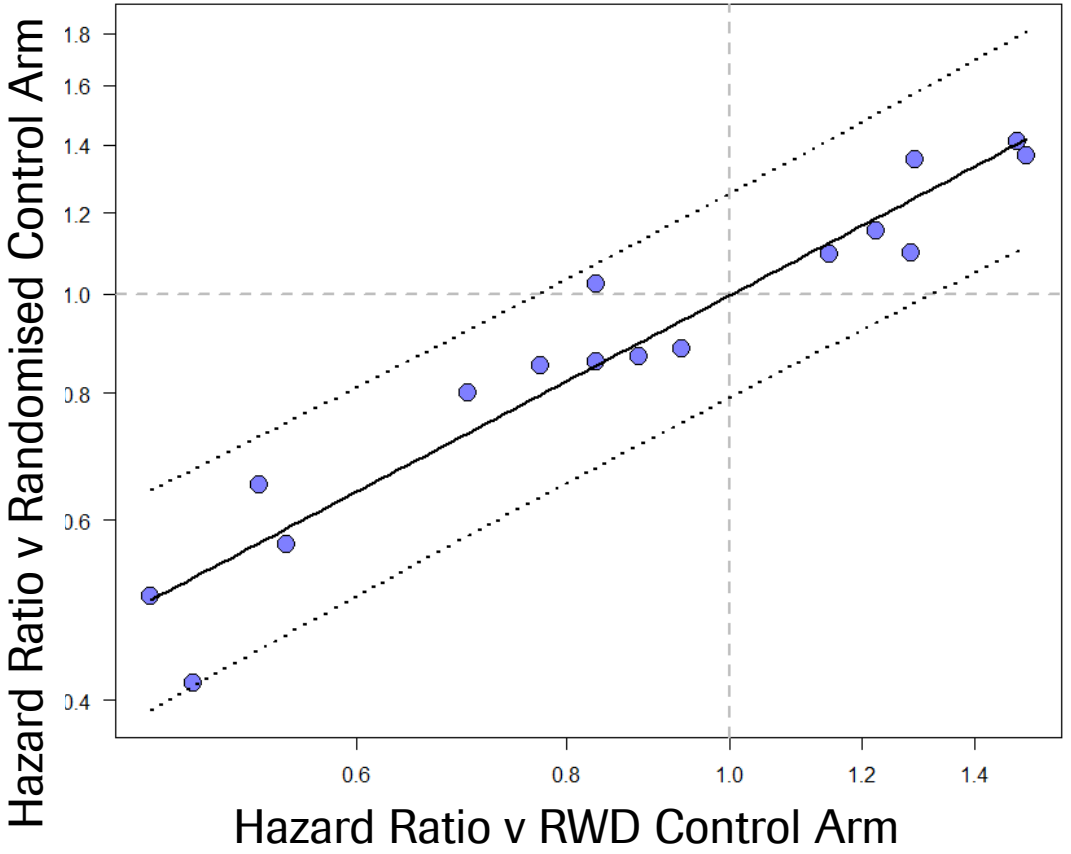
- **Measurement biases**
  - More events and data captured in clinical trials
  - Variables measured in different ways
  - Some variables (e.g. ECOG) captured in clinical trials but not in clinical practice
  - Clinical trial data may be collected on a more regular basis
- **Unmeasured confounders – Unknown unknowns**
  - We see that absolute results in clinical trials often vary more than we may expect
  - An external control may also have unexpected differences
  - No absolute bound on the size of these differences

Idea – Surrogacy & Surrogate Threshold Effect

Compare effects in two **endpoints**, e.g. PFS & OS
Objective : To be able to use PFS for **decision making**



Compare effects using two **control arms**, e.g. RCT & RWD
Objective : to be able to use RWD Control for **decision making**

