Evidence-synthesis methods for personalizing the choice of treatment

2nd Year Examination, Graduate School for Health Sciences, University of Bern

Michael Seo[1,2], Orestis Efthimiou[1,3]

May 10, 2021

[1]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[2]Graduate School for Health Sciences, University of Bern, Bern, Switzerland

[3]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

Overview

- First paper: Comparing methods for estimating patient-specific treatment effects in individual patient data meta-analysis
- Second paper: Combining individual patient data from randomized and non-randomized studies to predict real-world effectiveness of interventions
- Side project: The Kilim plot: A tool for visualizing network meta-analysis results for multiple outcomes
- Future research: Developing prediction models when there are systematically missing predictors in an individual patient data meta-analysis

First paper

Received: 2 April 2020 | Revised: 28 September 2020 | Accepted: 7 December 2020

DOI: 10.1002/sim.8859

RESEARCH ARTICLE

Statistics in Medicine WILEY

Comparing methods for estimating patient-specific treatment effects in individual patient data meta-analysis

Michael Seo[1,2] | Ian R. White[3] | Toshi A. Furukawa[4] | Hissei Imai[4] | Marco Valgimigli[5] | Matthias Egger[1] | Marcel Zwahlen[1] | Orestis Efthimiou[1]

Compared standard approach to individual patient data (IPD) meta-analysis with six alternative methods: stepwise regression and five regression methods that perform shrinkage

Background

- Meta-analysis of individual patient data (IPD) is increasingly used to synthesize data from multiple trials.
- IPD meta-anaysis offers several advantages as compared to meta-analysing aggregated data, such as a better capacity to **individualize treatment recommendations**.
- For this paper, we utilize one-stage approach to IPD meta-analysis as they are generally considered more flexible.

IPD meta-analysis

One-stage approach

For patient i randomized in study j to receive treatment t_{ij}, we assume the outcome is categorical

$$y_{ij} \sim Bernoulli(p_{ij})$$

where the linear predictor on the log-odds scale is

$$log(\frac{p_{ij}}{1-p_{ij}}) = \alpha_j + \beta \boldsymbol{x_{ij}} + \gamma \boldsymbol{x_{ij}} t_{ij} + d_j t_{ij}$$

$$d_j \sim Normal(\delta, \tau^2)$$

where α_j is study-specific intercept, β are main effects of covariates $\boldsymbol{x_{ij}}$, γ are coefficients for treatment-covariate interactions, and d_j is the treatment effect which we assume to have random effects with average treatment effect δ and heterogeneity τ^2

Patient-specific treatment effects

$$log(\frac{p_{ij}}{1-p_{ij}}) = \alpha_j + \beta \boldsymbol{x_{ij}} + \boldsymbol{\gamma x_{ij}} t_{ij} + d_j t_{ij}$$

$$d_j \sim Normal(\delta, \tau^2)$$

The usual goal of an IPD meta-analysis is to estimate the average treatment effect δ and to identify important treatment-covariate interactions γ. In this paper, we focus on patient-specific treatment effects which we define as

$$\gamma \boldsymbol{x_{ij}} + \delta$$

We explore whether the **selection/shrinkage** of treatment-covariate interactions (i.e. **effect modifiers**) in an IPD meta-analysis can lead to better estimates of patient-specific treatment effects.

Variable selection and shrinkage methods

1. Stepwise variable selection: bidirectional stepwise regression considers both adding and removing one **effect modifier** at each step, and takes the best option according to the criterion such as AIC

2. LASSO: uses L_1 penalty term in the optimization function. The penalty term leads to a shrinkage of the effect modification

$$\min_{\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta} -l(\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta) + \lambda\Big[\sum_{m=1}^{p} \mid \gamma_m \mid \Big]$$

where $l(\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta)$ is the log binomial likelihood

3. Ridge regression: uses L_2 penalty term instead

$$\min_{\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta} -l(\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta) + \lambda\Big[\sum_{m=1}^{p} \gamma_m^2\Big]$$

Variable selection and shrinkage methods (2)

4. Adaptive LASSO: adaptive weights are used for penalizing different coefficients in the L_1 penalty

$$\min_{\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta} -l(\alpha_j,\boldsymbol{\beta},\boldsymbol{\gamma},\delta) + \lambda\Big[\sum_{m=1}^{p} \hat{w}_m \mid \gamma_m \mid \Big]$$

where $\hat{w}_m = 1/\hat{\gamma}_m$ and $\hat{\gamma}_m$ were obtained from fitting ridge regression

5. Bayesian LASSO: penalizes effect modifiers using a Laplace prior distribution. We assign a vague prior distribution on the scale parameter of the Laplace prior distribution

$$\pi(\boldsymbol{\gamma}) = \prod_{m=1}^{p} \frac{\lambda}{2} e^{-\lambda|\gamma_m|}$$

$$\lambda^{-1} \sim \mathit{Uniform}(0,5)$$

Variable selection and shrinkage methods (3)

