Precision medicine in cardiorenal and metabolic diseases with routinely collected clinical data: a novel insight

Sheyu Li [1], Qingyang Shi[1], Valentyn Litvin[2] and Charles F. Manski[3,*]

[1]Department of Endocrinology and Metabolism, MAGIC China Centre, Cochrane China Centre, Chinese Evidence-based Medicine Centre, West China Hospital, Sichuan University, Chengdu 610041, China
[2]Department of Economics, Northwestern University, Evanston, IL 60208, USA
[3]Department of Economics and Institute for Policy Research, Northwestern University, Evanston, IL 60208, USA
*Correspondence: Charles F. Manski, cfmanski@northwestern.edu

Dear Editor,

Evidence-based medicine guides modern clinical practice. Quantitative evidence synthesis facilitates clinical decision-making after comprehensive interviews between clinicians and patients. Individualised concern is always critical, not only for the diversity of patients' values and preferences, but also for questioning whether the treatment response of the individual is in line with the population. Precision medicine aims to choose the right treatment for each person.

The success of precision medicine in oncology relies on the biological progression that a particular genetic mutation determines the development and treatment response of the neoplasm. However, patients with cardiorenal and metabolic diseases have thus far benefited much less from genetic or omics technology. Although genetics and metabolic parameters show an association with individualised drug response, only a few clinicians have changed their practice.[1] This suggests that genetic factors are associated with but may not dominantly determine drug response in patients with cardiorenal and metabolic diseases.[2] Turning back to routinely collected clinical data, which has supported decision-making by clinicians for more than a thousand years, could be helpful.

Well-conducted randomised controlled trials provide high-certainty evidence for clinicians and patients during their daily practice.[3] Subgroup analyses, now routine in reporting results, help identify potential treatment response diversity in patients with different covariates. However, most of them fail to reach statistical significance or clinical credibility given a large number of covariates of interest but a limited number of trial participants. Without credible support from subgroup analyses, clinicians have to treat people as if all respond equally to a drug with the same relative effect. Even when some independent subgroup analyses yield fruitful results, clinicians can find it difficult to interpret this information quantitatively for a single patient with a particular combination of these covariates.

Systematic review and meta-analysis summarise the evidence from randomised trials to support clinical practice guidelines. Without individual patient data in most cases, however, meta-analysis relies on summary effects of the published trials that represent the marginal effects on the trial population and corresponding distribution of patients' characteristics. The pooled effect of a meta-analysis barely explains the heterogeneity of treatment effects among different groups of people when the study-level subgroup analyses or meta-regressions are subject to ecological bias (*i.e.* aggregation bias).

Data sharing of completed randomised controlled trials allows meta-analyses at the individual patient level to categorise people based on their routinely measured parameters. Statisticians and epidemiologists developed a buddle fashion approach to achieve this. Vo and colleagues addressed case-mix heterogeneity of treatment response by direct standardisation or inverse probability weighting to a target population.[4] Coincidentally, Dahabreh and colleagues proposed a framework to transport inferences from the trial participants to a target population using either outcome regression or inverse probability weighting requesting five assumptions.[5] Both methods require individual patient data in all included trials. When individual patient data is available in some but not all trials, Phillippo and colleagues' multilevel meta-regression linked statistical parameters for trials with aggerate data to the parameters for trials with individual data by integrating the covariates' distribution to form the likelihood, assuming the consistency of parameter distributions across trial populations.[6] This assumption is reasonable in some cases but strong. The number and proportion of the available individual patient data heavily affect the degree of validity of estimations and restricted the use of this approach.

These novel approaches work, but they require sharing of at least some individual patient data of the randomised controlled trials and also strong statistical assumptions. In 2017, the International Committee of Medical Journal Editors (ICMJE) released a statement regarding data sharing policy, and some medical journals requested that authors upload individual-level patient data to accompany their articles.[7] The Blood Pressure Lowering Treatment Trialists' Collaboration (BPLTTC) group set a good example by integrating the individual patient data of most large trials

Table 1 Examples of bounded-variation assumptions.