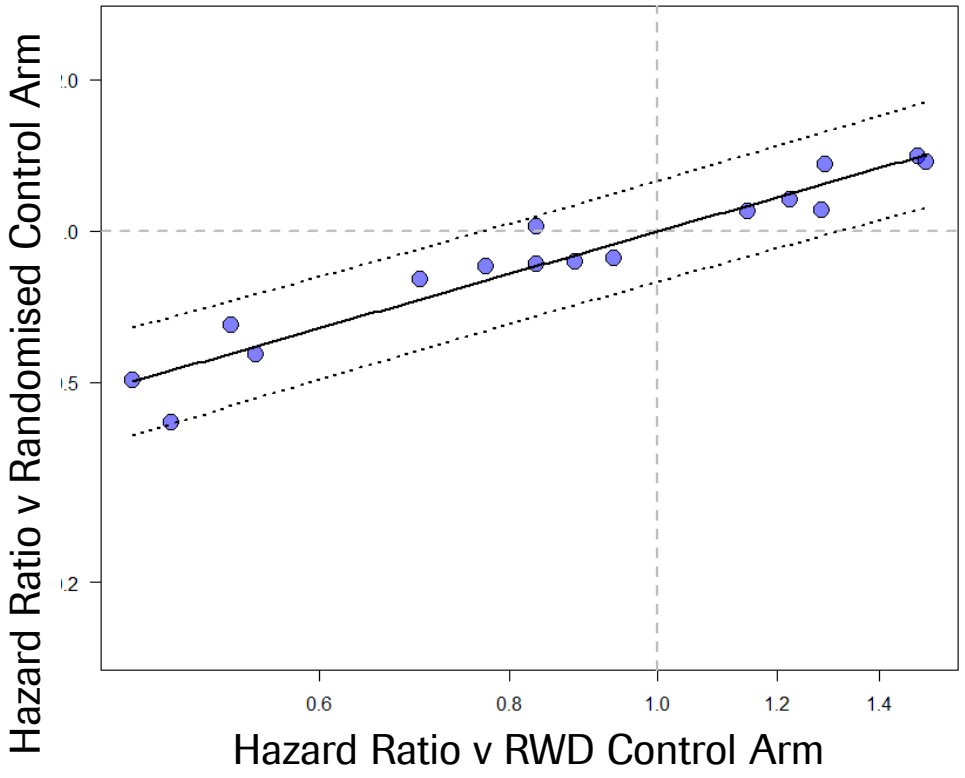


Note : RWD data shown in this presentation is artificial data

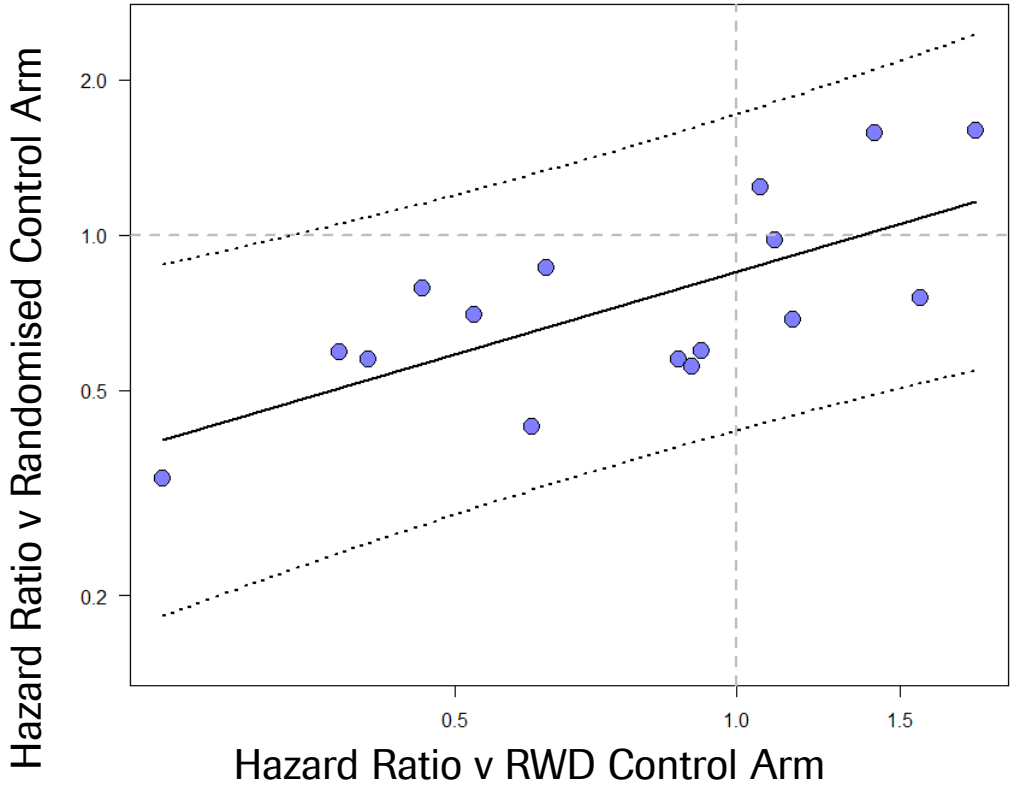
Surrogate threshold effect : An alternative measure for meta-analytic surrogate endpoint validation, Burzykowki & Buyes, Pharmaceutical Statistics 2006;5;173-186

