Predicting real world effectiveness of interventions, combining individual patient data from multiple randomized and non-randomized studies

Michael Seo, Orestis Efthimiou

Institute of Social and Preventive Medicine

University of Bern, Switzerland

Project funded by the Swiss National Science Foundation (Grant No. 180083)

Background

- There has been growing interest in using data from **non-randomized studies** (**NRSs**) 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**.
- Currently, there is a gap in methods for combining **individual patient data** (IPD) from RCTs and NRSs, when aiming to make **predictions** about the real-world effects of medical interventions.

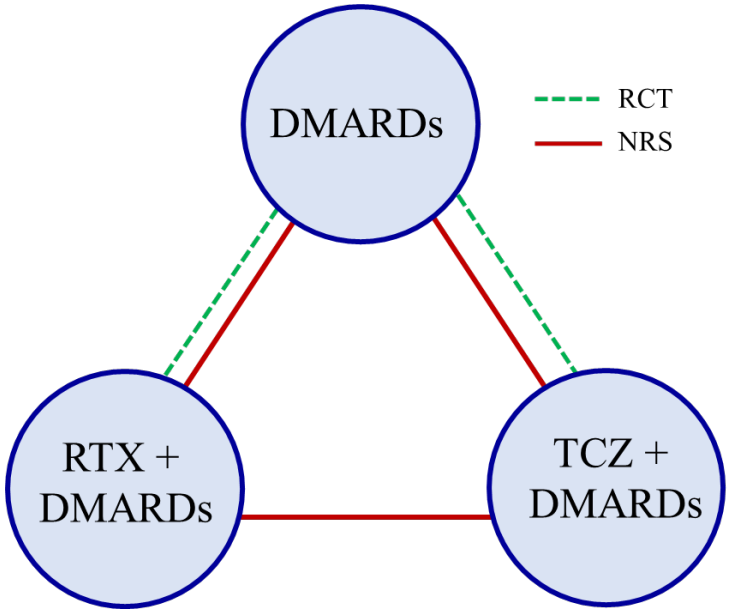
Aims

1. to describe a general framework for developing a **prediction model** that combines IPD from multiple randomized and non-randomized studies, regarding a range of alternative interventions, and

2. to illustrate how to implement this framework in practice, using a real example in **rheumatoid arthritis** (**RA**).

Clinical example: treatments for rheumatoid arthritis (RA)

- We used IPD from **3 RCTs and 2 NRSs** (a Swiss and a British registry) on patients diagnosed with RA (N=2524 from RCTs, N=3126 from NRSs). Data reside in different servers.
- The outcome of interest was the Disease Activity Score 28 (**DAS28**).
- The treatments that we considered are disease-modifying anti-rheumatic drugs (**DMARDs**), Rituximab (**RTX**) + **DMARDs**, and Tocilizumab (**TCZ**) + **DMARDs**.
- **Nine** covariates available:
  - *gender, age, disease duration, BMI, baseline rheumatoid factor, number of previous DMARDs and anti-tumor necrosis factor (TNF) agents, baseline health assessment questionnaire (HAQ) disability index, baseline erythrocyte sedimentation rate (ESR), and baseline DAS28.*



Overview of modelling approaches (1)

- We developed a **series of competing prediction models**, borrowing methods from individual patient data network meta-analysis.
- We used a **two-stage approach**
  - *at the first stage, we analyse each study separately, using the same model*
  - *at the second stage we meta-analyse the study-specific estimates in order to estimate the parameters of our prediction models.*
- Although one-stage approaches are generally more flexible, a two stage approach is often necessary due to **data restrictions.**
  - *this was the case for the RA dataset, where the data from different studies were situated in different servers.*

Overview of modelling approaches (2)

- We focus on **linear models**.
- We set our models in a **Bayesian setting**, which allows increased flexibility (especially in the meta-analytical part, i.e. where we synthesize study-specific estimates)
- We especially focus on the distinction between the **prognostic effect** of the covariates (*presumably better informed by NRSs*) and **relative treatment effects** (*presumably better informed by RCTs*)

Overview of modelling approaches (3)

- The **prediction model** we aim to build is of the following form:

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

Where

- $\boldsymbol{x}$ is a vector of covariates
- t is the treatment variable, A is the reference treatment
- α is the predicted effect for $\boldsymbol{x} = \boldsymbol{0}$, $t = A$
- $\boldsymbol{\beta}$ encompasses the prognostic ability of the covariates
- $\boldsymbol{\gamma_{tA}}$ expresses the effect modification of treatment t vs. A
- δ_{tA} is the relative treatment effect t vs. A for $\boldsymbol{x} = \boldsymbol{0}$

Approach I: target NRS only

- We only use the data from a NRS of interest, to build a prediction model (i.e. **no meta-analysis involved**).
  - *E.g. we only use the data from the Swiss registry to build a prediction model for the Swiss population*
- We use the linear model of the previous slide.
- We can **penalize** (i.e. shrink) the coefficients of the effect modification (treatment-covariate interactions, i.e. parameters $\boldsymbol{\gamma}$ in the prediction equation) using a **Laplace prior distribution** ('Bayesian LASSO').
- We use this as a reference method, to gauge the performance of all advanced, meta-analytic models.

Approach II: meta-analysis disregarding study design

- Using this approach we meta-analyse without distinguishing between RCTs and NRSs
- Assume study j had only two treatment arms, treatment W and the reference A and that patient i in this study j had an outcome y_{ij}. First stage model is as follows:

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

$$m_{ij} = \begin{cases} a_j + \boldsymbol{b_j x_{ij}} & if \;\; t_{ij} = A \\ a_j + \boldsymbol{b_j x_{ij}} + \boldsymbol{c_{j,WA}}\, \boldsymbol{x_{ij}} + d_{j,WA} & if \;\; t_{ij} = W \neq T_{ref} \end{cases}$$

Approach II: meta-analysis disregarding study design

- After fitting we obtain the estimate $\widehat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \left(\hat{a}_j, \widehat{\boldsymbol{b}}_{\boldsymbol{j}}, \hat{\boldsymbol{c}}_{\boldsymbol{j,WA}}, \hat{d}_{j,WA}\right)$ and the corresponding variance-covariance matrix $\hat{\boldsymbol{S}}_{\boldsymbol{j}}$.

- These estimates contribute to the likelihood of the **second stage (meta-analysis) model**, via a multivariate random effects (network) meta-analysis model. We will call this **Approach II(a).**

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{j}} \sim N(\boldsymbol{\lambda}_{\boldsymbol{j}}, \hat{\boldsymbol{S}}_{\boldsymbol{j}})$$

$$\boldsymbol{\lambda}_{\boldsymbol{j}} \sim N\left(\left(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}_{\boldsymbol{WA}}, \delta_{WA}\right)', \boldsymbol{\Sigma}\right)$$

$$(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}_{\boldsymbol{WA}}, \delta_{WA}, \boldsymbol{\Sigma}) \sim (vague\ prior\ distributions)$$

Parameters needed for the prediction model

Approach II: meta-analysis disregarding study design

- We can additionally shrink (i.e. via a Laplace prior) treatment-covariate interactions at the first stage. We will call this **Approach II(b)**
- We can change the model by only using the target population for estimating model intercept (α) and main effects (β). We call this **Approach II(c):**
  - *We pick a NRS population for which we want to make predictions, e.g. the Swiss registry*
  - *We estimate α and β from this study, and include it in our prediction model*
- A motivation is that the average disease progression may be **setting-specific**.
- On the other hand, estimated average relative treatment effects (γ) and effect modification (δ) have the potential to improve upon aggregating results from different RCTs and NRSs.

