Subgroup Analysis European Special Interest Group

B. Bornkamp[1] B. Ratitch[2] K. Sechidis[1] D. Svensson (SIG Lead)[3]
on behalf of the Subgroup Analysis European Special Interest Group

[1] Novartis Pharma AG, Basel, Switzerland

[2] Bayer Inc., Mississauga, ON, Canada

[3] AstraZeneca, Gothenburg, Sweden

EFSPI Regulatory Statistics Workshop, Basel, 2024

Primary Focus

The main focus of the Subgroup Analysis SIG is on methods for assessing heterogeneity of treatment effects (HTE) across a population of patients. *Heterogeneity* can be defined as a nonrandom explainable variability in the direction and magnitude of individual treatment effects, including both beneficial and adverse effects [5]. Assessing the presence of treatment effect variability and explaining it by patient and disease characteristics may help to identify patient subgroups that benefit most from a given treatment and to guide the drug development processes and clinical care.
Typical goals of the HTE assessment are:

- Rigorously and comprehensively assessing available data for evidence against an a-priory assumption of treatment effect homogeneity
- Advancing understanding of the treatment's mechanism of action
- Supporting decisions on potential population enrichment or population changes in future clinical development studies
- Salvaging a failed study and/or mitigating an unexpected safety concern
- Supporting reimbursement applications
- Providing a comprehensive characterization of treatment effects in scientific publications
- Supporting development of precision medicine solutions

There are two main types of subgroup analysis:

- *Pre-specified subgroups*: Treatment effects can be evaluated across a (typically small) number of patient subgroups pre-specified in advance based on existing clinical knowledge or hypotheses (e.g., males vs. females).
- *Data-driven discovery*: HTE can be assessed in a manner where the analysis methodology is pre-specified but relevant patient subgroups and their characteristics are discovered in the process of data analysis.

The SIG is currently focused primarily on principled methods for data-driven subgroup analysis.

About the Special Interest Group

Creation and Affiliation

The SIG was formed in 2015 as an industry response to the new EMA ICHE14 Guideline on subgroups, which primarily focused on methodology and interpretation of regulatory consistency assessments, i.e., exploratory pre-specified subgroups. Since then, the area of HTE assessment underwent significant development, and the scope of the SA-SIG has correspondingly become wider.
The SA-SIG currently consists of ∼30 members from ∼11 companies/institutions.

Activities

The members of the SA-SIG meet approximately monthly to discuss methodological aspects, recent publications, share case studies, and listen to invited speakers. Members also self-organize into work streams to collaborate on publications, presentations, and research. Furthermore, SA-SIG organized a session dedicated to HTE in PSI 2021-2024, and in November 2021 a webinar dedicated to modern approaches to subgroup identification.

Topics of Interest

- **Processes** for setting objectives, engaging with stakeholders, analysis planning, data preparation, interpretation of findings, and reporting
- **Methods** for data-driven HTE discovery, inference, and robustness assessment
- **Software tools** and **practical applications**
- Identification of **methodological challenges** and pitfalls, **benchmarking** of methods

Inside SIG scope

a. Regulatory consistency assessment of pre-specified subgroups
b. Promoting theoretical understanding of methodology for data-driven subgroup identification;
c. Recommendations for practical ways of working with the data: how to best approach the assessment of HTE
d. Conducting joint work on e.g, benchmarking methodology via simulations
e. Following research advances on HTE also in other industries (such as econometrics/financial applications)
f. Inspiring statisticians/data scientists working in the HTE field