We May Expect Different Endpoints / Indications to Exhibit Different Levels of Variability

Potential Relationship For RCT & RWD Overall Survival



Potential Relationship For RCT & RWD PFS

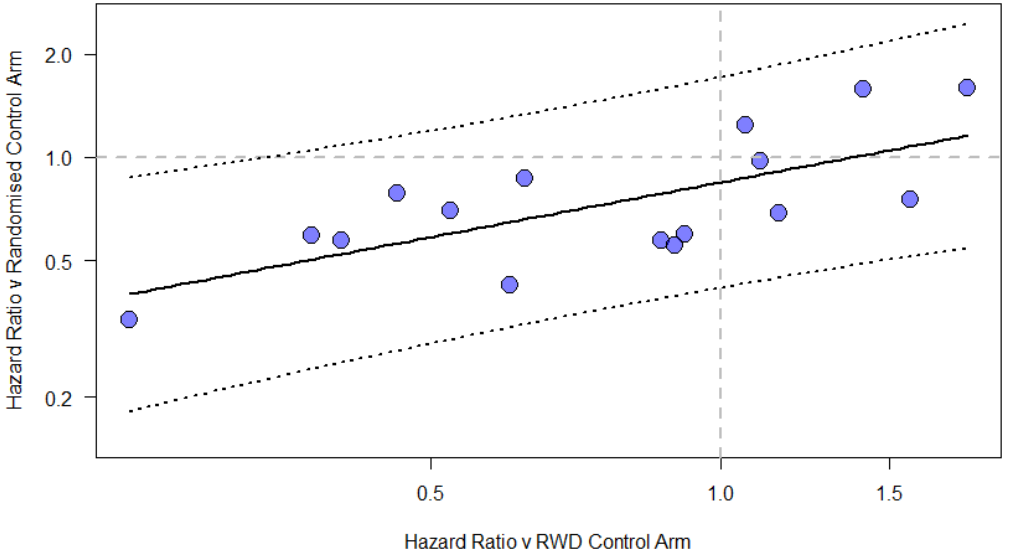


May expect as the methodology for collecting an outcome is refined, or as the understanding of key covariates for an indication grows allowing improved propensity scoring models to be fitted, that the variability will decrease

However, Surrogacy isn't a Perfect Analogy to our Situation

Classic Surrogacy	Comparing Control Arms
Deliberately different endpoints	Endpoints designed to be similar
Same patients	Different patients But same patients in treated arm
Same treatments	Same treatments
Plotting : T_{TRT} *v* T_{CTRL} against S_{TRT} *v* S_{CTRL} If T & S are unrelated, these will be unrelated	Plotting : T_{TRT} *v* T_{CTRL_RAND} against T_{TRT} *v* T_{CTRL_RWD} The common term T_{TRT} will induce a correlation

