Evidence synthesis for a single randomized controlled trial and observational data in small populations

Steffen Unkel, Christian Röver, Tim Friede

Department of Medical Statistics
University Medical Center Goettingen, Germany

InSPiRe
Innovative methodology for small populations research

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-602144.

Motivation

- Usually two independent randomized controlled trials (RCTs) are required to demonstrate efficacy and safety for marketing authorization.

- In small populations the conduct of a single RCT with a sufficient sample size might be difficult or not feasible.

- This is particularly the case
  1. in paediatric studies,
  2. if the intervention is to treat a rare disease, or
  3. if randomization is challenging.

- In situations where randomization is difficult to achieve, methods that incorporate data from other sources in the estimation of the treatment effects may be beneficial.

Examples

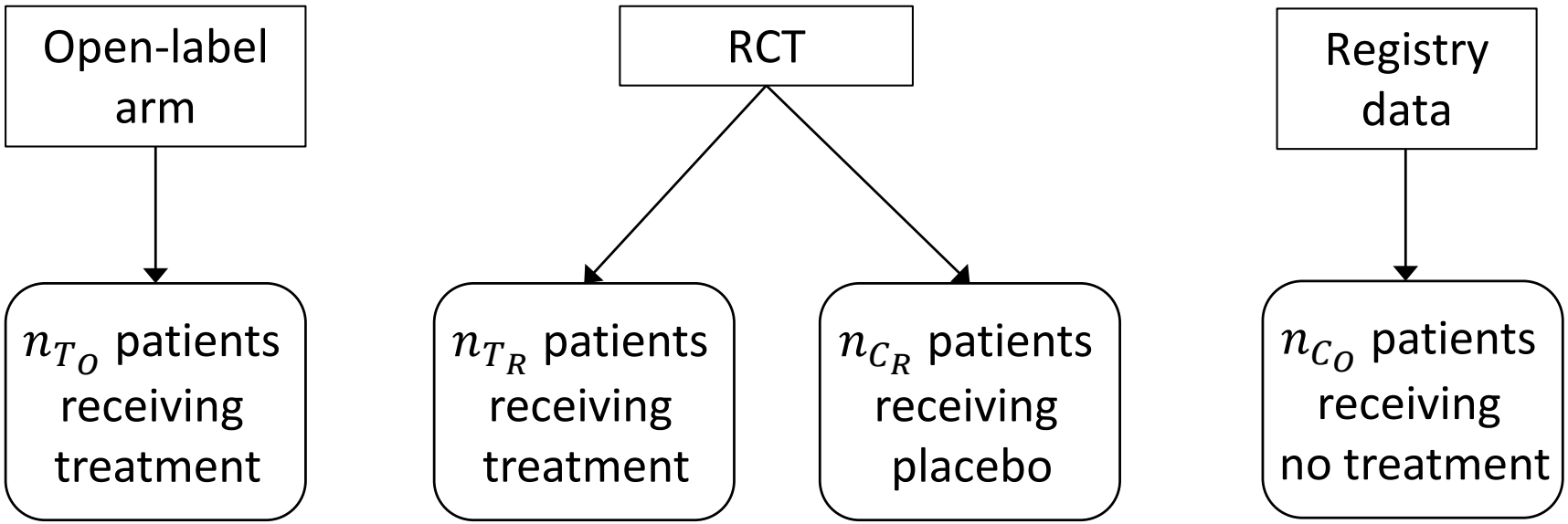
Examples where the required number of patients were not randomized include

1. several RCTs in patients with ankylosing spondylitis.
   Patients were randomly assigned (in a 4:1 ratio) to either treatment or placebo. To support the small placebo control, data from eight previous trials in patients with ankylosing spondylitis were included (Baeten et al. 2013).

2. an RCT in patients with Creutzfeldt-Jakob disease
   Meta-analysis combining evidence on the effects of a certain treatment in patients with Creutzfeldt-Jakob disease from both a randomized study and a non-randomized study (Varges et al. 2017).

3. the EARLY PRO-TECT trial in paediatric Alport patients (Gross et al. 2012a).

