WATCH: a Workflow for Assessing Treatment effeCt Heterogeneity in drug development for clinical trial sponsors

Konstantinos (Kostas) Sechidis
Associate Director of Data Science
Novartis

PSI 2024 Conference, Amsterdam
June 19th, 2024

NOVARTIS | Reimagining Medicine

arXiv > stat > arXiv:2405.00859

Statistics > Applications

[Submitted on 1 May 2024]

WATCH: A Workflow to Assess Treatment Effect Heterogeneity in Drug Development for Clinical Trial Sponsors

Konstantinos Sechidis, Sophie Sun, Yao Chen, Jiarui Lu, Cong Zang, Mark Baillie, David Ohlssen, Marc Vandemeulebroecke, Rob Hemmings, Stephen Ruberg, Björn Bornkamp

This paper proposes a Workflow for Assessing Treatment effeCt Heterogeneity (WATCH) in clinical drug development targeted at clinical trial sponsors. The workflow is designed to address the challenges of investigating treatment effect heterogeneity (TEH) in randomized clinical trials, where sample size and multiplicity limit the reliability of findings. The proposed workflow includes four steps: Analysis Planning, Initial Data Analysis and Analysis Dataset Creation, TEH Exploration, and Multidisciplinary Assessment. The workflow aims to provide a systematic approach to explore treatment effect heterogeneity in the exploratory setting, taking into account external evidence and best scientific understanding.

Many we have been in similar situation ...

The trial failed! We needed 80% efficacy, and it's only 70% ...

... but for women over 50, the efficacy is 95%. Can we just give it to them?

Example inspired by a talk of Prof Richard Samworth

Subgroup analysis in clinical trials

Finding subgroups is the “hardest problem there is” (Stephen Ruberg)

- High chances of false negatives due to insufficient sample size
  Cinical trials not designed for assessing subgroup treatment effects or testing interactions (underpower).
- High chances of false positives due to multiplicity
  Performing multiple comparisons on unreliable or noisy subgroup treatment effects and selecting “the best” can introduce bias (selective inference).

Traditionally in drug development the term subgroup analysis is used for investigations related Treatment Effect Heterogeneity (TEH).

Issues with Replicability of Subgroup Findings

Article

July 3, 1991

Analysis and Effects in Sub Clinical Trials

Salim Yusuf, DPhil, MRCP; Janet

» Author Affiliations

JAMA. 1991;266(1):93-98. doi:10

Table 2.—Claims of Subgroup Effects on Mortality in the 65 Randomized Trials of β-Blockers in Acute Myocardial Infarction*

Study	Subgroup Benefit Claimed	Prior Hypothesis	Confirmed in Other Trials	Overall *P* Value	Test for Heterogeneity	Correction for Multiplicity	*P*
			Early Initiation of Treatment				
1. Barber et al[20]	Tachycardia at entry >100 beats per min	No	No	Not significant	–	–	–
2. MIAMI[21]	"High-risk" patients	No	No	NS	–	–	–
			Late Initiation of Treatment				
3. Anderson et al[22]	Treatment beneficial in patients <65 y and harmful in those >65 y	Unclear	No, most trials show similar reductions in relative risk among younger and older patients.[23]	NS	–	–	–
4. Hjalmarson et al[24]	Benefit observed only in patients with HR >65 beats per min (never formally published)	No	No, the MIAMI trial included only this group. Overall results were not significant. The impact of HR on effect of treatment was tested in ISIS-1.[25] No differential was found.	<.03	–	–	–
5. Wilhelmsson et al[26]	Benefit only in patients with "electrical" or "mechanical" complications	No	No, although one other study[27] observed a similar result. Many other studies and the Beta-Blocker Pooling Project[28] failed to identify this subgroup as benefiting preferentially.	NS	–	–	–
6. Multicenter International[29]	Benefit only in patients with anterior MI before entry	No	No	<.08	–	–	–
7. Taylor et al[30]	Benefit only among those with treatment initiated within 6 mo of MI, while those treated later appeared harmed	No	No	NS	–	–	–
8. Beta-blocker Heart Attack Trial[27]	Benefit only in patients with "electrical" or "mechanical" complications prior to randomization	No	Not consistently	<.003	–	–	+
9. Yusuf et al[23] (pooled data)	β-blockers without ISA more effective than those with ISA	No	Uncertain. Three new trials appear to contradict this conclusion. A trial of metoprolol was unpromising. Two studies, one of acebutolol and one of oxprenolol, both with ISA, were favorable	<.0001	+(*P*<.02)	–	–

*Only subgroups that are "proper" are included. HR indicates, heart rate; MI, myocardial infarction; and ISA, intrinsic sympathomimetic activity.

