Disclosure

The presenter has advised that the following presentation is subject to no conflicts of interest and has nothing to disclose.

“Anchoring effect” in Random Effects Meta Analysis: a cautionary tale

The George Institute for Global Health

Gian Luca Di Tanna, Abdul Salam, Anna Campain and Anthony Rodgers

Head of Statistics, Australia

GDiTanna@georgeinstitute.org.au GLDiTanna

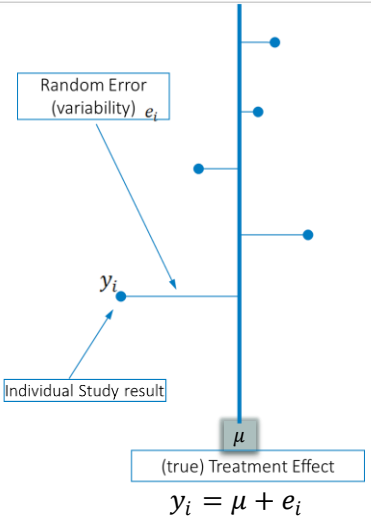
With Acknowledgements to Carinna Hockham

Introduction

- Meta Analysis refers to the statistical synthesis of results from a series of systematically reviewed studies addressing similar clinical questions
- I will refer in this talk to head-to-head Meta Analysis of randomised controlled trials
- With the massive growth in number of Meta Analysis, increasing importance in updated results: "***how does this new trial change the overall result?***"
- This talk introduces the concept of ***anchoring*** – an undesirable situation in which, with certain methods, a previous pooled effect is impervious to change with the advent of new trial evidence

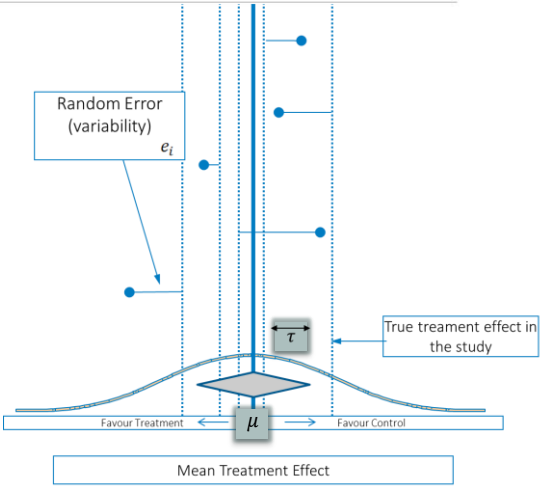
Models available (and their assumptions): the Fixed Effect Model (FEM)

- We rely on the pooled ("weighted") effect, based on the assumption that each study is a near replication of a single experiment
- ***Differences among study results are due only to chance***
- A FEM would be appropriate if the objective is to determine whether the treatment had an effect in the observed studies (i.e., a ***conditional inference***) or when it is believed that the true treatment effects in each study are the same



Models available (and their assumptions): the Random Effects Model (REM)

- REM assumes that the observed treatment effect for a study is a combination of a treatment effect common to all studies plus a component specific to that study alone
- REM also assumes that ***all studies are randomly drawn from a pool*** of all possible evidence
- Common error to adopt REM in presence of statistical heterogeneity (SH):
  - Lack of SH does not preclude heterogeneity
  - Presence of SH does not confirm it is quantitatively important



$$y_i = \mu + \delta_i + e_i$$
$$e_i \sim N(0, v_i)$$
$$\delta_i \sim N(0, \tau^2)$$

The George Institute for Global Health

FEM vs REM

- Both models have strong assumptions
- The model choice depends on the objective of the analysis and knowledge of the included studies
- FEM are commonly used when there are too few studies with which to estimate the τ from the data alone
- REM most commonly used because:
  - Correctly, it reflects an extra source of potential variability
  - **Incorrectly, it is assumed to 'take care' of heterogeneity**
- The most widely used REM is based on an estimator developed by DerSimonian and Laird (DL) in mid-1980s. It is the standard (or only) method available in any software

Which REM? Which τ?

- DL confidence intervals too narrow (high false positive rates) when few studies and/or large between-study differences
- Hartung-Knapp and Sidik-Jonkman method (HKSJ) makes small sample size adjustments to the variance estimates
  - consistently better error rates than DL method
  - many suggest it should be the standard for REM

- **There are >15 methods to estimate τ and >5 methods to estimate its uncertainty with variability across difference software packages**
- The imprecision in the estimates of τ are not acknowledged in both FEM and REM
- Authors have advocated the routine use of prediction intervals to represent the heterogeneity in the same metric of the intervention effect size measure, and to give a range of values that can be expected in future settings

Why not go Bayesian?

