

Assessing heterogeneity of treatment effect in multiple sclerosis trials

Maria Pia Sormani , Jeremy Chataway, David M Kent and Ruth Ann Marrie ; On Behalf of the Participants in the International Conference on Innovations in Clinical Trial Design & Enhancing Inclusivity of Clinical Trial Populations

Abstract: Multiple sclerosis (MS) is heterogeneous with respect to outcomes, and evaluating possible heterogeneity of treatment effect (HTE) is of high interest. HTE is non-random variation in the magnitude of a treatment effect on a clinical outcome across levels of a covariate (i.e. a patient attribute or set of attributes). Multiple statistical techniques can evaluate HTE. The simplest but most bias-prone is conventional one variable-at-a-time subgroup analysis. Recently, multivariable predictive approaches have been promoted to provide more patient-centered results, by accounting for multiple relevant attributes simultaneously. We review approaches used to estimate HTE in clinical trials of MS.

Keywords: Multiple sclerosis, clinical trial

Date received: 8 March 2023; revised: 8 May 2023; accepted: 10 May 2023

Introduction

Outcomes in multiple sclerosis (MS) are well recognized to be heterogeneous. This heterogeneity may also extend to therapeutic responses and adverse effects and can be accounted for by multiple factors including genetic differences, as well as age, sex, liver or renal function, health behaviors, and disease severity.¹ For example, genetic variation influences the expression of thiopurine S-methyltransferase (TPMT) enzyme. Individuals with a homozygous deficiency of TPMT activity are at high risk of myelosuppression when treated with azathioprine at usual doses.^{2,3} Genetic variation in the CYP2C9 gene influences metabolism of siponimod. The use of siponimod is contraindicated in individuals with the *3/*3 CYP2C9 genotype due to dangerously prolonged metabolism, and reduction of the maintenance dose from 2 mg daily to 1 mg daily is required for individuals with *1/*3 or *2/*3 genotypes.⁴ Nevertheless, typically evidence from the overall results of clinical trials, which provide average effects in populations, is used to predict potential outcomes in an individual patient, who is assumed to be similar to patients treated in the trial.

Heterogeneity of treatment effect (HTE) refers to non-random variation in the magnitude or direction of

the effect of treatment on a clinical outcome of interest in different patient subgroups defined by one or more covariates.¹ For example, the effect of a disease-modifying therapy (DMT) might be larger for younger individuals with relapsing remitting MS who have multiple gadolinium-enhancing lesions than for older individuals with secondary progressive MS who do not have gadolinium-enhancing lesions. In epidemiology, the equivalent concept is effect modification. HTE assessment is the cornerstone of precision medicine, which seeks to predict the optimal treatments at the individual level according to patient-specific characteristics. The most widely used, simple and most biased HTE assessment uses the one variable-at-a-time subgroup analysis. Predictive approaches to HTE analyses provide individualized predictions of treatment benefit considering multiple relevant patient characteristics simultaneously and are foundational to personalization in evidence-based medicine.⁵

In December 2022, a workshop was held by an international group sponsored by the European Committee on Treatment and Research in MS and the U.S. National MS Society. Participants included members of the International Advisory Committee on Clinical Trials, as well as external participants with expertise

Multiple Sclerosis Journal

2023, Vol. 29(9) 1158–1161

DOI: 10.1177/

13524585231189673

© The Author(s), 2023.



Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

RA Marrie

Departments of Internal
Medicine and Community
Health Sciences, Max Rady
College of Medicine, Rady
Faculty of Health Sciences,
University of Manitoba,
GF532-820 Sherbrook Street,
Winnipeg, MB R3A 1R9,
Canada.

rmarrie@hsc.mb.ca

Maria Pia Sormani

Department of Health
Sciences, University of
Genoa, Genoa, Italy/IRCCS
Ospedale Policlinico San
Martino, Genova, Italy

Jeremy Chataway

Queen Square Multiple
Sclerosis Centre, Department
of Neuroinflammation,
UCL Queen Square Institute
of Neurology, Faculty of
Brain Sciences, University
College London, London,
UK/National Institute for
Health Research, University
College London Hospitals,
Biomedical Research
Centre, London, UK/
Medical Research Council
Clinical Trials Unit at UCL,
Institute of Clinical Trials
and Methodology, University
College London, London, UK

David M Kent

Predictive Analytics and
Comparative Effectiveness
(PACE) Center, Institute for
Clinical Research and Health
Policy Studies, Tufts Medical
Center, Boston, MA, USA

Ruth Ann Marrie

Departments of Internal
Medicine and Community
Health Sciences, Max Rady
College of Medicine, Rady
Faculty of Health Sciences,
University of Manitoba,
Winnipeg, MB, Canada

in multiple areas relevant to clinical trials, including biostatistics. Herein, we describe HTE assessment methods in MS clinical trials as discussed in that workshop.