Issues with Replicability of Subgroup Findings

Original Investigation

April 2017

Evaluati
and Cor
Random

Joshua D. Wallach, [illegible]
PhD[1]; Ewout W. Ste[illegible]

» Author Affiliations | Article Information

JAMA Intern Med. 2017;177(4):554-560. doi:10.1001/jamainternmed.2016.9125

RESULTS Sixty-four eligible RCTs made a total of 117 subgroup claims in their abstracts. Of these 117 claims, only 46 (39.3%) in 33 articles had evidence of statistically significant heterogeneity from a test for interaction. In addition, out of these 46 subgroup findings, only 16 (34.8%) ensured balance between randomization groups within the subgroups (eg, through stratified randomization), 13 (28.3%) entailed a prespecified subgroup analysis, and 1 (2.2%) was adjusted for multiple testing. Only 5 (10.9%) of the 46 subgroup findings had at least 1 subsequent pure corroboration attempt by a meta-analysis or an RCT. In all 5 cases, the corroboration attempts found no evidence of a statistically significant subgroup effect. In addition, all effect sizes from meta-analyses were attenuated toward the null.

Exploratory assessment of the TEH is important

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

31 January 2019
EMA/CHMP/539146/2013
Committee for Medicinal Products for Human Use (CHMP)

Guideline on the investigation of subgroups in confirmatory clinical trials

STATISTICAL PERSPECTIVES ON SUBGROUP ANALYSIS: TESTING FOR HETEROGENEITY AND EVALUATING ERROR RATE FOR THE COMPLEMENTARY SUBGROUP

Mohamed Alosh[1], Mohammad F. Huque[2], and Gary G. Koch[3]
[1]*Division of Biometrics III, Office of Biostatistics, OTS, CDER, FDA, Silver Spring, Maryland, USA*
[2]*Office of Biostatistics, OTS, CDER, FDA, Silver Spring, Maryland, USA*
[3]*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*

AN OVERVIEW OF STATISTICAL AND REGULATORY ISSUES IN THE PLANNING, ANALYSIS, AND INTERPRETATION OF SUBGROUP ANALYSES IN CONFIRMATORY CLINICAL TRIALS

Robert Hemmings
Medicines and Healthcare Products Regulatory Agency, London, United Kingdom

" ... ignoring the problem, and similarly routinely dismissing results of subgroup analysis, is no scientific solution."

Exploratory assessment of the TEH is important,
... but very challenging

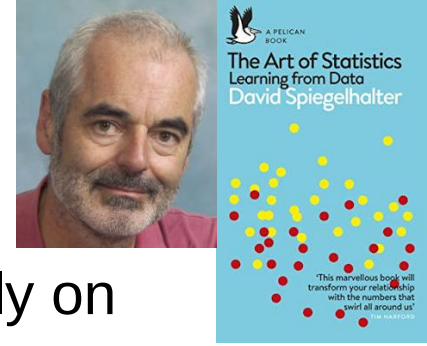
Data alone don't speak for "themselves" on TEH

- If study was not planned for TEH investigation (which is often the case)
- Need to take external information into account: historical trials from similar drugs in same indication or same drug in other indication (replication), pre-clinical, mechanistic understanding, clinical information (biological plausibility).