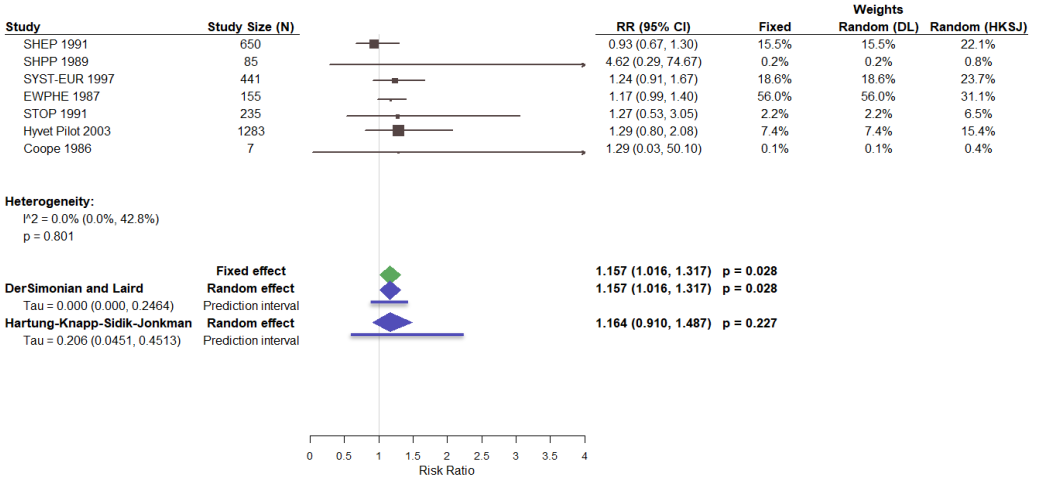
- Bayesian random-effects models provide a better accounting of the uncertainty at a cost of being slightly more computationally intensive
- We can incorporate external evidence on the treatment effects, τ and their relative uncertainty
- Usually the base-case scenarios are around the use of non-informative priors. *But vague priors can be informative in any case*
- **It's generally agreed to use a non-informative prior** on the treatment effect (although we may like to employ historical priors as we accumulate further evidence)
- **It is desirable to use informative prior distributions on τ**: there are various "off-the-shelf" priors (see Turner et al., 2014) to better capture between-study heterogeneity expected in various research settings

Motivating example

- Meta Analysis on antihypertensive drugs in those aged >80 years (Gueyffier et al, Lancet 1999)
- Comprising 7 (subgroup analysis) trials and 2,856 patients
- It showed statistically significant reduction of non-fatal events
- …but borderline statistically significant adverse mortality results