6. Stochastic search variable selection (SSVS): uses a mixture prior on γ_m

$$\pi(\gamma_m \mid I_m) = (1 - I_m)N(0, \eta^2) + I_m N(0, g\eta^2)$$

$$\eta \sim \mathit{Uniform}(0, 5),\ g = 100$$

where $I_m = 1$ indicate presence of the interaction term of covariate m in the model and η^2 is the variance when the interaction term is absent

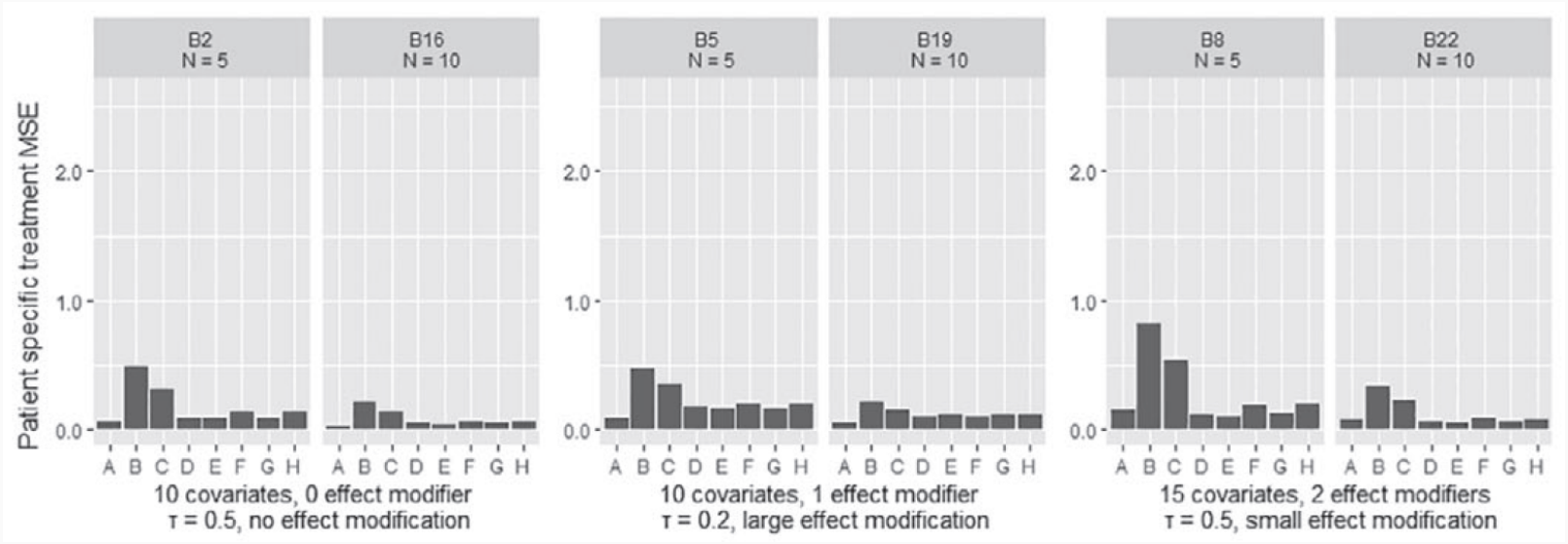
Simulations

- We explored different configurations regarding the number of covariates, the number of included studies, and the magnitude of effect modifiers.
- Our main performance metric was mean squared error (MSE) of the patient-specific treatment effect, averaged over all simulated patients

$$\frac{1}{n}\sum_{i=1}^{n}\left[(\hat{\gamma}\boldsymbol{x}_{ij}^{\boldsymbol{em}}+\hat{\delta})-(\gamma\boldsymbol{x}_{ij}^{\boldsymbol{EM}}+\delta)\right]^2$$

where $\boldsymbol{x}_{ij}^{\boldsymbol{em}}$ are the effect modifiers that the model identified and $\boldsymbol{x}_{ij}^{\boldsymbol{EM}}$ are the true effect modifiers.

Simulation results



- A is the true model used in data generation; B is the standard IPD meta-analysis model; C is the stepwise regression; D-H are shrinkage methods described
- In most scenarios, shrinkage methods gave lower MSE of the patient-specific treatment effect as compared with the standard approach and stepwise regression.

Real clinical example: Drug-eluting or bare-metal stents for percutaneous coronary intervention

- IPD from 8 RCTs. 11,133 patients who have undergone percutaneous coronary intervention for coronary artery disease
- Patients randomized to receive either drug-eluting versus bare metal stents. The outcome is cardiac death or myocardial infarction at 1-year after randomization.
- 9 covariates including age, number of implanted stents, gender, stent placement in the left anterior descending artery, etc

Real clinical example results

TABLE 3 Estimated treatment effect (odds ratio and 95% CI) for different subgroup population in Stent dataset

