Methods for the inclusion of real-world evidence in network meta-analysis

Chanatpon Aonnuam

Background

Background

- When assessing novel health technologies, randomized controlled trials (RCTs) have historically served as the gold standard, providing the foundation for meta-analyses in the evaluation of these advancements.
- There is an increasing emphasis on integrating real-world evidence (RWE) derived from observational studies, particularly in rare disease domains or circumstances where the design of RCTs proves to be formidable.

Background

- RWE can provide a substantial source of evidence, offering a more representative view of clinical practice.
- Bridging the gap between the efficacy observed in controlled trials and the effectiveness in real-life scenarios.

Background

- Various methods, such as naïve pooling, power transform prior approach, and hierarchical modeling have been employed to combine evidence from different sources.
- These methods, initially introduced in standard pairwise meta-analysis, were later generalized to network meta-analysis (NMA).

Objective

- The paper's objective is to explore the use of NMA to combine estimates from both RCTs and RWE, employing methods that differentiate between study designs to address potential biases in RWE.

Methods

Methods

- The methodology is applied to an illustrative example in relapsing-remitting multiple sclerosis (RRMS), where a systematic literature review identifies data sources from both RCTs and RWE on the effectiveness of disease-modifying therapies (DMTs) in RRMS patients.
- The results illustrate how these methodologies can be employed to combine data from different sources, providing insights into their impact on treatment effect estimates and associated uncertainty.

PICO

- P : RRMS patients
- I : Disease-modifying therapies (DMTs)
- C : Placebo
- O : Annualized relapse rate ratio (ARRR)

Illustrative example and sources of evidence

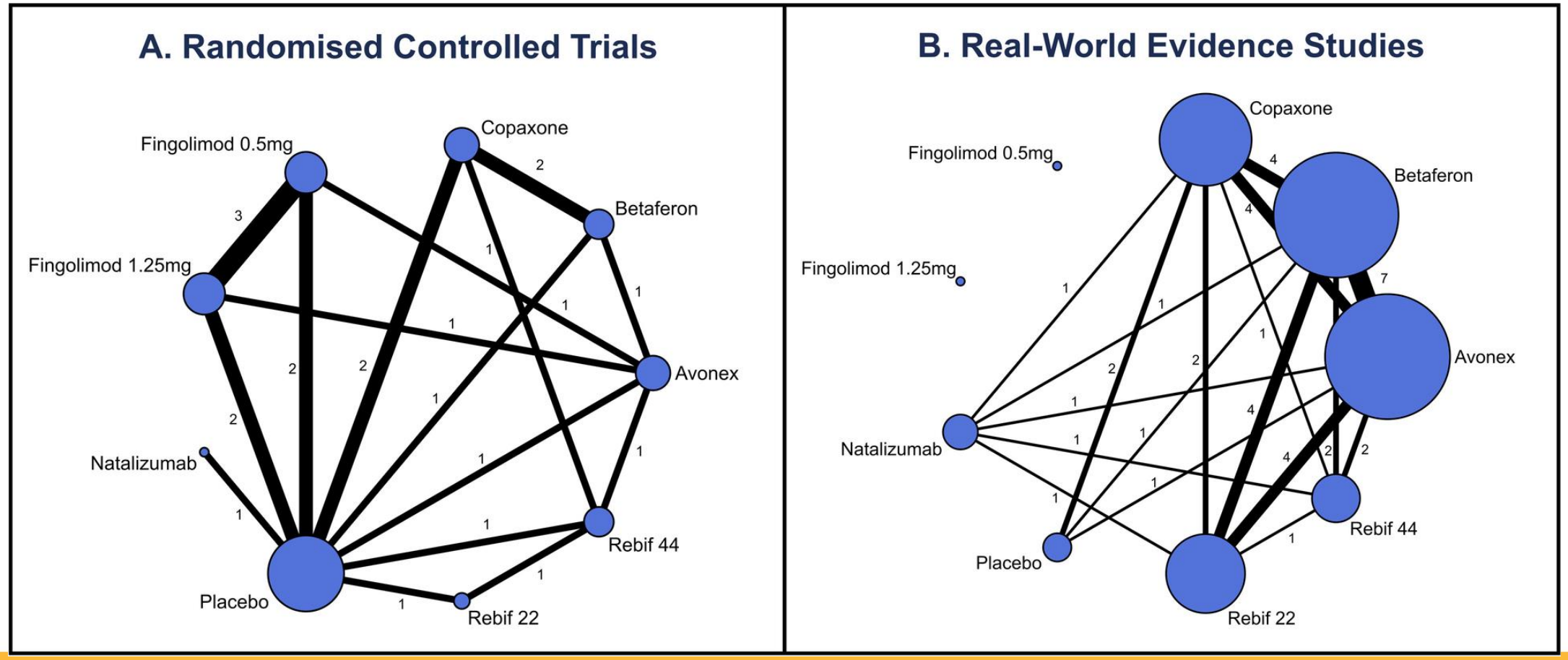
- In a motivating example, DMTs used in patients with RRMS were considered. A systematic review was carried out to identify studies, both randomised and observational, of different DMTs with a main focus on effectiveness of fingolimod to illustrate how the inclusion of RWE in NMA would impact the estimates of effectiveness of fingolimod in the context of a technology appraisal.