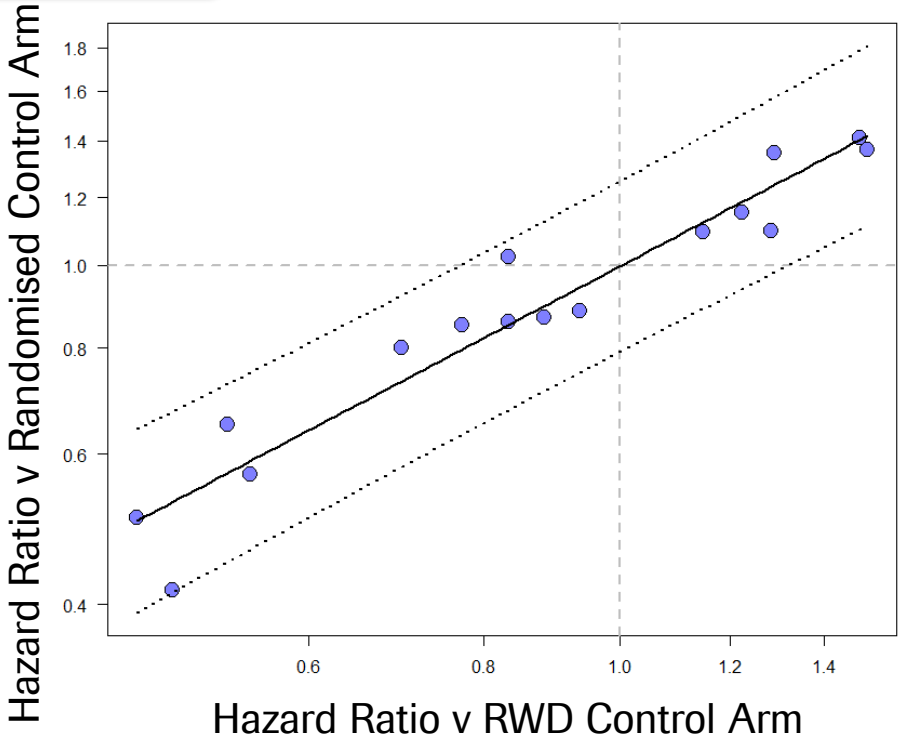
Issue : Most of the variability in such a plot, will represent variability in the performance of the novel treatment comparators.
No reason to think the difference between control arms should be related to how good the treatment being compared to is
For our aim of understanding the relationship between the different flavours of controls this is a distraction



For a More Rigorous Examination of Bias & Variability, Rotate Through 45% -> Bland-Altman Plot

Unlike surrogacy, as units on both axes are on the same scales, taking differences and averages makes sense