Subgroups	GLMM-full (95% CI)	STEP (95% CI)	LASSO (95% CI*)	ridge (95% CI*)	Adaptive LASSO (95% CI*)	Bayesian LASSO (95% CrI)	SSVS (95% CrI)
80 years old, female, diabetes, unstable CAD, multivessel disease, ladtreated, overlapping stents, mean stent diameter < 3 mm, and 5 stents	0.28 (0.11, 0.73)	0.52 (0.32, 0.86)	0.63 (0.17, 0.98)	0.43 (0.18, 0.94)	0.42 (0.15, 0.97)	0.54 (0.24, 1.18)	0.55 (0.24, 1.27)
50 years old, male, no diabetes, stable CAD, no multivessel disease, no ladtreated, no overlapping stents, mean stent diameter ≥ 3 mm, and 1 stent	1.84 (0.98, 3.45)	1.24 (0.94, 1.63)	1.05 (0.80, 2.61)	1.32 (0.84, 2.27)	1.26 (0.84, 2.52)	1.22 (0.75, 2.00)	1.24 (0.73, 2.11)

Note: All other abbreviations as per Table 2. CI, confidence interval; CrI, credible interval. CI denoted with a star (*) were estimated using bootstrap. An odds ratio smaller than 1 favors the drug eluding stents.

- Treatment should be assigned to the first group rather than to the second.
- GLMM-full gave large estimates of patient-specific treatment effects; they are probably an overestimation of the true treatment effects.

Conclusion

- We recommend that future IPD meta-analysis that aim to estimate patient-specific treatment effects using multiple effect modifiers should use **shrinkage methods**, whereas stepwise regression should be avoided.
- We recommend the use of Bayesian shrinkage models over the frequentist ones. The Bayesian models allow for greater flexibility, for example in setting up random effect structures, and allow incorporating prior information regarding some of the model's parameters.

Second paper

Combining individual patient data from randomized and non-randomized studies to predict real-world effectiveness of interventions

Michael Seo[1,2 *], Thomas P.A. Debray[3,4], Yann Ruffieux[1], Sandro Gsteiger[5], Sylwia Bujkiewicz[6], Axel Finckh[7], Matthias Egger[1,8], Orestis Efthimiou[1,9]

- Developed models that combine individual patient data from randomized controlled trials and non-randomized studies when aiming to predict real-world outcomes for a set of treatments

Background

- There has been growing interest in using data from non-randomized studies (NRS) to complement evidence from randomized controlled trials (RCTs) in medical decision-making.
- RCTs are the best source of evidence regarding relative treatment effects but they often employ strict experimental settings, which may hamper their ability to predict outcomes in **'real-world' clinical settings** (i.e. NRS).
- In this paper, we focus on two-stage approach to IPD meta-analysis. Two-stage approach is often necessary in practice due to data restrictions, e.g. when data reside on different servers.

Single NRS model - Approach I

- We can build a prediction model using a single NRS that is representative of the population we want to predict and disregard all others sources of information.
- If we assume that the study had only two treatment arms, W and A (the reference)

$$y_i \sim N(m_i, \sigma^2)$$

$$m_i = \alpha + \boldsymbol{\beta x_i} + \boldsymbol{\gamma x_i} t_i + \delta t_i$$

or equivalently,

$$m_i = \alpha + \boldsymbol{\beta x_i} + \boldsymbol{\gamma_{WA} x_i} + \delta_{WA}$$

Single NRS model - Approach I (2)

- The reason why we introduce different notation is because in this paper, we are considering multiple treatments.
- If the study had three treatment arms, Z, W, and A,

$$y_i \sim N(m_i, \sigma^2)$$

$$m_i = \alpha + \beta \boldsymbol{x_i} + \boldsymbol{\gamma_{WA} x_i} + \boldsymbol{\gamma_{ZA} x_i} + \delta_{WA} + \delta_{ZA}$$

- But for simplicity of our notation, we will assume that studies have only two treatment arms, W and A.

Design-naive IPD meta-analysis - Approach II

First stage

For each study j, the same model described in Single NRS model is fitted (now indexed with study j).

$$y_{ij} \sim N(m_{ij}, \sigma_j^2)$$

$$m_{ij} = a_j + \boldsymbol{b_j x_{ij}} + \boldsymbol{c_{j,WA} x_{ij}} + d_{j,WA}$$

We define

$$\hat{\theta}_j = (\hat{a}_j, \hat{\boldsymbol{b}}_{\boldsymbol{j}}, \hat{\boldsymbol{c}}_{\boldsymbol{j,WA}}, \hat{d}_{j,WA})$$

and the corresponding variance-covariance matrix $\hat{\boldsymbol{S}}_{\boldsymbol{j}}$.

Design-naive IPD meta-analysis - Approach II (2)

Second stage

These estimates, $\hat{\theta}_j$ and $\hat{\boldsymbol{S}}_{\boldsymbol{j}}$, contribute to the likelihood of the second stage of the meta-analysis. Assuming random effect for $\theta^{(\delta)}_{j,WA}$,

$$\hat{\theta}_j \sim N((\alpha, \beta, \boldsymbol{\gamma_{WA}}, \theta^{(\delta)}_{j,WA})^T, \hat{\boldsymbol{S}}_{\boldsymbol{j}})$$

$$\theta^{(\delta)}_{j,WA} \sim N(\delta_{WA}, \tau^2)$$

It may be difficult to estimate τ^2, the heterogeneity of the treatment effect, when only a few studies are available. A further simplication would be to assume fixed treatment effect.

$$\hat{\theta}_j \sim N((\alpha, \beta, \boldsymbol{\gamma_{WA}}, \delta_{WA})^T, \hat{\boldsymbol{S}}_{\boldsymbol{j}})$$

We use the Bayesian framework and assume vague priors for all the parameters.