Subgroup analyses

Traditionally, clinical trials use pre-planned or post-hoc subgroup analyses to address HTE, which may or may not be formally tested using a statistical interaction term between treatment and the subgrouping covariate. These approaches have substantial limitations, including limited power to detect true effects (type II error), and multiple comparisons (increasing type I error). Moreover, “these analyses are also incongruent with the way clinical decision-making occurs at the level of the individual patient, because patients have multiple attributes simultaneously that can affect the tradeoffs between the benefits and harms of the intervention.”⁵

To evaluate the methods used for subgroup analysis in MS trials, we used an existing systematic review⁶ which examined racial and ethnic characteristics of participants in phase 3 trials of DMTs for relapsing-remitting MS conducted between 1995 and June 2006.⁶ We used a standardized form to extract study name, year, intervention, comparator, and subgroup analyses (subgroup definition, if *a priori* or *post hoc*, whether analyses involved stratification or interaction terms). We did not evaluate observational studies, although subgroup analyses are also relevant to those study designs, because the focus of the workshop was on clinical trials.

The primary systematic review identified 45 trials (44 publications) of which 31 (68.9%) conducted subgroup analyses; only 11 used interaction terms, and none used multivariable models (Supplementary Table e1). Most tested for statistical significance of treatment effects within strata (subgroup-specific analysis), rather than for the contrast in treatment effects between strata, thereby increasing false positive rates.⁷ The few that used formal tests of interaction were likely underpowered. Moreover, even if the interaction test was correctly done, some studies reported subgroup-specific *p* values.⁸

In a simulation study, trials with 80% power to detect an overall treatment effect had power to detect an interaction effect of similar magnitude as the overall effect of 29%.⁹ A 2017 meta-analysis of 64 clinical trials reporting ≥ 1 positive subgroup found that 46 subgroups (33 trials) included an interaction test that supported statistically significant heterogeneity.¹⁰ Only 5 of 46 (10.9%)

subgroups were tested for reproducibility, none successfully.¹⁰ These empirical results conform to what is expected theoretically: Weak theory and noisy data (i.e. exploratory analysis in a low power setting) are a recipe for generating false positive findings.¹¹

Predictive HTE approaches

Importantly, HTE, effect measure modification, and statistical interaction are “scale-dependent” concepts; their presence or absence depends on the scale selected to measure treatment effect¹² as discussed elsewhere.^{5,12} Treatment effect estimates in randomized controlled trial (RCT) are usually described on a relative scale (i.e. odds ratios for binary outcomes, or hazard ratios for time-to-event outcomes). The analysis of HTE is usually conducted on a relative treatment effect scale because of statistical convenience and because relative effects are understood to be the most transportable ones.¹² However, for clinical decision-making, it is most important to interpret variations in effects on the absolute risk difference scale.

The PATH (Predictive Approaches to Treatment effect Heterogeneity) statement⁵ was created to encourage and guide predictive HTE analysis and describes two distinct approaches to predictive HTE analysis. With a “risk-modeling” approach, a multivariable model that predicts risk for an outcome is identified from external sources (an “external model”) or developed directly on the trial population without a term for treatment assignment (an “internal model”). This prediction model is applied to disaggregate patients within trials to examine risk-based variation in treatment effects. It takes advantage of the fact that the risk of the outcome is a determinant of the treatment effect and that the absolute risk difference varies across strata even when the relative treatment effect is the same (Supplementary Figure e1). With the “effect modeling” approach, a model is developed on RCT data that include a term indicating treatment assignment and interaction terms between treatment assignment and ≥ 1 covariate (e.g. sex).⁵ The advantage of risk modeling is that subgroup identification is “blinded” to treatment assignment and thus separate from treatment effect estimation. Effect modeling, in contrast, uses treatment effect estimation to identify patients who benefit most. It may thus be more prone to overfitting and bias, unless precautions such as rigorous internal validation approaches are undertaken.¹³

HTE assessment in MS trials

Despite the evidence that treatment responses are heterogeneous in MS, few specific biomarkers exist to guide treatment choice; these are more often related

to safety than efficacy. Thus, it is challenging to identify drug responders or non-responders. Post-hoc subgroup analyses of clinical trials in MS have not found specific markers of response to approved drugs; a meta-analysis² including all published post-hoc subgroup analyses of clinical trials in relapsing remitting MS (RRMS) indicated generic predictors of higher response to DMTs, like younger age, lower Expanded Disability Status Scale, and presence of Gadolinium-enhancing lesions at baseline.