Approach III: weigh studies by design

- When aggregating $\alpha, \beta, \gamma, \delta$ at the second stage, we use a **weighting scheme according to study design**.
- With this approach, the variance of the estimates of average relative treatment effects (δ) and effect modification (γ) obtained from NRSs is **inflated** after dividing by a factor w_j, with $0 < w_j < 1$.
  - $w_j = 0$ *means excluding the NRS*
  - $w_j = 1$ *corresponds to no down-weighting of the NRS*
- By doing so, we effectively **decrease** the impact of NRSs in the estimation of all relative treatment effects.
- The motivation is that RCTs are considered to be a potentially better source of information regarding relative effects

Approach III: weigh studies by design

The second stage model is as follows:

Variance inflation

$$(\hat{\boldsymbol{c}}_{\boldsymbol{j,WA}}, \hat{d}_{j,WA})' \sim \begin{cases} N\left((\boldsymbol{\gamma_{WA}}, \delta_{WA})', \dfrac{\hat{\boldsymbol{S}}_{\boldsymbol{j}}}{w_j}\right), & if\ study\ j\ is\ a\ NRS \\ N\left((\boldsymbol{\gamma_{WA}}, \delta_{WA})', \hat{\boldsymbol{S}}_{\boldsymbol{j}}\right), & if\ study\ j\ is\ a\ RCT \end{cases}$$

$$(\boldsymbol{\gamma_{WA}}, \delta_{WA}) \sim (vague\ prior\ distributions)$$

Approach III: weigh studies by design

- Since our aim is to predict outcomes for "real-world" patients, estimates for study intercept and the main effects of the covariates (α, β) are aggregated **only using NRSs.** We will call **this approach III(a)**

$$\left(\hat{a}_j, \widehat{\boldsymbol{b}}_{\boldsymbol{j}}\right)' \sim N\left((\alpha, \boldsymbol{\beta})', \widehat{\boldsymbol{S}}_{\boldsymbol{j}}\right), \qquad if\ study\ j\ is\ a\ NRS$$

$$(\alpha, \boldsymbol{\beta}) \sim (vague\ prior\ distributions)$$

- Instead of aggregating NRSs to estimate study intercept and the main effects, we can only use the target NRS of interest. We will call this **approach III(b).**
- We explored models where $w_j = 0.25, 0.5, 0.75$

	Description	Absolute effects (α, β) are estimated from	Relative effects (γ, δ) are estimated from
Approach I	No meta-analysis; We build a prediction model using one NRS alone	Single NRS	Single NRS
Approach II(a)	First stage: we fit a linear model separately in each study. Second stage: we fit a design naïve analysis	RCTs and NRSs	RCTs and NRSs
Approach II(b)	Same as II(a), but where at the first stage we penalize coefficients of effect modification using Bayesian LASSO	RCTs and NRSs	RCTs and NRSs
Approach II(c)	Same as Approach II(b), but only use target NRS for intercept and main effects	Single NRS	RCTs and NRSs
Approach III(a)	First stage: as per Approach II(b) Second stage: weight of NRSs is reduced when estimating relative effects. Absolute effects are estimated using only NRSs.	Only NRSs	Mainly RCTs, but also NRSs
Approach III(b)	Same as Approach III(a), but only use target NRS for intercept and main effects	Single NRS	Mainly RCTs, but also NRSs

Assessing the performance of the prediction models

- We use an **internal** and an **internal-external cross-validation** approach.
- We calculate mean square error (MSE) and bias

$$MSE = \frac{1}{N}\sum_{i,j}\left(\hat{y}_{ij} - y_{ij}\right)^2$$

$$Bias = \frac{1}{N}\sum_{i,j}\left(\hat{y}_{ij} - y_{ij}\right)$$

- Calibration line for the multiple treatments: $y_{ij} = a\ \hat{y}_{ij,A}\ I_A\left(t_{ij}\right) + b\ \hat{y}_{ij,B}\ I_B\left(t_{ij}\right) + \cdots$ where $\hat{y}_{ij,X}$ is the predicted outcome of a patient if administered treatment X

Results – Internal validation

- We develop the model using the entire datasets to train the model
- We predict results for each target NRS datasets (i.e. Swiss/British registry) and compare with observations