Design-adjusted IPD meta-analysis - Approach III

- The variance of the estimates of treatment effects and effect modification obtained from NRS is inflated by dividing by a factor w_j, with $0 < w_j < 1$. By doing so, we effectively decrease the impact of NRS in the estimation of all relative treatment effects.

$$(\hat{\boldsymbol{c}}_{\boldsymbol{j},\boldsymbol{WA}}, \hat{d}_{j,WA})^T \sim \begin{cases} N\left((\boldsymbol{\gamma}_{\boldsymbol{WA}}, \delta_{WA})^T, \frac{\hat{\boldsymbol{S}}_{\boldsymbol{j}}}{w_j}\right) & \textit{if study } j \textit{ is a NRS} \\ N\left((\boldsymbol{\gamma}_{\boldsymbol{WA}}, \delta_{WA})^T, \hat{\boldsymbol{S}}_{\boldsymbol{j}}\right) & \textit{if study } j \textit{ is a RCT} \end{cases}$$

- The motivation is that RCTs are the most reliable sources of information for relative treatment effects because randomization helps us avoid issues related to confounding.

Design-adjusted IPD meta-analysis - Approach III (2)

- A range of values for weights can be used and model performance measures can be assessed to decide on the optimal weights.
- Moreover, since we aim to predict outcomes for NRS population, estimates for the model's intercept and the main effects of covariates are aggregated only using NRS.

$$(\hat{a}_j, \hat{\boldsymbol{b}}_{\boldsymbol{j}})^T \sim N\Big((\alpha, \beta)^T, \hat{\boldsymbol{S}}_{\boldsymbol{j}}\Big), \textit{ if study j is a NRS}$$

Possible extensions: shrinkage

- Shrinkage methods in general are known to improve prediction accuracy.
- We penalize the coefficients of the effect modification using Bayesian LASSO at the **first stage** and aggregate penalized coefficients at the second stage.
- The prior for effect modifiers of treatment W vs. the reference A is given by:

$$\pi(\boldsymbol{\gamma_{WA}}) = \prod_{k=1}^{n_{cov}} \frac{\lambda}{2} e^{-\lambda|\gamma_{k,WA}|}$$

$$\lambda^{-1} \sim Uniform(0, 5)$$

Possible extensions: calibration

- After second stage is used to aggregate parameters for each IPD meta-analysis, we **replace** α and β with estimates from a single NRS model that reflects the target population (i.e. Approach I).
- A motivation for this is that data sampled from the patient population of interest might be the best source of evidence for predicting the **reference treatment outcome**.
- Conversely, estimates of the intercept term obtained from RCTs might be less representative for the target population because RCTs are performed in a highly controlled settings.

Prediction equation

For each approaches, use the estimated/aggregated parameters, $\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma_{WA}}, \delta_{WA}$, to make predictions for a new patient with covariates $\boldsymbol{x}$ and treatment t using:

$$y_{pred}(\boldsymbol{x}, t) = \alpha + \boldsymbol{\beta x} + \boldsymbol{\gamma_{tA} x} + \delta_{tA}$$

or equivalently,

$$y_{pred}(\boldsymbol{x}, t) = \begin{cases} \alpha + \boldsymbol{\beta x} & \text{if } t = A \\ \alpha + \boldsymbol{\beta x} + \boldsymbol{\gamma_{tA} x} + \delta_{tA} & \text{if } t \neq A \end{cases}$$

Summary of approaches used

Approach	Description	Prediction for new patient with covariates x, treatment t	Meta-analysis	α, β in prediction model estimated from:	γ, δ in prediction model estimated from:	Shrinkage at 1st stage
I	We only use a single NRS.	$y_{pred}(x,t) = a + \beta x + \gamma_{tA} x + \delta_{tA}$	✗	A single NRS	A single NRS	✓
IIa	**1st stage**: fit a model in each study separately. **2nd stage**: fit a design-naïve NMA.		✓	All RCTs & NRS	All RCTs & NRS	✗
IIb	Same as IIa, but at 1st stage we use penalized estimation.		✓	All RCTs & NRS	All RCTs & NRS	✓
IIc	Same as IIb, but intercept and main effects (α, β) are estimated from a single NRS.		✓	A single NRS	All RCTs & NRS	✓
IIIa	**1st stage**: as per IIb **2nd stage**: use weights according to study design.		✓	All NRS	Mainly RCTs, but also NRS	✓
IIIb	Same as IIIa, but intercept and main effects are estimated from a single NRS.		✓	A single NRS	Mainly RCTs, but also NRS	✓

Performance metrics

- A common practice to measure a prediction model's performance for a continuous outcome is thorugh MSE, bias, and R-squared.
- Calibration slope for outcome: inspect the agreement between observed and predicted patient outcomes

$$y_{ij} = \eta_0 + \sum_{k=1}^{N_T} \eta_k \hat{y}_{ij} I(t_{ij} = k)$$

  where N_T is the total number of treatments and $I(t_{ij} = k)$ is an indicator function that equals 1 if the treatment assigned was k.
- This 'calibration line' compares the observed outcome (y_{ij}) with the predicted outcome ($\hat{y}_{ij}$) across different treatments
- Having η_0 close to 0 and η_1, ..., η_{N_T} close to 1 indicate good performance of the predictions.