Trial design that mimics the Alport trial



Endpoints

- The primary efficacy endpoint in the EARLY PRO-TECT Alport trial is "time-to-progression to the next disease level".
- This time-to-event endpoint will be assessed in 6-monthly intervals over the treatment period of 3 years.
- The second efficacy endpoint "albuminuria after 3 years corrected for baseline albuminuria for patients randomized to receive ramipril compared to placebo" is continuous.
- One might also think of binary endpoints such as "progression to the next disease level within 3 years (yes/no)".

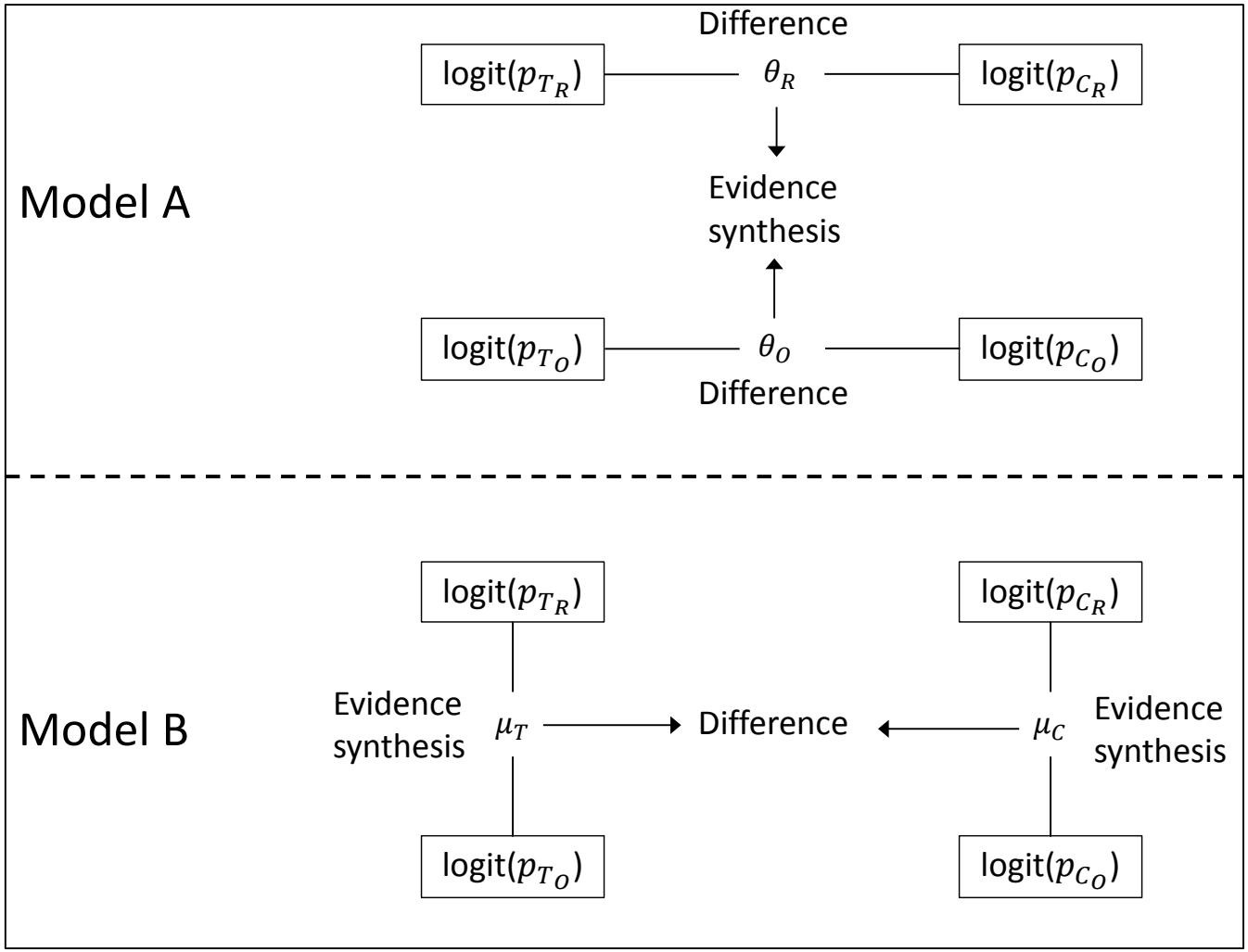
Data

- We consider a binary endpoint.
- Randomized arms: let X_{i_R} be the number of events and p_{i_R} denote the probability of an event in group i $(i = T, C)$.
- Non-randomized arms: let X_{i_O} be the number of events and p_{i_O} denote the probability of an event in group i $(i = T, C)$.
- Binomial model:

$$X_{i_j} \sim \mathcal{B}(n_{i_j}, p_{i_j}) \ , \quad i = T, C; j = R, O \ .$$

- Let $\theta_R = \log\left(\frac{p_{T_R}(1-p_{C_R})}{p_{C_R}(1-p_{T_R})}\right)$ and $\theta_O = \log\left(\frac{p_{T_O}(1-p_{C_O})}{p_{C_O}(1-p_{T_O})}\right)$ denote the log odds ratio for the randomized and observational data, respectively.

Model frameworks



Arm-based versus contrast-based synthesis of data

Commentary

Research Synthesis Methods

Received 6 July 2015, Revised 28 July 2015, Accepted 13 August 2015 Published online 13 October 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1184

Absolute or relative effects? Arm-based synthesis of trial data

S. Dias* and A. E. Ades

We congratulate Hwanhee Hong and colleagues on another fascinating paper (Hong *et al.,* 2015a) arguing the case for arm-based models for meta-analysis.

The standard approach to meta-analysis is the *contrast-based* model where the information that is pooled over trials is the information of the trial-specific *relative* treatment effect, expressed for example as a log relative risk, log odds ratio, or as a mean treatment difference. In an *arm-based* model, it is the *absolute* log risk, log odds, or mean outcome on each arm that are pooled.

Methods for evidence synthesis

1. The power prior approach assigns a weight to the external data somewhere in between the cases of irrelevance and full equality.

2. Bias allowance models assume that the external data are potentially biased and the potential bias is modelled using an extra variance component that represents the bias.

3. Meta-analytic approaches or hierarchical models for evidence from different study designs are an extension of standard random-effects meta-analysis that explicitly model between-study-type variability.

Hierarchical models

- The hierarchical structure of model A may be stated as

$$y_j|\theta_j, s_j \sim \mathcal{N}(\theta_j, s_j^2) \ ,$$
$$\theta_j|\mu, \tau \sim \mathcal{N}(\mu, \tau^2) \ , \quad j = R, O \ ,$$

  where y_j is an estimate of θ_j and s_j is its standard error.

  The θ_j differ from study to study and are distributed around a common mean μ with between-study-type variability or heterogeneity τ.

- The framework for model B consists of two hierarchical structures with parameters (μ_T, τ_T) and (μ_C, τ_C).

  The overall treatment effect is computed as a contrast: $\mu_T - \mu_C$.