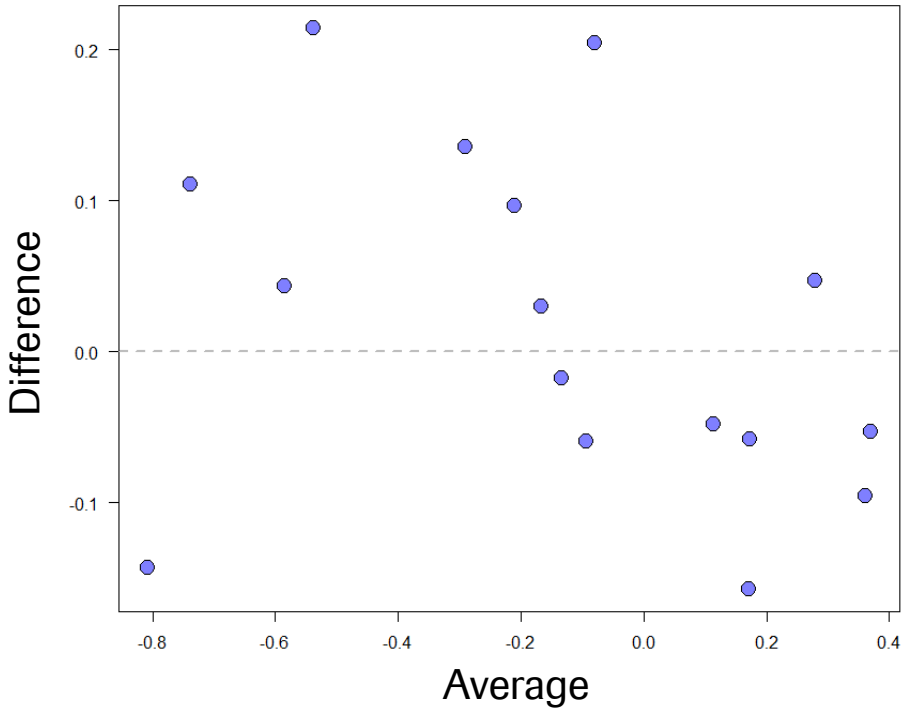
T_T v T_{C_RCT}



T_T v T_{C_RWD}

T_{C_RCT} v T_{C_RWD}

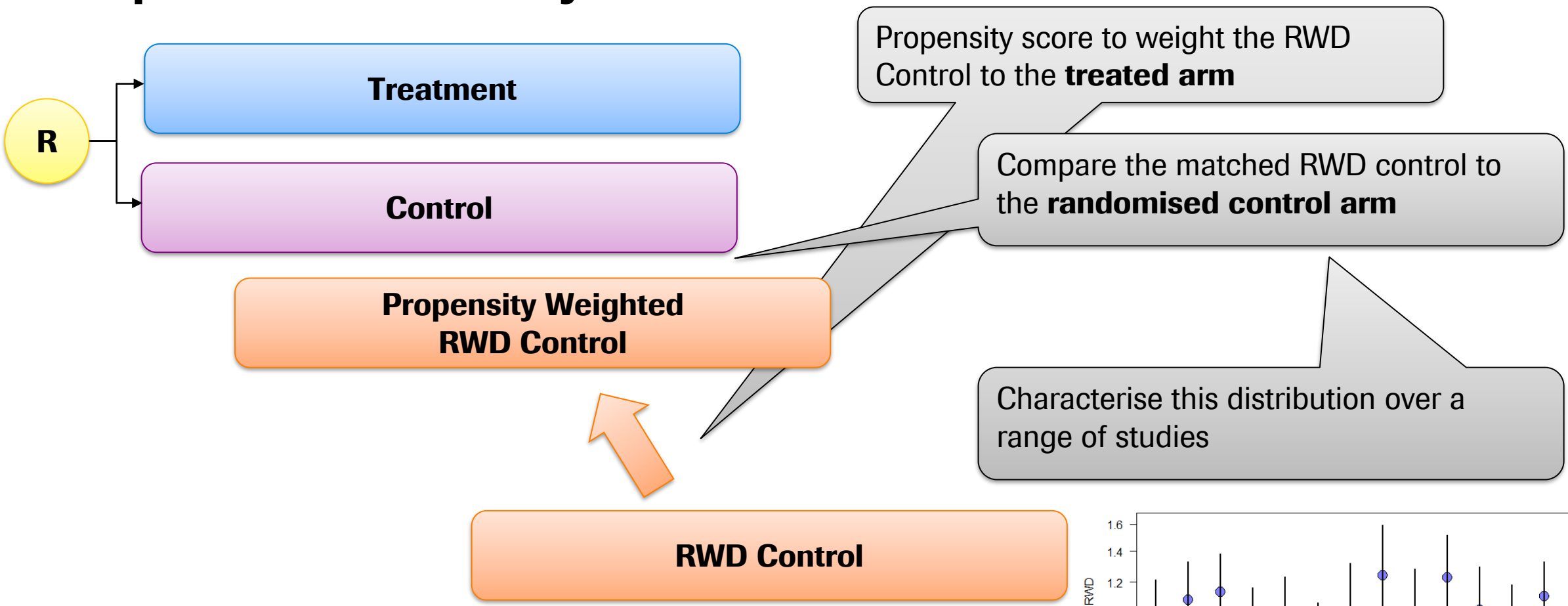
This is what we are interested in

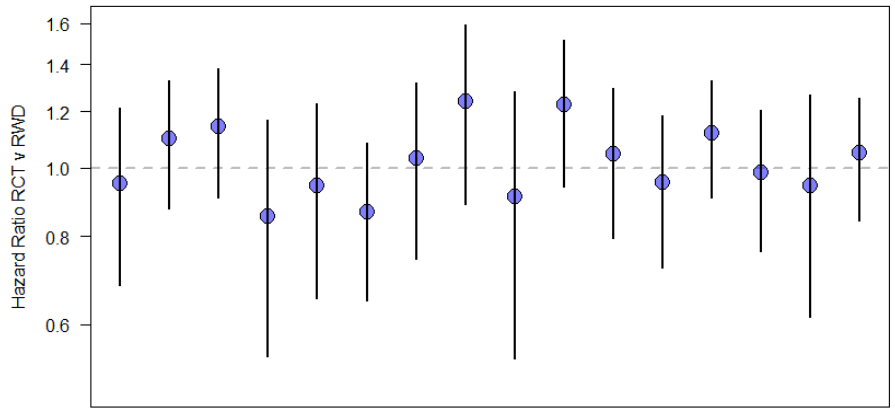


T_T v 0.5*(T_{C_RCT} + T_{C_RWD})

This is a nuisance parameter

Proposal : Focus Directly on the Differences Between RCT & RWD





Once Characterised : How Do We Use This Distribution?

Using the Approximation :

$$Log(HR\ Trt\ v\ Rand\ Cont) \approx Log(HR\ Trt\ v\ RWD\ Cont) - Log(HR\ RWD\ Cont\ v\ Rand\ Cont)$$

Systematic Bias

$$E[Log(HR\ Trt\ v\ Rand\ Cont)] \approx E[Log(HR\ Trt\ v\ RWD\ Cont)] - E[Log(HR\ RWD\ Cont\ v\ Rand\ Cont)]$$

Study Specific Bias

$$Var[Log(HR\ Trt\ v\ Rand\ Cont)] \approx Var[Log(HR\ Trt\ v\ RWD\ Cont)] + Var[Log(HR\ RWD\ Cont\ v\ Rand\ Cont)]$$

Standard Analysis

Distribution of CT v RWD differences

Comparing Back To Surrogacy

Surrogacy

Prediction Variance : $Var(\beta + b_0|\mu_{S0}, \alpha_0, \vartheta) \approx f\{Var(\hat{\mu}_{S0}, \hat{\alpha}_0)\} + f\{Var(\hat{\vartheta})\} + (1 - R^2_{trial})Var(b_0)$

Variability from measurement of surrogate

Variability from uncertainty of model

Residual lack of fit of the model

No model, assumed constant

External Controls

$Var[Log(HR\ Trt\ v\ Rand\ Cont)] \approx Var[Log(HR\ Trt\ v\ RWD\ Cont)] + Var[Log(HR\ RWD\ Cont\ v\ Rand\ Cont)]$

Variability of assessed comparison

Residual differences between RCT & RWD Controls

What Would be the Impact on the Analysis?