Illustrative example and sources of evidence

- The literature search was limited to studies reported prior to January 2010, when fingolimod was given licensing authorization.
- Data were extracted on the effect of each treatment on relapse rate

Network structure

Figure 1 illustrates the network diagram of direct comparisons between interventions in both the RWE and RCT data.



Network structure

- The nodes represent individual interventions analyzed and the interconnecting lines represent the direct comparisons between interventions.
- The numbers along the lines represent the number of studies for each comparison in either the RCTs or RWE.

Network structure

- In total there were 23 studies included, 14 of them being RCTs and 9 of them being the RWE studies.
- This example the average sample size in each arm for the RWE was 186 participants, compared to the 288 participants in the RCT arms.

Network meta-analysis

- In the Network Meta-Analysis (NMA), a random-effects model with adjustment for multi-arm trials was used.
- Assuming consistency in the network, treatment effects for each contrast were represented as differences of basic parameters.
- A Bayesian approach with prior distributions on model parameters, such as baseline study effects , basic, and between-study variance was adopted.
- For multi-arm studies, correlation between treatment effects relative to a common baseline treatment was considered.

Naïve pooling approach

- The above NMA model was initially used to combine data from RCTs with RWE by including the observational studies at ‘face-value’.
- Data from all studies, regardless of the study design, were combined in the NMA described above.

Naive pooling using standard NMA

Table 1 Matrix table of annualised relapse rate ratios (95% credible intervals) for network meta-analysis (NMA) using naïve pooling random-effects models

Treatment	Placebo	Natalizumab	Fingolimod 1.25	Fingolimod 0.5	Avonex	Rebif 22	Rebif 44	Copaxone	Betaferon
Placebo		0.41 (0.29, 0.57)	0.46 (0.35, 0.59)	0.42 (0.32, 0.53)	0.78 (0.65, 0.94)	0.77 (0.61, 0.95)	0.75 (0.60, 0.93)	0.60 (0.50, 0.71)	0.70 (0.58, 0.84)
Natalizumab	0.32 (0.26, 0.38)		1.16 (0.74, 1.68)	1.05 (0.67, 1.52)	1.99 (1.34, 2.74)	1.95 (1.30, 2.72)	1.90 (1.28, 2.64)	1.53 (1.02, 2.11)	1.78 (1.20, 2.46)
Fingolimod 1.25	0.46 (0.40, 0.54)	1.48 (1.15, 1.89)		0.91 (0.69, 1.17)	1.73 (1.30, 2.27)	1.70 (1.22, 2.32)	1.66 (1.20, 2.26)	1.33 (0.97, 1.77)	1.56 (1.14, 2.07)
Fingolimod 0.5	0.42 (0.36, 0.49)	1.36 (1.04, 1.37)	0.92 (0.77, 1.08)		1.92 (1.44, 2.52)	1.88 (1.35, 2.57)	1.84 (1.33, 2.51)	1.48 (1.08, 1.96)	1.72 (1.27, 2.29)
Avonex	0.83 (0.72, 0.96)	2.67 (2.61, 3.38)	1.81 (1.50, 2.16)	1.98 (1.64, 2.38)		0.98 (0.80, 1.20)	0.96 (0.78, 1.18)	0.77 (0.63, 0.93)	0.90 (0.76, 1.05)
Rebif 22	0.72 (0.60, 0.86)	2.32 (1.75, 2.99)	1.57 (1.24, 1.97)	1.72 (1.35, 2.16)	0.87 (0.70, 1.08)		0.99 (0.78, 1.23)	0.79 (0.62, 0.98)	0.92 (0.74, 1.12)
Rebif 44	0.68 (0.59, 0.78)	2.18 (1.69, 2.75)	1.48 (1.21, 1.80)	1.62 (1.32, 1.97)	0.82 (0.69, 0.96)	0.95 (0.78, 1.14)		0.81 (0.64, 0.99)	0.94 (0.75, 1.15)
Copaxone	0.65 (0.57, 0.75)	2.09 (1.62, 2.65)	1.42 (1.16, 1.73)	1.56 (1.27, 1.90)	0.79 (0.66, 0.94)	0.91 (0.73, 1.12)	0.96 (0.82, 1.13)		1.17 (0.98, 1.41)
Betaferon	0.67 (0.57, 0.77)	2.15 (1.65, 2.71)	1.45 (1.18, 1.77)	1.59 (1.29, 1.96)	0.81 (0.67, 0.95)	0.93 (0.74, 1.16)	0.99 (0.82, 1.17)	1.03 (0.89, 1.17)	