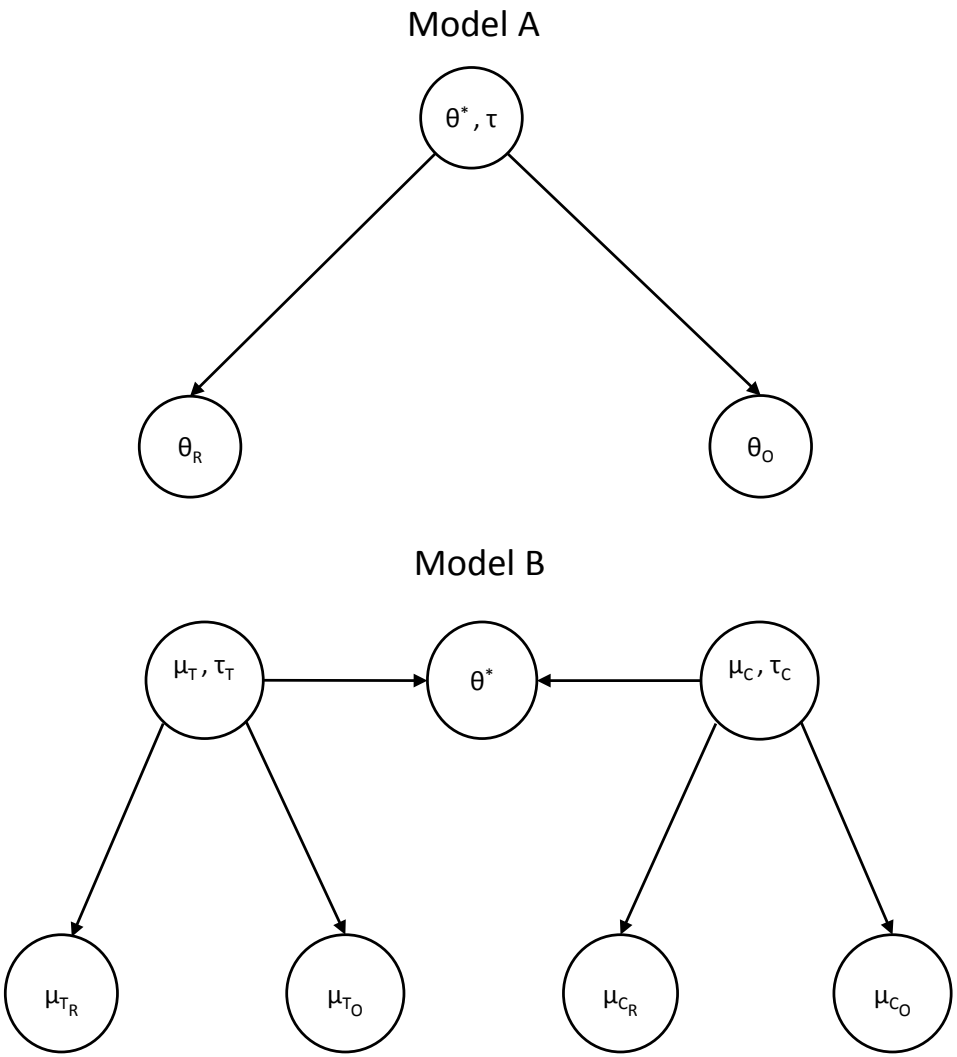


Figure: Hierarchical structures for model A (top) and model B (bottom).

Fitting Bayesian hierarchical models

- We use a Bayesian approach for fitting the hierarchical models.
- Inference for μ and τ is captured by the joint posterior distribution, from which the marginal distribution of μ is used to derive point estimates and probability intervals for μ.
- Our approach requires prior distributions for μ and τ:
  - For μ one may use a noninformative (improper) uniform prior or a normal prior with mean zero and large variance.
  - For τ we use half-normal (HN) prior distributions.
- The R package `bayesmeta` provides a collection of functions to facilitate Bayesian inference in the random-effects meta-analysis model.

Generating data

RCT	Treatment	Control
No event	31	9
Event	9	11
$\sum$	$n_{T_R} = 40$	$n_{C_R} = 20$

Log odds ratio $y_R = 1.4374$
Standard error $s_R = 0.5877$

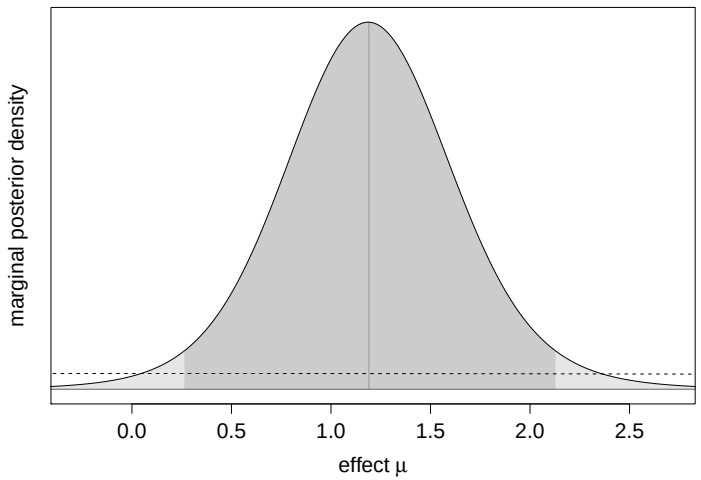
Observational data	Treatment	Control
No event	29	29
Event	11	31
$\sum$	$n_{T_O} = 40$	$n_{C_O} = 60$