The George Institute for Global Health

Meta Analysis



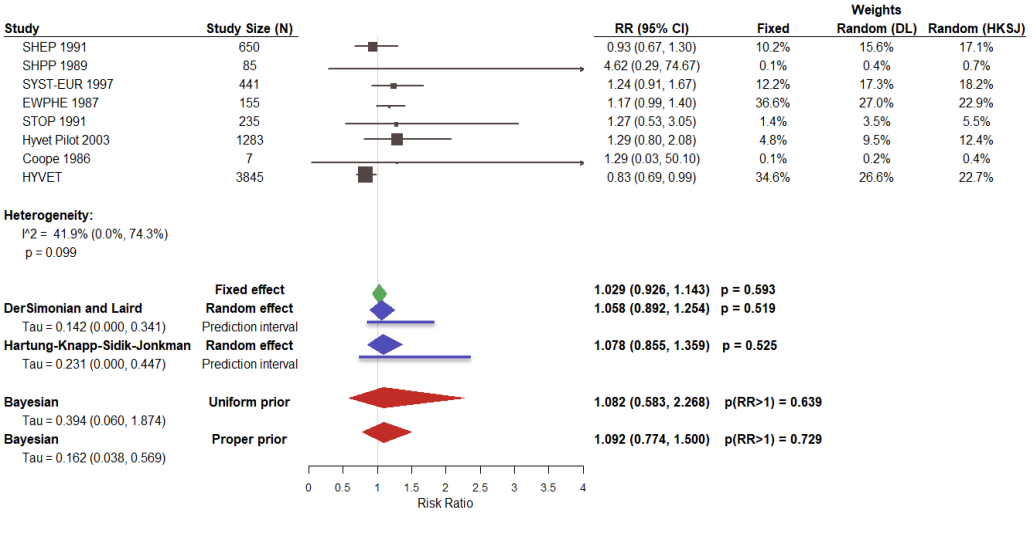
The George Institute for Global Health

The follow-up story

- Meta Analysis on antihypertensive drugs in those aged >80 years (Gueyffier et al, Lancet 1999)
- Comprising 7 (subgroup analysis) trials and 2,856 patients
- It showed statistically significant reduction of non-fatal events
- …but borderline statistically significant adverse mortality results
- Then came HYVET trial with 3,845 patients aged >80 years (NEJM, 2008)
- Statistically significant reduction in nonfatal events and mortality associated with hypertensive treatment

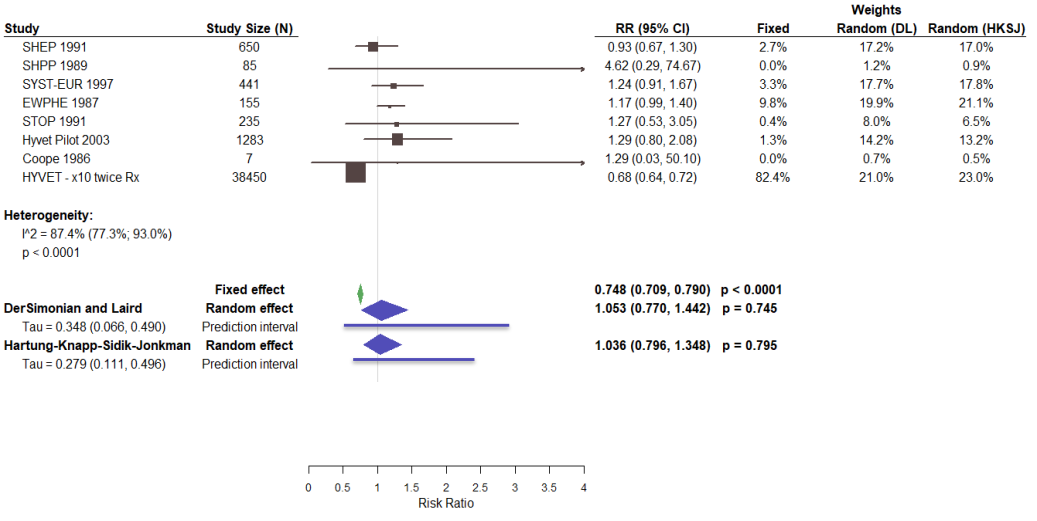
The George Institute for Global Health

Meta Analysis with HYVET



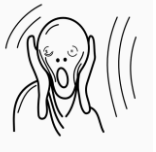
The George Institute for Global Health

What if HYVET was 10x bigger with 2x treatment effect?



Summary table

Scenario	FEM RR (95%CI)	Weight given to HYVET trial	DL-REM RR (95%CI)	Weight given to HYVET trial	HKSJ-REM RR (95%CI)	Weight given to HYVET trial	Bayesian Meta Analysis RR (95% CrI)
Prior to HYVET	1.16 (1.02; 1.32)	-	1.16 (1.02; 1.32)	-	1.16 (0.91; 1.49)	-	1
With HYVET	1.03 (0.93; 1.14)	35%	1.06 (0.89; 1.25)	27%	1.08 (0.86; 1.36)	23%	1
With HYVET (10x sample size)	0.87 (0.83; 0.92)	84%	1.06 (0.86; 1.32)	26%	1.07 (0.85; 1.34)	25%	1
With HYVET (100x sample size)	0.83 (0.82; 0.85)	98%	1.07 (0.86; 1.33)	26%	1.07 (0.85; 1.34)	25%	0



Final remarks

- Every model comes with assumptions
- Be careful with the standard approaches - FEM tends to over-react, **REM tends to remain anchored**
- **Do not rely on a single model**, challenge the findings
- Do not neglect the between-study variance and its uncertainty
- **Bayesian Meta Analysis** is a flexible option but do not miss the opportunity to employ adequate/informative priors (at least for τ) to overcome the weaknesses of FEM and REM (taken in isolation)

The George Institute for Global Health

Thank you!

The George Institute for Global Health