- Combining evidence from both RCTs and real-world data in NMA can elevate uncertainty levels.
- In the comparison of fingolimod 0.5 mg with Avonex, the 95% credible interval of ARRR widened from (1.64 to 2.38) with only RCT data to (1.44 to 2.52) with combined data.
- This increased uncertainty is likely attributed to heightened between-study heterogeneity when integrating evidence from both sources.

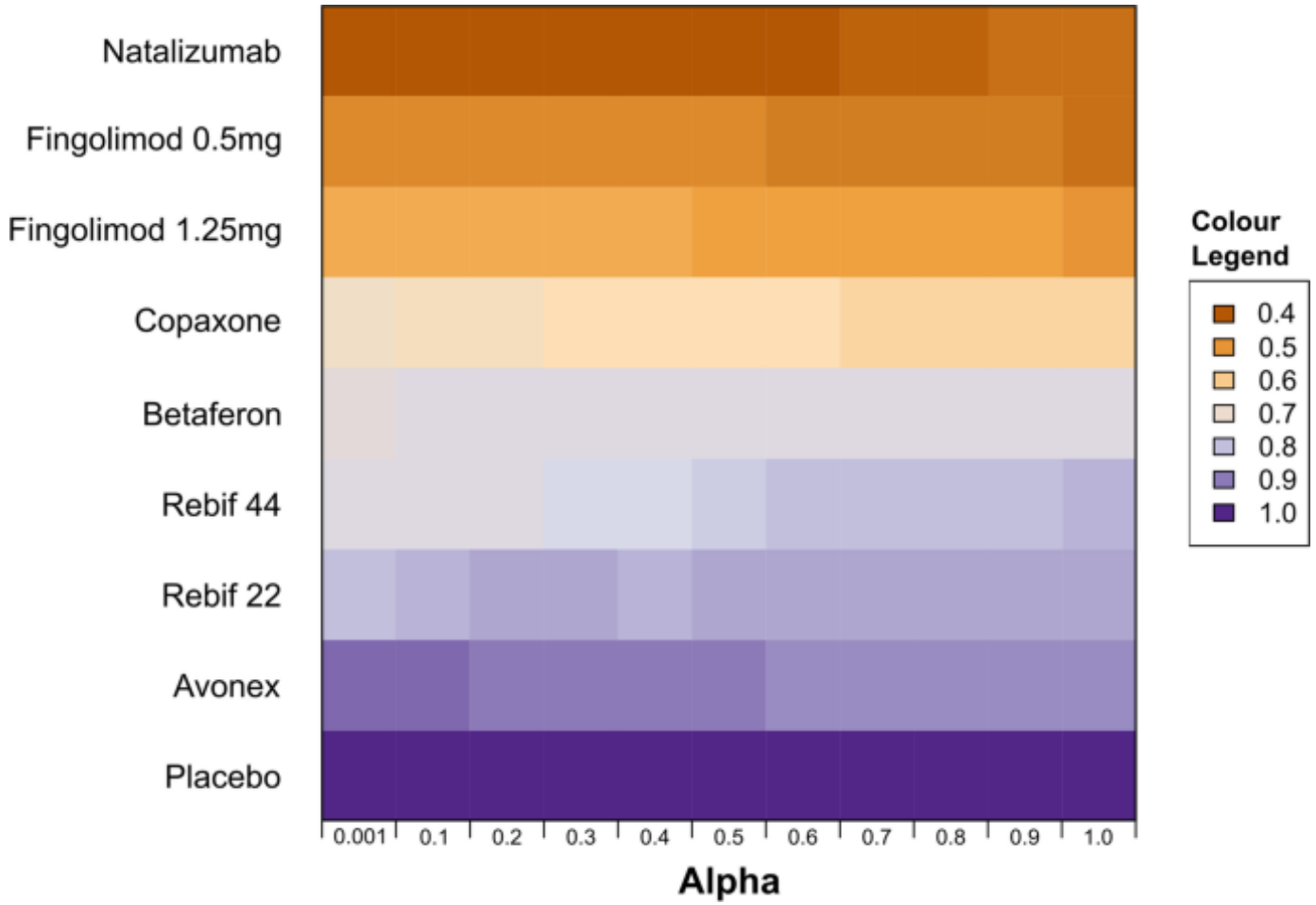
Power prior approach

- To address study design differences between RCTs and observational studies, a 'power transform prior' approach was used, introducing a down-weighting factor (alpha) to adjust the contribution of RWE in the NMA.
- Varying alpha from zero to one allows different levels of discounting RWE.
- The overall joint posterior distribution, considering the ARRRs, is then determined by combining the likelihood contributions of RWE and RCT data.
- This approach provides flexibility in weighing the impact of RWE on the analysis.

Table. Annualised relapse rate ratios (95% credible intervals) of each active treatment compared to placebo for values alpha using the power prior model with between study heterogeneity standard deviation estimates

Alpha	Natalizumab	Fingolimod 1.25mg	Fingolimod 0.5mg	Avonex	Rebif 22	Rebif 44	Copaxone	Betaferon	Between study SD
0.001	0.32 (0.26, 0.39)	0.46 (0.40, 0.54)	0.42 (0.36, 0.50)	0.83 (0.72, 0.95)	0.71 (0.60, 0.86)	0.68 (0.59, 0.78)	0.65 (0.57, 0.75)	0.67 (0.58, 0.77)	0.055