Performance metrics (2)

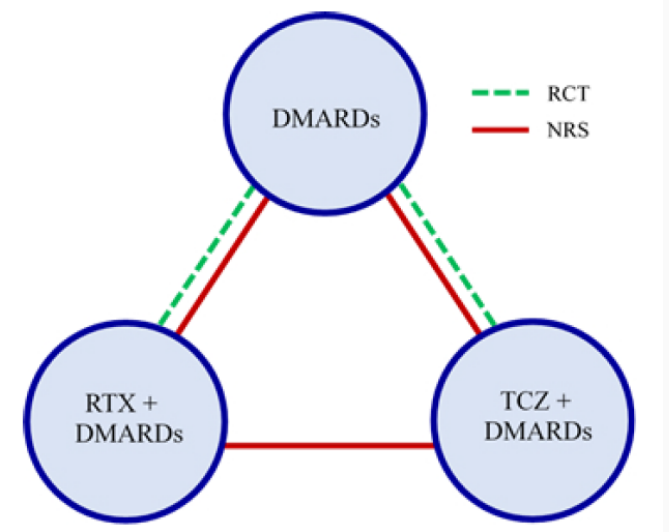
- Calibration slope for benefit: reshape the calibration line using the predicted benefit (i.e. the difference between predicted outcomes under different treatments).

$$y_{ij} = \kappa_0 + \kappa_1 \hat{y}_{ij,A} + \kappa_2 (\hat{y}_{ij,B} - \hat{y}_{ij,A}) I(t_{ij} = B) + \kappa_3 (\hat{y}_{ij,C} - \hat{y}_{ij,A}) I(t_{ij} = C) + ...$$

- Similarly, having κ_0 close to 0 and κ_1,κ_2,... close to 1 indicate good performance of the predictions.

Clinical example: treatments for rheumatoid arthritis (RA)

- We used IPD from 3 RCTs and 2 NRS (a Swiss and a British registry) on patients diagnosed with RA (N = 2524 from RCTs, N=3126 from NRS).
- The outcome of interest was the Disease Activity Score 28 (DAS28).
- The treatments that we consider are disease-modifying anti-rheumatic drugs (DMARDs), Rituximab (RTX) + DMARDs, and Tocilizumab (TCZ) + DMARDs.