Bounded-variation assumption	Description
Bound on the rate/risk of treatment outcome	The range between the minimal and maximal rates/risks of the outcome, reflecting the possible border of natural incidence.
Bound on absolute difference across covariates	The range between the minimal and maximal absolute rate/risk differences of outcomes in people with different covariates, reflecting the largest possible absolute magnitude of the covariate impact on the outcome.
Bound on relative difference across covariates	The range between the minimal and maximal rate/risk ratios of outcome in people with different covariates, reflecting the largest possible relative magnitude of the covariate impact on the outcome.
Bound on absolute difference across interventions	The range between the minimal and maximal absolute rate/risk differences of outcome in people receiving different interventions, reflecting the largest possible absolute magnitude of the intervention impact on the outcome.
Bound on relative difference across interventions	The range between the minimal and maximal rate/risk ratios of outcome in people receiving different interventions, reflecting the largest possible relative magnitude of the intervention impact on the outcome.

for hypertension.[8] Nevertheless, this data sharing was restricted within the collaborating team and the trial data remain mainly inaccessible to the public or research peers outside the group. Publishers and researchers have to await respones to their increasing requests for data-sharing until the solutions to various difficult concerns, including but not restricting to patient privacy, technical barriers, and conflicts of interest.

Strong assumptions facilitate statistical inference, but imply flawed conclusions if researchers ignore their violation. These violations happen but are often ignored in clinical practice. For example, traditional meta-analysis deals with heterogeneity between studies by assuming a random-effect distribution of treatment effects, typically following a normal distribution. However, several sources of heterogeneity exist and are not necessarily normally distributed. When this is the case, the pooled overall effect does not have a clear clinical interpretation from a patient-centred perspective.[9]

Our recent working paper applies partial identification, a credible method that has served econometrics for decades, to precision medicine utilising routinely reported subgroup analyses of binary data in randomised controlled trials.[10] Partial identification analysis makes relatively weak, reasonable assumptions and uses them to generate bounds (*i.e.* interval or set value) on the effects of interest.[11,12] The basic idea is to start from set-valued identification regions obtained in the absence of assumptions and then impose realistic assumptions to shrink the regions to be sufficiently informative or yield clinically useful conclusions.

Applying the method to a large randomised controlled trial of empagliflozin in adults with type 2 diabetes at high cardiovascular risk, the team selected three patient covariates (age, sex, and glycated haemoglobin [HbA1c] at baseline) that may impact response to the drug empagliflozin. Each covariate allows the anticipation of the conditional treatment effect, and we call these short outcomes. For a particular patient with a certain combination of several binary covariates, combining the corresponding short outcomes implies long outcomes, which is the individualised treatment effect. Without any assumptions, the long outcome in the given example provided null information regarding the rate of major cardiovascular events (MACE, the primary outcome of the trial) in individuals with different combinations of the three covariates. Imposing bounded-variation assumptions can make the results meaningful to clinicians. Bounded-variation assumptions restrict the potential magnitude of the effects of interest and the degree to which the effects can vary with different patients' covariates. We propose five types of

bounded-variation assumptions that could be helpful for medical estimates (Table 1). In the preliminary exercise, we assumed that a single covariate could only limitedly contribute to the drug response and the variation could be no larger than 0.05. The analyses showed that older males with lower HbA1c and older females with higher HbA1c without receiving empagliflozin are at the highest risk of MACE, while young males with higher HbA1c and young females with lower HbA1c may be at the lowest risk without taking the drug.