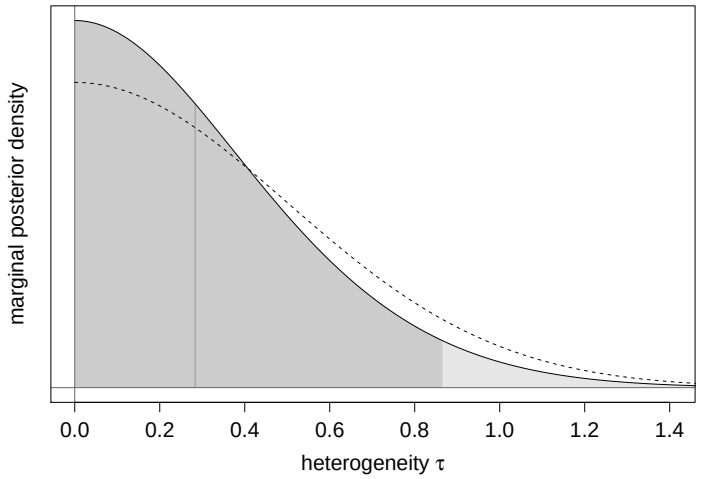
Log odds ratio $y_O = 1.0361$
Standard error: $s_O = 0.4383$

Fitting model A

- `bma <- bayesmeta(y, s, mu.prior.mean=0, mu.prior.sd=10, tau.prior=function(t){dhalfnormal(t,scale=0.5)})`
- Marginal posterior summary:

	tau	mu
mode	0.0000	1.1870
median	0.2833	1.1960
mean	0.3428	1.1931
sd	0.2680	0.4699
95% lower	0.0000	0.2637
95% upper	0.8651	2.1278





Fitting model B

- Compute estimates for the logits(p_{i_j}) ($i = T, C; j = R, O$) and associated standard errors.

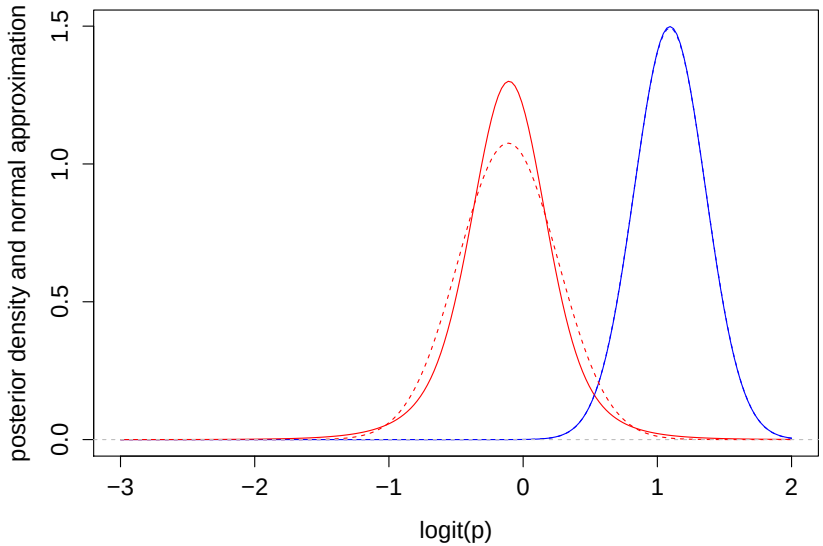
- ```
  bma.t <- bayesmeta(y=yt, s=st, labels=names(yt),
                     mu.prior.mean=0, mu.prior.sd=10,
                     tau.prior=function(t){dhalfnormal(t, scale=0.1)})

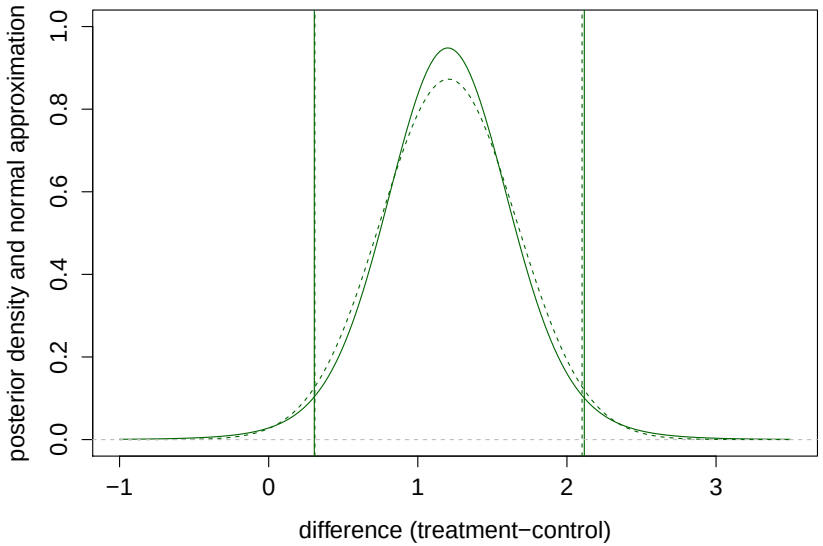
  bma.c <- bayesmeta(y=yc, s=sc, labels=names(yc),
                     mu.prior.mean=0, mu.prior.sd=10,
                     tau.prior=function(t){dhalfnormal(t, scale=0.5)})
  ```