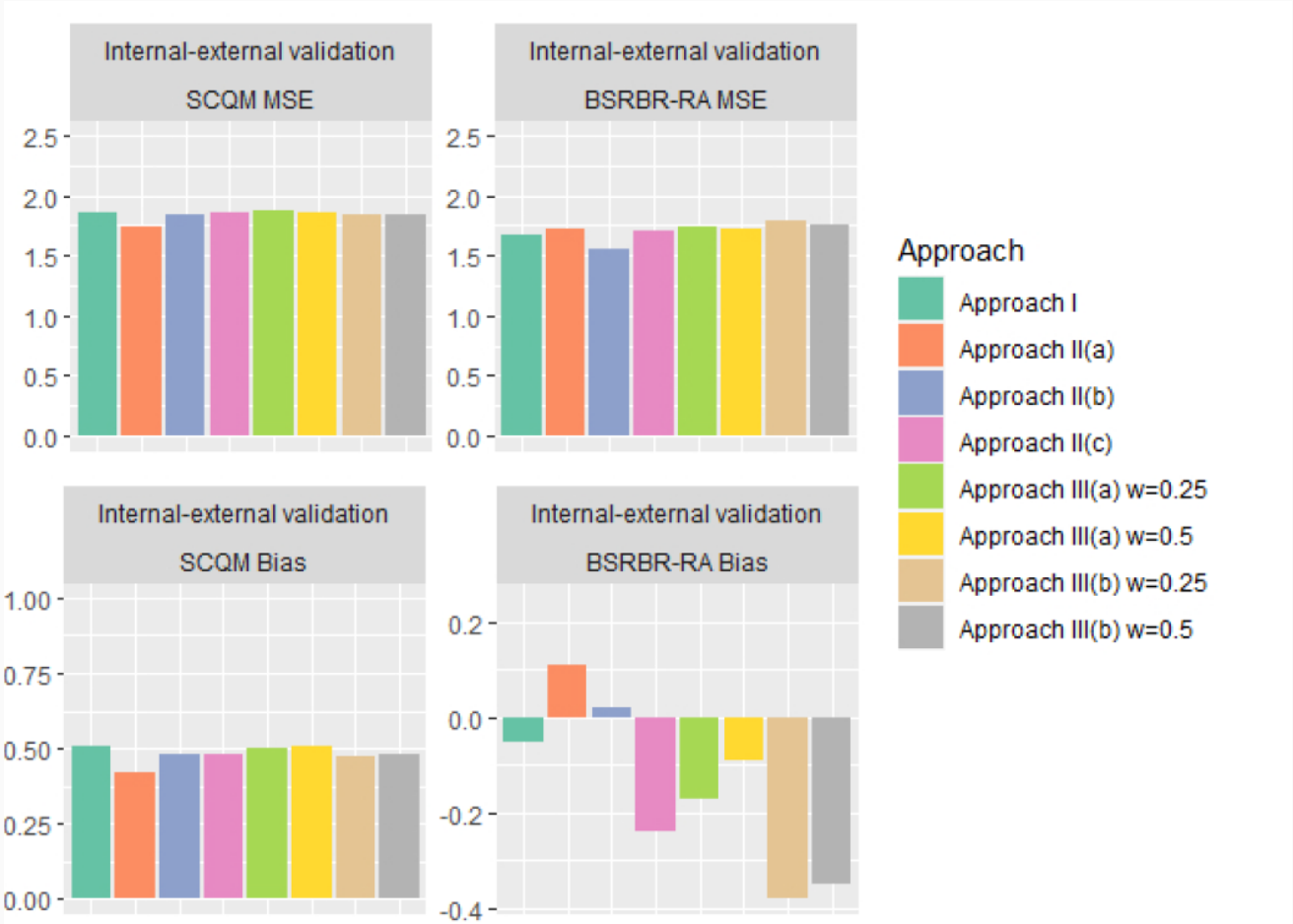


- Nine covariates including gender, age, disease duration, BMI, baseline rhematoid factor, etc

Internal-external cross validation

- We exclude one NRS from the analysis and use the rest of the data to train the models. We then use the left-out NRS to make predictions and compare them with observations. Finally, we cycle through all available NRS.
- For instance, for the RA case study, we leave out the British registry and use Swiss registry and RCTs to develop model and make predictions, vice versa.

Internal-external cross validation results (1)



Internal-external cross validation results (2)

Left-out Dataset	Performance metric	Approach I	Approach II(a)	Approach II(b)	Approach II(c)	Approach III(a) w = 0.25	Approach III(a) w = 0.5	Approach III(b) w = 0.25	Approach III(b) w = 0.5
SCQM	Calibration slope for outcome	η_0=0.25 η_1=0.78 η_2=0.88 η_3=1.04	η_0=0.40 η_1=0.76 η_2=0.86 η_3=0.96	η_0=0.35 η_1=0.76 η_2=0.88 η_3=1.02	η_0=0.34 η_1=0.76 η_2=0.88 η_3=1.05	η_0=0.23 η_1=0.78 η_2=0.93 η_3=1.05	η_0=0.18 η_1=0.79 η_2=0.92 η_3=1.08	η_0=0.33 η_1=0.76 η_2=0.91 η_3=1.03	η_0=0.32 η_1=0.76 η_2=0.89 η_3=1.04
	Calibration slope for benefit	κ_0=0.22 κ_1=0.79 κ_2=0.23 κ_3=0.45	κ_0=0.35 κ_1=0.78 κ_2=0.29 κ_3=0.53	κ_0=0.27 κ_1=0.78 κ_2=0.23 κ_3=0.46	κ_0=0.24 κ_1=0.78 κ_2=0.20 κ_3=0.43	κ_0=0.14 κ_1=0.80 κ_2=0.19 κ_3=0.44	κ_0=0.09 κ_1=0.82 κ_2=0.21 κ_3=0.44	κ_0=0.23 κ_1=0.79 κ_2=0.19 κ_3=0.44	κ_0=0.23 κ_1=0.79 κ_2=0.20 κ_3=0.44
	R squared	0.09	0.14	0.09	0.09	0.08	0.09	0.09	0.09
BSRBR-RA	Calibration slope for outcome	η_0=1.07 η_1=0.80 η_2=0.75 η_3=0.53	η_0=1.47 η_1=0.66 η_2=0.64 η_3=0.40	η_0=0.97 η_1=0.79 η_2=0.77 η_3=0.56	η_0=1.23 η_1=0.76 η_2=0.77 η_3=0.55	η_0=1.35 η_1=0.69 η_2=0.75 η_3=0.50	η_0=1.36 η_1=0.68 η_2=0.72 η_3=0.48	η_0=1.21 η_1=0.77 η_2=0.82 η_3=0.58	η_0=1.15 η_1=0.78 η_2=0.83 η_3=0.60
	Calibration slope for benefit	κ_0=1.11 κ_1=0.75 κ_2=0.59 κ_3=1.32	κ_0=1.10 κ_1=0.70 κ_2=0.36 κ_3=0.98	κ_0=0.87 κ_1=0.78 κ_2=0.59 κ_3=1.08	κ_0=1.02 κ_1=0.78 κ_2=0.47 κ_3=1.03	κ_0=1.17 κ_1=0.72 κ_2=0.38 κ_3=0.95	κ_0=1.23 κ_1=0.69 κ_2=0.39 κ_3=0.95	κ_0=1.00 κ_1=0.80 κ_2=0.45 κ_3=0.98	κ_0=0.98 κ_1=0.81 κ_2=0.48 κ_3=0.99
	R squared	0.27	0.25	0.32	0.25	0.23	0.24	0.21	0.23