Recently, predictive HTE approaches have been applied to MS trials. A proof-of-concept study used data from the pivotal studies of dimethyl-fumarate in RRMS (the DEFINE and CONFIRM trials).¹⁴ A treatment effect modeling approach¹⁵ allowed calculation of a patient-specific score derived from a linear combination of the baseline variables to predict the individualized size of treatment effect. This model indicates that a subgroup of “super-responders” to dimethyl-fumarate can be identified, with a relapse rate reduction higher than the average effect in the trials. A subsequent testing/validation procedure on three clinical trials successfully predicted a subgroup of RRMS patients that was responsive to laquinimod,¹⁶ a drug for which the average treatment effect in the original phase 3 studies was insufficient to obtain approval (Supplementary Table e2). This approach was also applied to an active controlled study (the CombiRx study¹⁷), comparing interferon-beta to glatiramer-acetate. The ability of the combination of patients’ characteristics to discriminate responders to interferon-beta and glatiramer-acetate was replicated using real-world data.¹⁷ The same methodology was applied to enable short proof-of-concept trials in progressive MS by using a deep-learning predictive model¹⁸ to predict those more likely to progress, allowing enrichment of study populations, thereby increasing statistical power. A recent study¹⁹ classified MS subtypes using an unsupervised machine learning algorithm on brain magnetic resonance imaging (MRI) scans acquired in clinical trials; the results suggested that MRI-based subtypes predict not only MS disability progression but also response to treatment, indicating a potential method to define groups of patients in clinical trials albeit requiring validation.

Recently, a paper applied a risk modeling approach in a network meta-analytic setting²⁰ to estimate the benefit of alternative treatment options for individual patients with MS. First, a prognostic model was developed to predict the baseline risk of the outcome. Second, the baseline risk score from the first stage was used as a single prognostic factor and effect modifier in a network meta-regression model and was found to modify treatment effects.

Conclusions

The need to identify “responders” to therapies is urgent in MS, since many treatments are available, and the response to each drug is highly heterogeneous. To personalize the use of DMT, we must ensure that we capture and share standardized demographic, clinical, and biomarker-based characteristics that may influence outcomes, including comorbidities and social determinants of health (Supplementary Table e3). Many studies erroneously try to characterize responders to therapies by examining the disease course during treatment, yet identification of differential treatment effects requires that HTE assessment be based on trial data; the PATH statement provides relevant guidelines that should be followed.

Declaration of Conflicting Interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: M.P.S. has received consulting fees from Biogen, Genzyme, GeNeuro, MedDay, Merck, Novartis, Roche, and Teva. In the last 3 years, J.C. has received support from the Efficacy and Evaluation (EME) Program, a Medical Research Council (MRC) and National Institute for Health Research (NIHR) partnership and the Health Technology Assessment (HTA) Program (NIHR), the UK MS Society, the US National MS Society, and the Rosetrees Trust. He is supported in part by the NIHR University College London Hospitals (UCLH) Biomedical Research Center, London, UK. He has been a local principal investigator for a trial in MS funded by MS Canada; a local principal investigator for commercial trials funded by Ionis, Novartis, and Roche; and has taken part in advisory boards/consultancy for Azadyne, Biogen, Lucid, Janssen, Merck, NervGen, Novartis, and Roche. In the last 3 years, D.M.K. has received consulting fees from the American Medical Group Association (AMGA) Analytics and the Canadian Stroke Consortium. He has received research funding from the National Institutes of Health (NIH), Patient-Centered Outcomes Research Institute (PCORI), Greenwall Foundation, and W. L. Gore and served on the Scientific Advisory Board for Optum Labs. R.A.M. receives research funding from Canadian Institutes of Health Research, Research Manitoba, MS Canada, Multiple Sclerosis Scientific Foundation, Crohn’s and Colitis Canada, National Multiple Sclerosis Society, Consortium of MS Centers, the Arthritis Society, and US Department of Defense. She is supported by the Waugh Family Chair in Multiple Sclerosis. She is a co-investigator on a study funded in part by Biogen Idec and Roche (no funds to her or her institution).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The International Advisory Committee on Clinical Trials in Multiple Sclerosis and the International Conference on Innovations in Clinical Trial Design & Enhancing Inclusivity of Clinical Trial Populations were supported by the National Multiple Sclerosis Society and the European Committee for Treatment and Research in Multiple Sclerosis. There was no involvement of the sponsors in the design, collection, analysis, or interpretation of data discussed at the conference. The opinions expressed are those of the authors.