Dataset	Performance metric	Approach I	Approach II(a)	Approach II(b)	Approach II(c)	Approach III(a) w = 0.5	Approach III(b) w = 0.5
Swiss registry	MSE	1.44	2.23	2.35	1.45	1.83	1.49
	Bias	0.10	0.88	0.94	0.08	0.59	0.01
	Calibration slope	a=0.97 b=0.97 c=1.06	a=0.79 b=0.85 c=0.78	a=0.78 b=0.82 c=0.76	a=0.97 b=1.01 c=1.06	a=0.85 b=0.86 c=0.95	a=0.97 b=1.07 c=1.38
British registry	MSE	1.39	1.48	1.48	1.70	1.56	1.44
	Bias	0.13	0.17	0.12	0.45	-0.35	0.15
	Calibration slope	a=0.98 b=0.98 c=0.93	a=1.01 b=0.95 c=0.80	a=1.03 b=0.96 c=0.79	a=0.98 b=0.89 c=0.71	a=1.08 b=1.09 c=1.07	a=0.98 b=0.96 c=0.85

Results – Internal-external validation

- We exclude the target NRS and use rest of the data to train the model. We then use the target NRS to make predictions and compare with observations.
- For Approach I, we only use non-target NRS dataset to train. For instance, to predict Swiss registry, we only used the British registry to train.

Dataset	Performance metric	Approach I	Approach II(a)	Approach II(b)	Approach II(c)	Approach III(a) w = 0.5	Approach III(b) w = 0.5
Swiss registry	MSE	3.08	4.23	4.18	3.07	3.21	3.05
	Bias	1.23	1.62	1.60	1.23	1.28	1.21
	Calibration slope	a=0.72 b=0.82 c=0.89	a=0.65 b=0.81 c=0.74	a=0.66 b=0.80 c=0.76	a=0.72 b=0.83 c=0.82	a=0.71 b=0.84 c=0.85	a=0.72 b=0.85 c=0.86
British registry	MSE	2.12	1.68	1.57	1.93	2.68	2.91
	Bias	-0.79	0.26	0.14	-0.61	-1.07	-1.18
	Calibration slope	a=1.29 b=1.19 c=1.10	a=1.02 b=0.91 c=0.77	a=1.06 b=0.94 c=0.77	a=1.29 b=1.11 c=0.98	a=1.24 b=1.40 c=1.33	a=1.29 b=1.46 c=1.38
Overall (i.e. weighted average of Swiss and British registry)	MSE	2.45	2.55	2.46	2.32	2.68	2.91
	Bias	-0.10	0.73	0.64	0.02	-1.07	-1.18

Possible extensions – future work

- Explore **dynamic weighting** of the evidence
  - *Instead of assigning fixed variance inflation factors (as per Approach III), we could use a flexible weighting scheme.*
  - *E.g. when RCT and NRS parameter estimates agree, the weight of the NRS will increase. When results disagree, NRS will be down-weighted.*
  - *This method did not work well for the RA example.*
- Explore the so-called '**ensemble**' methods
  - *We can take the predictions from different approaches and use an algorithm that combines them into a single model.*
  - *The model uses as covariates the predictions of the all previous prediction models*
  - *Ensemble method never does worse than selecting the single best model on the training data (but not necessarily on the test data)*

Conclusions

- We developed a range of **two-stage meta-analysis methods** that combine IPD from RCTs and NRSs, aiming to make predictions about ‘real-world’ effects of medical interventions.

- We discussed how to decide between competing models using internal and internal-external validation.

- In the RA example:
  - *Approach IIc had the best **internal-external performance**. This model would probably be the best to use in a new population.*
  - *However, Approach I (i.e. no meta-analysis involved) as well as other approaches performed almost equally well in some cases.*
  - *Thus, no model was a clear winner.*

- 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.

References

- Debray TPA, Moons KGM, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods*. 2015;6(4):293-309.
- Didden EM, Ruffieux Y, Hummel N, et al. Prediction of Real-World Drug Effectiveness Prelaunch: Case Study in Rheumatoid Arthritis. *Medical Decision Making*, *38*(6), 719–729.
- Efthimiou O, Debray TPA, Valkenhoef G van, et al. GetReal in network meta-analysis: a review of the methodology. *Research Synthesis Methods*. 2016;7(3):236-263.
- Efthimiou O, Mavridis D, Debray TPA, et al. Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine*. 2017;36(8):1210-1226.
- Park T, Casella G. The Bayesian Lasso. *Journal of the American Statistical Association*. 2008;103(482):681-686.
- Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal–external, and external validation. *Journal of Clinical Epidemiology*. 2016;69:245-247.