Methodological Challenges and Opportunities

- **Causal Inference** provides a framework for the underlying goal of identifying Conditional Average Treatment Effect (CATE) $\Delta(x_i) = E[Y_i(1) - Y_i(0)|X = x_i]$, i.e., the expected difference in outcomes Y had the patient i, characterized by baseline variables x_i, been treated with treatment 1 versus 0. The fundamental challenge is that in most settings we can observe only $Y_i(1)$ or $Y_i(0)$ but not both. One solution is to model and predict the unobserved outcomes using flexible, high-performing Machine Learning (ML) models f as building blocks $\Delta(x) = f(1, x) - f(0, x)$.
- **Multiple Hypothesis Testing** considerations for controlling false positive rate and quantification of uncertainty about findings are indispensable because the analysis involves multiple stages with many statistical decisions at each step. Resampling techniques for multiplicity control are more complex and computationally intensive in this setting.
- **"Honest" inference** is required to answer questions like: Is there any treatment effect heterogeneity? What is the magnitude of the treatment effect in discovered subgroups? What is the uncertainty about the estimates of CATE? Unbiased ("honest") inference is especially challenging if data-driven discovery and inference have to be done on the same data set.
- **Robustness of findings** needs to be assessed given that methods rely on stochastic algorithms and spurious results can be expected in finite samples. Important questions include: Does the method consistently choose the same variables as predictive of HTE? Does the method consistently classify patients into subgroups in the same way?
- **Semi-supervised Machine Learning** tackles a problem where the true target $\Delta(x)$ is unobservable, but the nuisance functions $f(1, x)$ and $f(0, 1)$ can be estimated from data. However, some ML aspects, e.g., model selection and overfitting control, become more challenging. Nevertheless, off-the-shelf ML methods can be tailored and made more robust for the task of HTE assessment.

Selected Publications and Presentations by SA-SIG Members

- A recent collaboration among many SA-SIG members and other experts involved resulted in a practical process-oriented guideline for a structured approach to of HTE assessment, including preparation of data, cross-functional communication with stakeholders, and interpretation of findings.
  K. Sechidis, S. Sun, Y Chen, J. Lu, C. Zhang, M. Baillie, D. Ohlssen, M. Vandemeulebroecke, R. Hemmings, S. Ruberg, B. Bornkamp, *WATCH: A Workflow to Assess Treatment Effect Heterogeneity in Drug Development for Clinical Trial Sponsors*, Under Review, [3].
- Another recent SA-SIG collaboration resulted in a tutorial-style overview of recent trends in the literature, covering how the methodology evolved since the previous tutorial paper [1]. This paper focus on causal inference based modelling for estimating CATE (Conditional Average Treatment Effect) and related Treatment Regimes, as well as overviewing such aspects as post-selection inference (e.g., so called nGATES inference, omnibus tests of heterogeneity, etc).
  I. Lipkovich, D. Svensson, B. Ratitch, A. Dmitrienko, *Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data*, Statistics in Medicine, 2024 [2].
- Another publication compares the various methods for assessing treatment effect heterogeneity, but also evaluates their performance in simulation scenarios that mimic real clinical trials. Furthermore it introduces an R package (benchtm), which can simulate synthetic biomarker distributions based on actual clinical trial data and create interpretable scenarios to benchmark methods for identifying and estimating treatment effect heterogeneity.
  S. Sun, K. Sechidis, Y. Chen, J. Lu, C. Ma, A. Mirshani, D. Ohlssen, M. Vandemeulebroecke, B. Bornkamp, Biometrical Journal, 2024, [4].