Without external replication or plausibility,
... data-based findings alone very speculative

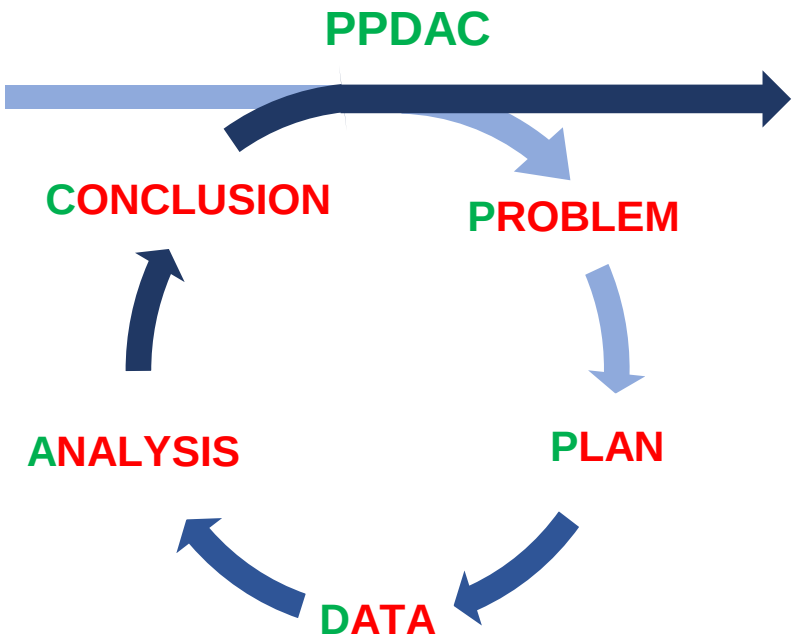
We need a systematic framework that considers all these issues

A systematic approach for exploratory projects



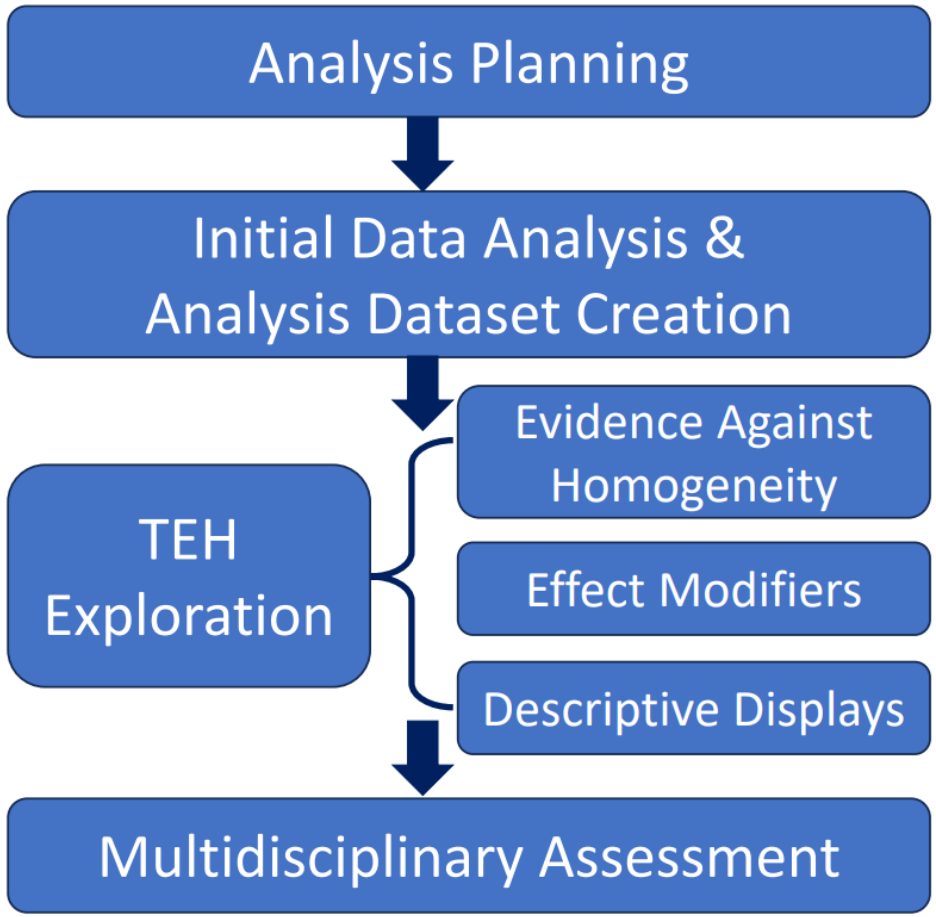
- Traditionally the focus of many researchers and analysts has been primarily on the "**ANALYSIS**" stage.

- However, to truly get value for our projects, we should approach them holistically following some structured frameworks like **PPDAC**



This systematic approach fosters transparency, reducing misunderstanding or errors, and allowing better replication of the findings.