Internal-external cross validation results (3)

- Removing Swiss registry for validation resulted in models IIa and IIIb performing best. When we left the British registry, we saw that model IIb performed best.
- This showed that utilizing information from RCTs can help us better predict real-world outcomes compared to using only data from NRS.
- Thus, if we aim to make predictions about patients in a new setting for which no data are currently available, we would recommend either **design naive or design adjusted** meta-analysis, with or without **shrinkage**, but probably not the use of **calibration**.

Conclusion

- We developed a range of **two-stage meta-analysis methods** that combined IPD from RCTs and NRSs, aiming to make predictions about 'real-world' effects of medical interventions.
- We concluded that aggregating data from multiple sources may potentially increase the performance of a clinical prediction model, but **comprehensive cross-validation** is required before employing it in clinical practice.

Kilim plot

SPECIAL ISSUE PAPER

Research Synthesis Methods WILEY

The Kilim plot: A tool for visualizing network meta-analysis results for multiple outcomes

Michael Seo[1] | Toshi A. Furukawa[2] | Areti Angeliki Veroniki[3,4,5] | Toby Pillinger[6,7] | Anneka Tomlinson[8,9] | Georgia Salanti[1] | Andrea Cipriani[8,9] | Orestis Efthimiou[1]

- The Kilim plot can be used to visualize results from NMA on multiple outcomes. It provides information regarding the strength of statistical evidence of treatment effects, and it illustrates absolute, rather than relative, effects of interventions.

Kilim plot (2)

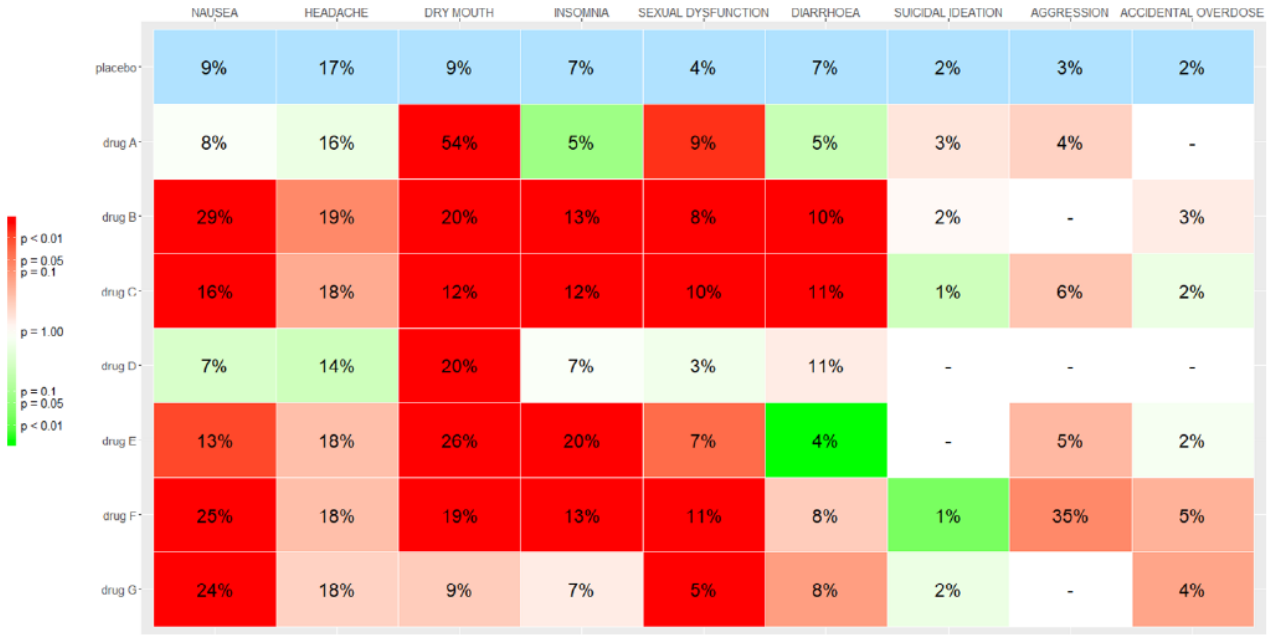


FIGURE 3 Kilim plot for comparing antidepressant drugs for nine outcomes. The numbers in each cell correspond to the estimated absolute event rates for each outcome and treatment. The colors correspond to the strength of statistical evidence regarding the relative effects vs placebo. A cell with a deep green color indicates strong evidence that the corresponding drug performs better than placebo for the corresponding outcome. Conversely, a deep red cell indicates strong evidence that the drug performs worse than placebo. Colors closer to white indicate lack of evidence on whether the drug performs better or worse than placebo [Colour figure can be viewed at

Future research

Developing prediction models when there are systematically missing predictors in an individual patient data meta-analysis

Michael Seo[1,2 *], Orestis Efthimiou[1,3]

We describe various approaches that can be used to develop prediction models in the case of systematically missing predictors across studies.