Method	Modeling type (1)	Application type (2)	Dimensionality (3)	Results type (4)	Inferential support (5)	Software (6)
Global outcome modeling						
Virtual Twins[51]	Freq/NP	RCT, OS	High	CATE, S	No	R [aVirtualTwins]
S-, T- X-learner[55]	Freq/SP, NP	RCT, OS	High	CATE, S	No	R [rlearner][a]
Global treatment effect modeling						
Interaction Trees[62]	Freq/NP	RCT	Medium	S	No	B
GUIDE[63]	Freq/NP	RCT	Medium	B,S	Yes	B
Model-based trees and forests[65]	Freq/NP	RCT	High	B, S	Yes	R [model4you]
Causal forests[71,72]	Freq/NP	RCT, OS	High	B, CATE	Yes[G, CATE]	R [grf]
Bayesian causal forests[160,230]	Bayes/NP	RCT, OS	High	CATE	Yes[CATE]	R [bcf, bartCausal]
Bayesian linear models[177,179]	Bayes/P	RCT	Low	CATE	Yes[CATE, S]	R [DSBayes, beanz]
Modified loss methods[78]	Freq/P, SP, NP	RCT, OS	High	CATE	No	R [personalized]
R-learner[83]	Freq/P, SP, NP	RCT, OS	High	CATE	No	R [rlearner][a]
Direct modeling of ITR						
AIPW estimator[96]	Freq/SP	RCT, OS	Medium	ITR	No	R [DynTxRegime]
OWL,[95] RWL,[100] AOL[101]	Freq/P, SP, NP	RCT, OS	High	ITR	No	R [DTRlearn2]
Tree-based ITR[99,116,118]	Freq/SP, NP	RCT, OS	Medium	ITR	No	R [T-RL,[b] policytree]
Direct subgroup identification						
SIDES and SIDEScreen[126,127]	Freq/NP	RCT	Medium	B, S	Yes[G, S]	B, R [SIDES, rsides]
TSDT[134]	Freq/NP	RCT	Medium	B, S	Yes[S]	R [TSDT]
PRIM[124]	Freq/NP	RCT	Medium	S	No	R [SubgrpID]
Sequential-Batting[131]	Freq/NP	RCT	Medium	S	Yes[S]	R [SubgrpID]
CAPITAL[120]	Freq/NP	RCT	Medium	S	No	R [policytree]
Bayesian Model Averaging[181]	Bayes/NP	RCT	Low	S	Yes[S]	R [subtee]

[a] Available at GitHub **xnie/rlearner**.
[b] Available at GitHub **Team-Wang-Lab/T-RL**.

Figure: Summary of ML methods that identify heterogeneous treatment effects using their key features as presented in [2].

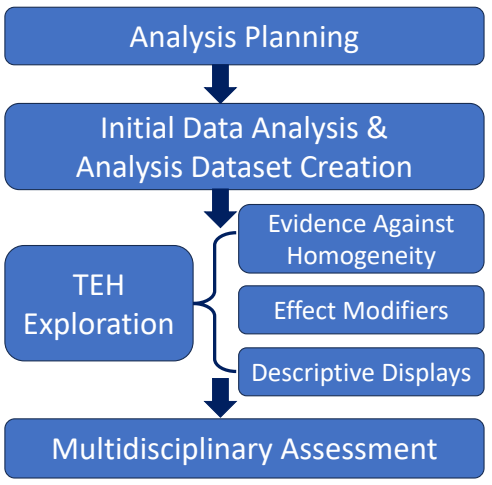

Figure: Overview of WATCH workflow [3] and the four main steps: (1) Analysis Planning, (2) Initial Data Analysis and Analysis Dataset Creation, (3) TEH Exploration, and (4) Multidisciplinary Assessment.

References

[1] I. Lipkovich, A. Dmitrienko, and R. B. D'Agostino Sr. "Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials". In: *Statistics in Medicine* 36 (2017), pp. 136–196.

[2] I. Lipkovich, D. Svensson, B. Ratitch, and A. Dmitrienko. "Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data". In: *Statistics in Medicine* n/a.n/a (2024).

[3] K. Sechidis, S. Sun, Y. Chen, J. Lu, C. Zang, M. Baillie, D. Ohlssen, M. Vandemeulebroecke, R. Hemmings, S. Ruberg, and B. Bornkamp. "WATCH: A Workflow to Assess Treatment Effect Heterogeneity in Drug Development for Clinical Trial Sponsors". In: *Under review* (2024). DOI: https://arxiv.org/abs/2405.00859.

[4] S. Sun, K. Sechidis, Y. Chen, J. Lu, C. Ma, A. Mirshani, D. Ohlssen, M. Vandemeulebroecke, and B. Bornkamp. "Comparing algorithms for characterizing treatment effect heterogeneity in randomized trials". In: *Biometrical Journal* 66.1 (2024), p. 2100337.

[5] R. Varadhan, J. Segal, C. Boyd, A. Wu, and C. Weiss. "A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research.". In: *J Clin Epidemiol.* 66(8) (2013), pp. 818–825.