- Compute the convolution, that is, the distribution of the difference (treatment - control).
  ```

Fitting model B (2)

Difference		Model A	
mean	standard error	mean	sd
1.2056	0.4571	1.1931	0.4699





	2.5%	97.5%
Normal.approx	0.3097	2.1015
Convolution	0.3059	2.1165
Model A	0.2637	2.1278

Simulation setup: meta-analysis scenario

- We investigate the performance of models A and B by means of a so-called general meta-analysis scenario.
- We assume that we observe four logit estimates (y_{T_R}, y_{C_R}, y_{T_O}, y_{C_O}) and associated standard errors (s_{T_R}, s_{C_R}, s_{T_O}, s_{C_O}).
- The underlying true effects are $\mu_{i_j} = \text{logit}(p_{i_j})$ $(i = T, C; j = R, O)$.
- The effects $(\mu_{T_R}, \mu_{C_R}, \mu_{T_O}, \mu_{C_O})^\top$ are assumed to follow a multivariate normal distribution with mean $(\mu_T, \mu_C, \mu_T, \mu_C)^\top$ and covariance matrix $\mathbf{\Sigma}$.

Simulation setup: meta-analysis scenario (2)

- We consider a scenario in which there are dependencies between the two randomized arms and between the two observational arms only.

- That is, the covariance matrix of $(\mu_{T_R}, \mu_{C_R}, \mu_{T_O}, \mu_{C_O})^\top$ is assumed to be

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2 - \frac{\tau_R^2}{2} & 0 & 0 \\ \sigma^2 - \frac{\tau_R^2}{2} & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \sigma^2 - \frac{\tau_O^2}{2} \\ 0 & 0 & \sigma^2 - \frac{\tau_O^2}{2} & \sigma^2 \end{pmatrix},$$

where $\tau_R^2 = \text{Var}(\mu_{T_R} - \mu_{C_R})$ and $\tau_O^2 = \text{Var}(\mu_{T_O} - \mu_{C_O})$.

Simulation setup: data and evaluation criteria

- Four groups motivated by the EARLY PRO-TECT study protocol:
  - $n_{T_R} = 40$: 8 failures / 32 successes
  - $n_{C_R} = 20$: 10 failures / 10 successes
  - $n_{T_O} = 40$: 8 failures / 32 successes
  - $n_{C_O} = 60$: 30 failures / 30 successes

- This leads to standard errors: $s_{RT} = 0.4$, $s_{RC} = 0.45$, $s_{OT} = 0.4$ and $s_{OC} = 0.26$

- Using 2000 simulation runs per parameter combination, we computed
  1. observed coverages for 95% confidence intervals for the pooled effect.
  2. lengths of meta-analytic confidence intervals relative to the interval length of the RCT.

Simulation setup: between-study heterogeneity

- For the between-study-type variability the choice of the prior distribution can be critical.
- For log-odds ratios, values for τ equal to 0.25, 0.5, 1 and 2 represent moderate, substantial, large, and very large heterogeneity.
- For example, $\exp(1.09\tau)$ is the median ratio of the maximum to the minimum of any random pair of odds ratios.
- Sensitivity analysis: we choose two half-normal priors for log-odds ratios with the following characteristics:

prior	median	95% interval
Half-normal(scale=0.5)	0.337	(0.016, 1.12)
Half-normal(scale=1.0)	0.674	(0.031, 2.24)

Coverages of confidence intervals (nominal level = 95%)

- Model A:

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	**98.9**	**99.2**	**98.7**	**97.0**	89.8	73.2
0.1	**99.1**	**99.2**	**98.8**	**96.9**	89.2	73.0
0.2	**98.9**	**99.0**	**98.7**	**96.7**	89.8	73.0
0.5	**97.9**	**97.7**	**97.5**	**95.4**	87.6	70.9
1	93.9	94.2	94.0	89.8	82.4	69.0
2	81.8	81.0	80.3	78.2	71.7	62.9

- Model B:

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	**99.5**	**99.6**	**99.3**	**98.2**	92.0	71.5