This approach does not require individual patient data, but it asks clinicians to understand bounded-variation assumptions and impose them based on clinical knowledge. The bounded variation represents the largest possible range of rate or effect size across treatment or covariates, possibly ranging from 0 to ∞. When assuming bounded variation across treatments, the clinicians should consider the theoretically maximal influence of a single patient character or an intervention on the effect size. Our illustrative exercise assumed a narrow-bounded variation to test the methodological feasibility of the approach. In practice, pooling analysis of multiple trials using a wider bounded variation could also yield a tight bound. An improper assumption of bounded variation may lead to a null bound that is not useful for practice, especially when the heterogeneity of a population or intervention is large. To solve real-world clinical problems, our new approach is a work in progress, but it has a bright potential.

In summary, precision medicine for cardiorenal and metabolic diseases calls for the utilisation of routinely collected data in practice. Individual-level patient data from large randomised controlled trials are essential to estimate personalised treatment effects precisely, but such data are largely inaccessible. When individual patient data are unavailable, a very recent approach adopting partial identification may informatively bound personalised treatment effects using weak assumptions based on clinical knowledge.

Conflict of interest

None declared.

References

1. Johnson JA. Warfarin pharmacogenetics: a rising tide for its clinical value *Circulation* 2012;**125** (16): 1964–6. https://doi.org/10.1161/CIRCULATIONAHA.112.100628.

2. Gloyn AL, Drucker DJ. Precision medicine in the management of type 2 diabetes. *The Lancet Diabetes & Endocrinology* 2018;**6** (11): 891–900. https://doi.org/10.1016/S2213-8587(18)30052-4.

3. Guyatt GH, Oxman AD, Vist GE, *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;**336** (7650): 924–6. https://doi.org/10.1136/bmj.39489.470347.AD.

4. Vo T-T, Porcher R, Chaimani A, *et al*. A novel approach for identifying and addressing case-mix heterogeneity in individual participant data meta-analysis. *Research Synthesis Methods* 2019;**10** (4): 582–96. https://doi.org/10.1002/jrsm.1382.

5. Dahabreh IJ, Petito LC, Robertson SE, *et al*. Toward causally interpretable meta-analysis: Transporting inferences from multiple randomized trials to a new target population. *Epidemiology* 2020;**31** (3): 334–44. https://doi.org/10.1097/EDE.0000000000001177.

6. Phillippo DM, Dias S, Ades AE, *et al*. Multilevel network meta-regression for population-adjusted treatment comparisons. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2020;**183** (3): 1189–210. https://doi.org/10.1111/rssa.12579.

7. Taichman DB, Sahni P, Pinborg A, *et al*. Data sharing statements for clinical trials—A requirement of the International Committee of Medical Journal Editors. *N Engl J Med* 2017;**376** (23): 2277–9. https://doi.org/10.1056/NEJMe1705439.

8. Rahimi K, Canoy D, Nazarzadeh M, *et al*. Investigating the stratified efficacy and safety of pharmacological blood pressure-lowering: an overall protocol for individual patient-level data meta-analyses of over 300 000 randomised participants in the new phase of the Blood Pressure Lowering Treatment Trialists' Collaboration (BPLTTC). *BMJ Open* 2019;**9**:e028698. https://doi.org/10.1136/bmjopen-2018-028698.

9. Manski CF. Toward credible patient-centered meta-analysis. *Epidemiology* 2020;**31** (3): 345–52. https://doi.org/10.1097/EDE.0000000000001178.

10. Li S, Litvin V, Manski CF. Partial identification of personalized treatment response with trial-reported analyses of binary subgroups. *arXiv* 2022. https://doi.org/10.48550/arXiv.2208.03381.

11. Manski CF. *Partial Identification of Probability Distributions*. New York: Springer-Verlag, 2003. https://doi.org/10.1007/b97478.

12. Mullahy J, Venkataramani A, Millimet DL, *et al*. Embracing uncertainty: The value of partial identification in public health and clinical research. *Am J Prev Med* 2021;**61** (2): e103–8. https://doi.org/10.1016/j.amepre.2021.01.041.