- To reflect the increased uncertainty :
  - A widening of confidence intervals, which will lead to :
  - a decrease in the alpha level required for statistical significance in the unadjusted analysis, or
  - an adjustment to the Minimal Detectable Difference, requiring a larger effect size

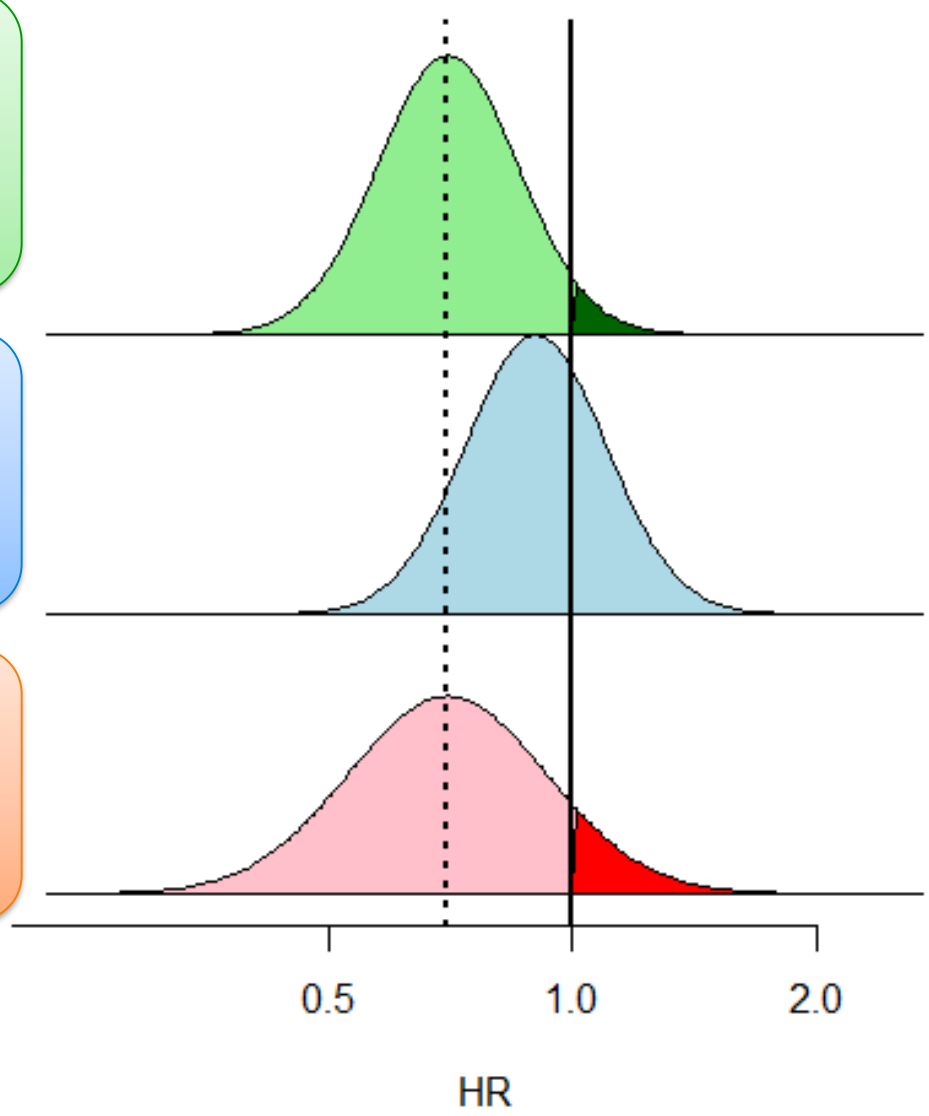
- A shift of the estimate of treatment effect to reflect any systematic biases :
  - To be conservative, propose an asymmetric approach
  - Don’t adjust if RWD control is on average a worse outcome than RCT control
  - Adjust if RWD control is on average a better outcome than RCT control