0.1	**99.5**	**99.6**	**99.3**	**97.9**	91.6	70.8
0.2	**99.4**	**99.5**	**99.2**	**98.1**	91.8	70.8
0.5	**98.9**	**98.4**	**98.3**	**97.0**	90.1	69.5
1	94.8	94.9	**95.3**	91.6	85.4	67.1
2	79.3	78.4	78.0	76.5	70.9	58.5

Lengths of confidence intervals (relative to RCT, in %)

- Model A (67% relative length means an effective sample size gain of $\approx 123\%$):

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	66.0	66.0	66.1	67.0	70.1	80.3
0.1	66.0	66.0	66.2	67.1	70.2	80.2
0.2	66.1	66.1	66.3	67.3	70.4	79.9
0.5	67.0	67.1	67.3	68.2	71.3	80.2
1	70.2	70.2	70.4	71.2	74.1	82.2
2	79.6	79.9	79.9	80.6	82.0	88.5

- Model B (80% relative length means an effective sample size gain of $\approx 56\%$):

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	79.1	79.3	79.4	79.1	79.9	80.3
0.1	79.4	79.5	79.5	79.5	79.8	80.1
0.2	79.5	79.4	79.4	79.5	79.7	80.1
0.5	79.6	79.5	79.1	79.7	79.9	79.7
1	79.4	79.9	79.9	79.7	80.3	79.9
2	80.0	80.1	80.0	80.2	80.0	79.3

Coverages (prior: $\tau \sim \text{HN}(\text{scale} = 1.0)$)

- Model A:

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	**100.0**	**100.0**	**100.0**	**99.9**	**98.5**	**96.2**
0.1	**100.0**	**100.0**	**99.9**	**99.7**	**98.8**	**95.5**
0.2	**100.0**	**99.9**	**100.0**	**99.7**	**98.6**	94.8
0.5	**100.0**	**99.8**	**99.6**	**99.6**	**97.6**	93.0
1	**99.7**	**99.7**	**99.5**	**98.2**	**95.1**	88.3
2	**98.1**	**98.4**	**98.0**	**95.3**	89.5	79.9

- Model B:

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	**100.0**	**100.0**	**100.0**	**100.0**	**99.2**	92.7
0.1	**100.0**	**100.0**	**100.0**	**99.8**	**99.1**	92.7
0.2	**100.0**	**100.0**	**100.0**	**99.9**	**98.6**	93.0
0.5	**99.9**	**99.9**	**100.0**	**99.9**	**98.1**	90.9
1	**99.8**	**99.7**	**99.7**	**99.1**	**95.8**	87.2
2	94.8	**96.2**	94.9	94.2	89.0	75.0

Lengths of intervals (prior: $\tau \sim \text{HN}(\text{scale} = 1.0)$)

- Model A:

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	94.9	94.7	94.9	97.2	103.5	118.8
0.1	95.0	94.8	94.9	97.3	102.6	118.6
0.2	95.0	94.8	95.5	97.4	103.5	118.4
0.5	96.7	97.0	97.1	99.4	104.3	119.8
1	102.5	102.7	103.4	104.1	109.5	121.6
2	118.9	119.1	118.6	119.1	123.2	132.4

- Model B:

τ_R \ τ_O	0	0.1	0.2	0.5	1	2
0	118.4	117.4	117.0	118.2	117.7	119.1
0.1	118.5	117.5	117.7	117.9	118.9	118.2
0.2	117.9	116.8	117.1	118.2	118.3	118.3
0.5	117.5	117.8	118.1	118.7	117.8	119.2
1	117.8	118.4	118.6	117.2	118.1	118.3
2	118.4	118.5	118.8	118.5	119.5	118.2

Summary and conclusions

- We have synthesized evidence from a single RCT and observational data in small populations.
- We presented two model frameworks within which evidence synthesis can be performed. Our simulation results indicate that framework A should be preferred over framework B.
- Recent computational advances in evidence synthesis facilitate the application of Bayesian hierarchical models.
- A meta-analysis of only two studies is a challenging problem, in particular the choice of a prior distribution for τ.
- Risk of bias due to lack of comparability of treatment groups or confounding.
- Adjustments for covariates can be done before the models are fitted.

Further work (some of it under way)

1. Simulations: consider a scenario, in which there is no correlation between the two observational arms, but a correlation between the randomized arms and open-label arm instead.
2. We will also consider continuous and time-to-event endpoints.