0.1	0.32 (0.27, 0.39)	0.46 (0.40, 0.53)	0.42 (0.36, 0.48)	0.79 (0.71, 0.90)	0.73 (0.64, 0.85)	0.69 (0.61, 0.78)	0.66 (0.58, 0.75)	0.68 (0.60, 0.78)	0.045
0.2	0.33 (0.27, 0.41)	0.46 (0.39, 0.53)	0.42 (0.36, 0.50)	0.78 (0.70, 0.88)	0.72 (0.62, 0.84)	0.69 (0.61, 0.79)	0.65 (0.57, 0.73)	0.68 (0.61, 0.77)	0.047
0.3	0.33 (0.27, 0.42)	0.45 (0.39, 0.54)	0.42 (0.35, 0.49)	0.77 (0.67, 0.88)	0.73 (0.64, 0.84)	0.70 (0.62, 0.81)	0.65 (0.58, 0.74)	0.69 (0.60, 0.78)	0.057
0.4	0.34 (0.27, 0.43)	0.45 (0.38, 0.54)	0.41 (0.35, 0.50)	0.77 (0.68, 0.89)	0.74 (0.63, 0.87)	0.72 (0.62, 0.84)	0.65 (0.56, 0.74)	0.69 (0.61, 0.80)	0.085
0.5	0.35 (0.28, 0.46)	0.46 (0.38, 0.55)	0.42 (0.34, 0.50)	0.78 (0.68, 0.90)	0.75 (0.63, 0.89)	0.72 (0.62, 0.85)	0.64 (0.55, 0.73)	0.70 (0.61, 0.81)	0.100
0.6	0.37 (0.29, 0.50)	0.46 (0.37, 0.57)	0.41 (0.33, 0.51)	0.78 (0.67, 0.92)	0.75 (0.62, 0.91)	0.73 (0.61, 0.88)	0.63 (0.53, 0.73)	0.70 (0.59, 0.82)	0.131
0.7	0.38 (0.29, 0.53)	0.46 (0.36, 0.57)	0.42 (0.33, 0.52)	0.78 (0.67, 0.92)	0.75 (0.62, 0.93)	0.73 (0.61, 0.89)	0.62 (0.52, 0.72)	0.70 (0.59, 0.82)	0.144
0.8	0.39 (0.29, 0.54)	0.46 (0.36, 0.58)	0.41 (0.32, 0.53)	0.78 (0.66, 0.93)	0.76 (0.62, 0.94)	0.74 (0.61, 0.91)	0.61 (0.51, 0.72)	0.70 (0.58, 0.83)	0.162
0.9	0.40 (0.30, 0.56)	0.46 (0.35, 0.59)	0.41 (0.32, 0.53)	0.78 (0.65, 0.94)	0.76 (0.61, 0.95)	0.74 (0.60, 0.92)	0.61 (0.51, 0.72)	0.70 (0.58, 0.83)	0.173
1.0	0.41 (0.30, 0.57)	0.45 (0.35, 0.59)	0.41 (0.32, 0.53)	0.78 (0.65, 0.93)	0.76 (0.61, 0.95)	0.74 (0.60, 0.93)	0.60 (0.49, 0.71)	0.69 (0.57, 0.83)	0.182