ORCID iDs

Maria Pia Sormani  <https://orcid.org/0000-0001-6892-104X>

Ruth Ann Marrie  <https://orcid.org/0000-0002-1855-5595>

Supplemental Material

Supplemental material for this article is available online.

References

- Goldstein DB, Need AC, Singh R, et al. Potential genetic causes of heterogeneity of treatment effects. *Am J Med* 2007; 120(4, Suppl. 1): S21–S25.
- Signori A, Schiavetti I, Gallo F, et al. Subgroups of multiple sclerosis patients with larger treatment benefits: A meta-analysis of randomized trials. *Eur J Neurol* 2015; 22(6): 960–966.
- Ford LT and Berg JD. Thiopurine S-methyltransferase (TPMT) assessment prior to starting thiopurine drug treatment; a pharmacogenomic test whose time has come. *J Clin Pathol* 2010; 63(4): 288–295.
- Díaz-Villamarín X, Piñar-Morales R, Barrero-Hernández FJ, et al. Pharmacogenetics of siponimod: A systematic review. *Biomed Pharmacother* 2022; 153: 113536.
- Kent D, Paulus J, Van Klaveren D, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement. *Ann Intern Med* 2020; 172: 35–45.
- Onuorah H-M, Charron O, Meltzer E, et al. Enrollment of non-white participants and reporting of race and ethnicity in phase III trials of multiple sclerosis DMTs: A systematic review. *Neurology* 2022; 98: e880–e892.
- Brookes ST, Whitley E, Peters TJ, et al. Subgroup analyses in randomised controlled trials: Quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001; 5(33): 1–56.
- Sormani MP and Bruzzi P. Reporting of subgroup analyses from clinical trials. *Lancet Neurol* 2012; 11(9): 747.
- Brookes ST, Whitley E, Egger M, et al. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57(3): 229–236.
- Wallach JD, Sullivan PG, Trepanowski JF, et al. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. *JAMA Intern Med* 2017; 177: 554–560.
- Kent D, Steyerberg E and van Klaveren D. Personalized evidence based medicine: Predictive approaches to heterogeneous treatment effects. *BMJ* 2018; 363: k4245.
- Kent D, van Klaveren D, Paulus J, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement: Explanation and elaboration. *Ann Intern Med* 2020; 172: W1–W25.
- van Klaveren D, Balan TA, Steyerberg EW, et al. Models with interactions overestimated heterogeneity of treatment effects and were prone to treatment mistargeting. *J Clin Epidemiol* 2019; 114: 72–83.
- Pellegrini F, Copetti M, Bovis F, et al. A proof-of-concept application of a novel scoring approach for personalized medicine in multiple sclerosis. *Mult Scler* 2020; 26(9): 1064–1073.
- Zhao L, Tian L, Cai T, et al. Effectively selecting a target population for a future comparative study. *J Am Stat Assoc* 2013; 108: 527–539.
- Bovis F, Carmisciano L, Signori A, et al. Defining responders to therapies by a statistical modeling approach applied to randomized clinical trial data. *BMC Med* 2019; 17: 113.
- Bovis F, Kalincik T, Lublin F, et al. Treatment response score to glatiramer acetate or interferon beta-1a. *Neurology* 2021; 96: e214–e227.
- Falet JP, Durso-Finley J, Nichyporuk B, et al. Estimating individual treatment effect on disability progression in multiple sclerosis using deep learning. *Nat Commun* 2022; 13(1): 5645.
- Eshaghi A, Young AL, Wijeratne PA, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun* 2021; 12: 2078.
- Chalkou K, Steyerberg E, Egger M, et al. A two-stage prediction model for heterogeneous effects of treatments. *Stat Med* 2021; 40: 4362–4375.

Visit SAGE journals online
[journals.sagepub.com/
home/msj](https://journals.sagepub.com/home/msj)

 SAGE journals