3. We estimated a pooled effect, θ^*. Other quantities of interest:
   - effect, θ_R, of an RCT in the light of observational data (shrinkage estimator),
   - effect, θ_{k+1}, of a future study (prediction / extrapolation).
4. Our frameworks bear some similarities to a comprehensive cohort-study design (Olschewski et al. (1992)). We may also want to consider a trial design with an additional observational open-label arm but no registry.

References – Applications


Baeten, D. et al. (2013):
Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial, *Lancet*, Vol. 382, pp. 1705-1713.


Varges, D. et al. (2017):
Doxycycline in early CJD - double-blinded randomized phase II and observational study, *Journal of Neurology, Neurosurgery, and Psychiatry*, Vol. 88, pp. 119-125.


Gross, O. et al. (2012a):
Safety and efficacy of the ACE-inhibitor ramipril in Alport syndrome: The double-blind, randomized, placebo-controlled, multicenter phase III EARLY PRO-TECT Alport trial in pediatric patients, *ISRN Pediatrics*, Volume 2012, Article ID 436046, 6 pages.


Gross, O. et al. (2012b):
Early angiotensin-converting enzyme inhibition in Alport syndrome delays renal failure and improves life expectancy, *Kidney International*, Vol. 81, pp. 494-501.

References – Methodology


Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004):
Bayesian Approaches to Clinical Trials and Health-Care Evaluation, Wiley.


Welton, N. J., Sutton, A. J., Cooper, N. J. et al. (2012):
Evidence Synthesis for Decision Making, Wiley.


Neuenschwander, B., Capkun-Niggli, G., Branson, M. et al. (2010):
Summarizing historical information on controls in clinical trials, *Clinical Trials*, Vol. 7, pp. 5-18.


Hampson, L. V., Whitehead, J., Eleftheriou, D. and Brogan, P. (2014):
Bayesian methods for the design and interpretation of clinical trials in very rare diseases, *Statistics in Medicine*, Vol. 33, pp. 4186-4201.


Unkel, S. et al.(2017):
A Bayesian hierarchical framework for evidence synthesis for a single randomized controlled trial and observational data in small populations, in preparation.


Röver, C. et al. (2017):
`bayesmeta`: Bayesian Random-Effects Meta-Analysis, R package version 1.4, `https://cran.r-project.org/package=bayesmeta`.

References – Relevant for further work


Saramago, P., Chuang, L.-H. and Soares, M.O. 2014:
Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data, *BMC Medical Research Methodology*, Vol. 14, No. 1, 105: DOI: 10.1186/1471-2288-14-105.


Lu, G., Ades, A.E., Sutton, A.J. et al. (2007):
Meta-analysis of mixed treatment comparisons at multiple follow-up times, *Statistics in Medicine*, Vol. 26, pp. 3681-3699.


Wandel, S., Neuenschwander, B., Röver, C. and Friede, T. (2017):
Using phase II data for the analysis of phase III studies: An application in rare diseases, *Clinical Trials*, [Epub ahead of print] doi: 10.1177/1740774517699409.


Olschewski, M., Schumacher, M. and Davis, K.B. (1992):
Analysis of randomized and nonrandomized patients in clinical trials using the comprehensive cohort follow-up study design, Controlled Clinical Trials, Vol. 13, pp. 226-239.


Schmoor, C., Caputo, A. and Schumacher, M. (2008):
Evidence from nonrandomized studies: a case study on the estimation of causal effects, *American Journal of Epidemiology*, Vol. 167, pp. 1120-1129.