Fig.3 Heat map displaying rankings for each treatment (based on absolute annualized relapse rates) for values of the down-weighting factor (alpha) using 'power prior' model. Orange represents highest ranking and purple represents lowest ranking



Hierarchical model approach

- In a Bayesian hierarchical model for NMA, adapting Schmitz et al., treatment effects from RCTs and RWE are considered exchangeable.
- The model separates RCT and RWE data, using a Poisson distribution.
- A power prior approach at the within-study level for RWE allows sensitivity analysis by down-weighting its contribution with the factor alpha.
- This combines RCT and RWE effects, providing an overall pooled ARRR combined effect.

Hierarchical model and hierarchical power prior model

- Table 2 presents outcomes from a hierarchical NMA incorporating an additional hierarchy for study design.
- While point estimates align broadly with the simpler 'power transform approach,' the hierarchical model generally exhibits greater uncertainty, seen in wider credible intervals.

Hierarchical model and hierarchical power prior model

- Table 2 Matrix table of annualised relapse rate ratios (95% credible intervals) for NMA using hierarchical models including randomised controlled trials and real-world evidence

Treatment	Placebo	Natalizumab	Fingolimod 1.25	Fingolimod 0.5	Avonex	Rebif 22	Rebif 44	Copaxone	Betaferon
Placebo		0.40 (0.26, 0.70)	0.46 (0.40, 0.54)	0.42 (0.36, 0.49)	0.83 (0.55, 1.26)	0.78 (0.51, 1.21)	0.79 (0.53, 1.27)	0.62 (0.40, 0.92)	0.72 (0.48, 1.13)
Natalizumab	0.35 (0.14, 0.74)		1.22 (0.65, 1.80)	1.12 (0.60, 1.65)	2.18 (1.04, 3.51)	2.04 (0.99, 3.36)	2.06 (1.04, 3.50)	1.62 (0.75, 2.60)	1.90 (0.93, 3.14)
Fingolimod 1.25	0.46 (0.40, 0.54)	1.76 (0.61, 3.40)		0.92 (0.76, 1.09)	1.71 (1.16, 2.79)	1.61 (1.09, 2.66)	1.62 (1.12, 2.80)	1.28 (0.84, 2.02)	1.50 (1.02, 2.51)
Fingolimod 0.5	0.42 (0.36, 0.49)	1.60 (0.57, 3.13)	0.92 (0.77, 1.07)		1.87 (1.27, 3.04)	1.76 (1.18, 2.91)	1.77 (1.22, 3.05)	1.39 (0.93, 2.21)	1.64 (1.11, 2.73)
Avonex	0.88 (0.44, 1.60)	3.39 (0.99, 7.53)	1.70 (0.96, 3.48)	1.86 (1.05, 3.79)		0.98 (0.54, 1.67)	0.99 (0.56, 1.76)	0.71 (0.41, 1.29)	0.84 (0.51, 1.56)
Rebif 22	0.79 (0.39, 1.56)	3.22 (0.92, 6.58)	1.50 (0.83, 3.43)	1.64 (0.91, 3.73)	1.00 (0.40, 2.11)		1.06 (0.59, 1.85)	0.76 (0.43, 1.36)	0.89 (0.53, 1.65)
Rebif 44	0.76 (0.40, 1.50)	3.17 (0.86, 6.86)	1.46 (0.86, 3.28)	1.60 (0.93, 3.57)	0.97 (0.38, 2.16)	1.09 (0.41, 2.38)		0.75 (0.41, 1.31)	0.88 (0.50, 1.59)
Copaxone	0.68 (0.34, 1.22)	2.98 (0.73, 5.65)	1.32 (0.72, 2.68)	1.45 (0.79, 2.94)	0.69 (0.31, 1.82)	0.76 (0.33, 2.01)	0.79 (0.33, 2.10)		1.23 (0.70, 2.16)
Betaferon	0.74 (0.39, 1.41)	2.91 (0.87, 6.49)	1.43 (0.84, 3.08)	1.56 (0.91, 3.39)	0.75 (0.38, 1.99)	0.84 (0.40, 2.25)	0.86 (0.40, 2.30)	1.23 (0.49, 2.74)	