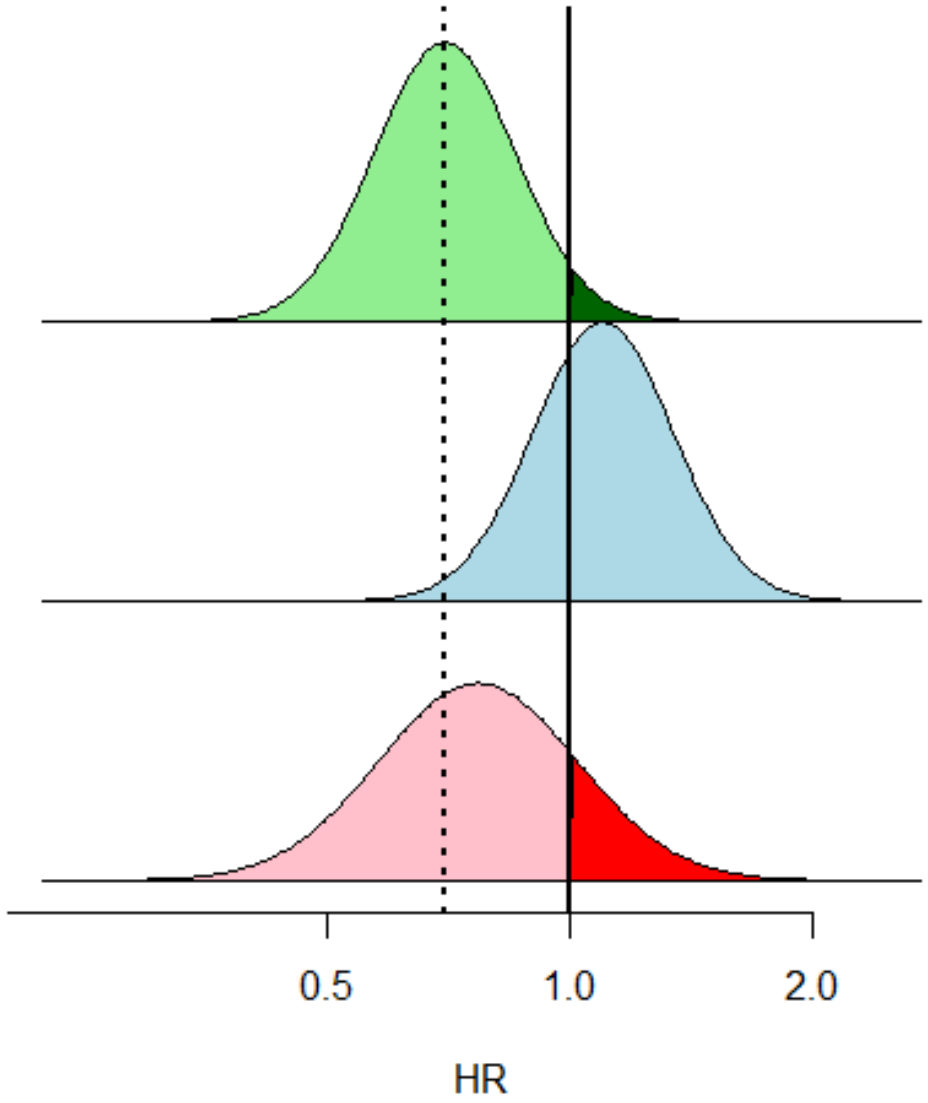
What Would be the Impact on the Analysis?

Treatment vs RWD Control Comparison

Additional Variability Associated With Non-Randomised Comparison

Overall Assessment Of Treatment Effect





Considerations & Points For Discussion

- **Variances are highly variable to estimate**
  - **→ Consider using t-distributions rather than normal to capture this uncertainty**
- **May have a non-normal distribution of comparisons**
  - **For example with one or two studies showing larger differences where matching has failed**
  - **→ Consider using mixture distributions to model the variability**
- **If the propensity scoring was to fail to deliver comparable populations in one study out of twenty, that is already spending our type 1 error**
  - **→ How many historical studies would we need to demonstrate the robustness of the relationship?**
- **Potential applications in internal and regulatory decision making**
  - **→ How may the required levels of evidence differ between these situations**

Doing now what patients need next