Background

- A variable is 'systematically missing' if it is wholly missing in some clusters.
- For analysis, we will use the one-stage model that was developed in previous projects.
- We briefly describe different approaches and how to perform internal-external cross validation.
- Let's consider a dataset where we have 4 covariates (X1, X2, X3, and X4) and X2 is systematically missing in half of the studies.

	X1	X2	X3	X4
Study 1	Complete	Complete	Complete	Complete
Study 2	Complete	Complete	Complete	Complete
Study 3	Complete	NA	Complete	Complete
Study 4	Complete	NA	Complete	Complete

Naive method

- The simplest approach is to develop a model using only the predictors measured in all studies.
- For the example dataset, we drop the variable X2 altogether.

Testing dataset

	X1	X3	X4
Study 1	Complete	Complete	Complete

Training dataset

	X1	X3	X4
Study 2	Complete	Complete	Complete
Study 3	Complete	Complete	Complete
Study 4	Complete	Complete	Complete

Imputation method

- Another approach is to impute systematically missing predictors using a hierarchical model to account for clustering of patients in studies and use the imputed datasets for model development
- Recently, there has been methods developed to impute systematically missing variables (Resche-Rigon and White 2018)
- Internal-external cross validation requires some more thought.

Testing dataset (Study with no systematically missing variable)

	X1	X2	X3	X4
Study 1	Complete	Complete	Complete	Complete

Training dataset

	X1	X2	X3	X4
Study 2	Complete	Complete	Complete	Complete
Study 3	Complete	Imputed	Complete	Complete
Study 4	Complete	Imputed	Complete	Complete

Imputation method (2)

- Imputation is done in the training set. We do not impute systematically missing variables in the testing dataset.
- Prediction model should not contain systematically missing variables in testing dataset.

Testing dataset (Study with systematically missing variable)

	X1	X2	X3	X4
Study 3	Complete	NA	Complete	Complete

Training dataset

	X1	X2	X3	X4
Study 1	Complete	Complete	Complete	Complete
Study 2	Complete	Complete	Complete	Complete
Study 4	Complete	Imputed	Complete	Complete

Separate prediction method

- Develop a separate prediction model for each study and then synthesize predictions in a multi-study emsemble.
- Prediction model for each study should contain variables that are not systematically missing in both training and testing dataset.

Testing dataset (Study with no systematically missing variable)

	X1	X2	X3	X4
Study 1	Complete	Complete	Complete	Complete

Training dataset (Study with systematically missing variable)

	X1	X2	X3	X4
Study 3	Complete	NA	Complete	Complete

Future aims

- We will explore in simulations the relative performance of these approaches and we will use a real dataset of 13 trials in psychotherapies for depression to illustrate all methods.
- We would like to conclude that more advanced approaches have the potential to lead to better prediction models when there are systematically missing predictors.

References

Debray, T. P. A., Moons, K. G. M., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R. H. H., Reitsma, J. B., and on behalf of the GetReal methods review group (2015). Get real in individual participant data (ipd) meta-analysis: a review of the methodology. *Research Synthesis Methods*, 6(4):293–309.

Didden, E.-M., Ruffieux, Y., Hummel, N., Efthimiou, O., Reichenbach, S., Gsteiger, S., Finckh, A., Fletcher, C., Salanti, G., and Egger, M. (2018). Prediction of real-world drug effectiveness prelaunch: Case study in rheumatoid arthritis. *Medical Decision Making*, 38(6):719–729. PMID: 30074882.

Efthimiou, O., Mavridis, D., Debray, T. P. A., Samara, M., Belger, M., Siontis, G. C. M., Leucht, S., Salanti, G., and on behalf of GetReal Work Package 4 (2017). Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine*, 36(8):1210–1226.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Lykou, A. and Ntzoufras, I. (2013). On bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, 23.

Neelon, B., James, A., and Malley, O. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics & Biostatistics*, 01.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

References (cont.)

Resche-Rigon, M. and White, I. R. (2018). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research*, 27(6):1634–1649. PMID: 27647809.

Seo, M., Furukawa, T. A., Veroniki, A. A., Pillinger, T., Tomlinson, A., Salanti, G., Cipriani, A., and Efthimiou, O. (2021a). The kilim plot: A tool for visualizing network meta-analysis results for multiple outcomes. *Research Synthesis Methods*, 12(1):86–95.

Seo, M., White, I. R., Furukawa, T. A., Imai, H., Valgimigli, M., Egger, M., Zwahlen, M., and Efthimiou, O. (2021b). Comparing methods for estimating patient-specific treatment effects in individual patient data meta-analysis. *Statistics in Medicine*, 40(6):1553–1573.

Steyerberg, E. and Harrell, F. (2015). Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemiology*, 69:245–247.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.