hierarchical power prior model

- For instance, comparing natalizumab to placebo, the ARRR was 0.41 (0.30, 0.57) with the power prior approach , whereas the hierarchical model showed 0.40 (0.26, 0.70).
- The hierarchical model explicitly considers study design differences, introducing additional variability.
- The inclusion of RWE in the hierarchical model had no impact on fingolimod's effectiveness estimate, reflecting the absence of RWE for this treatment and the model's flexibility in accommodating additional variability.

Discussion

- The study recognizes differences between RCTs and RWE studies. Unlike a simple over- or underestimation of treatment effects, this research reveals a nuanced pattern with both over- and underestimations for different treatments, aligning with previous findings.

Discussion

- The research extends methods introduced by Schmitz et al. (2013), adapting them for count data with the Poisson likelihood.
- The hierarchical model is also expanded to down-weight observational studies using a modified power prior approach.

Discussion

- While both models are deemed useful for addressing heterogeneity between study designs and potential RWE bias, the study's results do not significantly differ from naïve pooling or basic power transform prior approaches in the specific example.

Discussion

- Wider credible intervals are attributed to increased between-study design heterogeneity when including RWE. Although hierarchical models are considered more appropriate, caution is advised, and a sensitivity analysis comparing results with naïve pooling is recommended.

Discussion

- In the illustrative example, the inclusion of RWE heightens overall uncertainty in treatment effects, supporting prior findings.
- Greater heterogeneity across RWE studies, particularly when assessing the effectiveness of fingolimod 0.5 mg in the general population, contributes to increased uncertainty in the combined analysis compared to RCTs.

Discussion

- The study suggests further evaluation of these methods in various settings, including simulation studies.
- Extending the hierarchical modeling approach to accommodate different types of RWE may mitigate potential uncertainty increases due to broader evidence bases.

Discussion

- For decision-makers, the methods offer assessments on a larger evidence base, encompassing diverse patient demographics and clinical characteristics.
- While the inclusion of RWE can enhance decision-making evidence, analysts and decision-makers need to evaluate RWE credibility on a case-by-case basis, determine the acceptability of the analysis type, and interpret and use results appropriately.

Limitations

- Sample Size Discrepancy
  - Smaller sample sizes in RWE studies, compared to larger RCTs, might impact the uncertainty of effect estimates when weighting studies.
- Single Illustrative Example
  - The study is based on a single illustrative example, and results could differ in other clinical areas. Nonetheless, it underscores the importance of comparing the combined RCT and RWE data analysis to traditional NMA of RCT data alone to understand the effectiveness vs. efficacy gap.

Limitations

- Likelihood Model Choice
  - The use of a Poisson likelihood for data analysis might introduce increased uncertainty, which could potentially be reduced by employing a negative binomial likelihood to account for over-dispersion in modeling count data.
- Absence of Meta-regression
  - The study does not include meta-regression, which could explain some between-study heterogeneity, but its effectiveness may be limited by available covariate information and the number of studies in the NMA.

Limitations

- Aggregate-level Data:
  - The NMAs in the example use aggregate-level data only and obtaining individual patient data (IPD) from RWE studies could improve adjustments for potential allocation bias, thereby reducing between-study heterogeneity and uncertainty in pooled effectiveness estimates.
  - However, obtaining IPD from observational studies can be challenging due to data-sharing regulations.

Conclusions

- The 'power transform prior' and hierarchical models in NMA had minimal impact on ARRR effect estimates.
- The degree of RWE inclusion in NMAs significantly increased uncertainty around effect estimates due to heightened between-study heterogeneity.
- Hierarchical NMA models added further uncertainty, considering different study types (RCTs and RWE).

Conclusions

- A comprehensive simulation study is recommended to assess the models' ability to accurately estimate treatment effects and address biases introduced by RWE in various scenarios.
- RWE is valuable for HTA decision-making, particularly in rare diseases with limited clinical trial data.

Conclusions

- RWE inclusion in meta-analysis can inform clinical development planning, potentially influencing future trial design and reducing required patient numbers.

Thank you