Workflow for Assessing Treatment effeCt Heterogeneity - WATCH



arXiv > stat > arXiv:2405.00859

Statistics > Applications

[Submitted on 1 May 2024]

WATCH: A Workflow to Assess Treatment Effect Heterogeneity in Drug Development for Clinical Trial Sponsors

Konstantinos Sechidis, Sophie Sun, Yao Chen, Jiarui Lu, Cong Zang, Mark Baillie, David Ohlssen, Marc Vandemeulebroecke, Rob Hemmings, Stephen Ruberg, Björn Bornkamp

This paper proposes a Workflow for Assessing Treatment effeCt Heterogeneity (WATCH) in clinical drug development targeted at clinical trial sponsors. The workflow is designed to address the challenges of investigating treatment effect heterogeneity (TEH) in randomized clinical trials, where sample size and multiplicity limit the reliability of findings. The proposed workflow includes four steps: Analysis Planning, Initial Data Analysis and Analysis Dataset Creation, TEH Exploration, and Multidisciplinary Assessment. The workflow aims to provide a systematic approach to explore treatment effect heterogeneity in the exploratory setting, taking into account external evidence and best scientific understanding.

Biometrical Journal

RESEARCH ARTICLE | Full Access

Comparing algorithms for characterizing treatment effect heterogeneity in randomized trials

Sophie Sun, Konstantinos Sechidis, Yao Chen, Jiarui Lu, Chong Ma, Ardalan Mirshani, David Ohlssen, Marc Vandemeulebroecke, Björn Bornkamp

First published: 27 November 2022 | https://doi.org/10.1002/bimj.202100337

Pharmaceutical Statistics
The Journal of Applied Statistics in the Pharmaceutical Industry

MAIN PAPER | Full Access

Predicting subgroup treatment effects for a new study: Motivations, results and learnings from running a data challenge in a pharmaceutical corporation

Björn Bornkamp, Silvia Zaoli, Michela Azzarito, Ruvie Martin, Carsten Philipp Müller, Conor Moloney, Giulia Capestro, David Ohlssen, Mark Baillie

First published: 07 February 2024 | https://doi.org/10.1002/pst.2368

benchtm

Analysis Planning

Gather background information on study, drug and disease

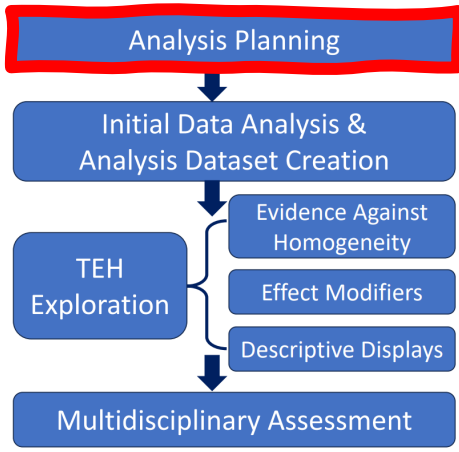
- ✓ Protocol(s), SAP, scientific publications

Involve and align with stakeholders & subject matter experts on

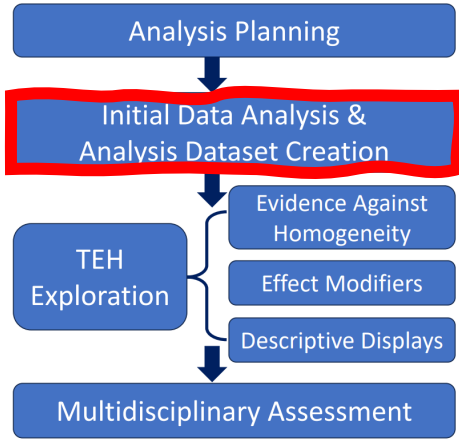
- ✓ Outcome variables
- ✓ Studies to include
- ✓ Baseline variables/biomarkers to include
- ✓ What is TEH (scale) we are interested in, e.g.
  - treatment effect of drug vs placebo, or
  - between different dosages

A-priori evidence for treatment effect modification

- for each variable, document the level of external evidence using categories: *none*, *low*, *moderate*, *high*



Initial data analysis (IDA)



Explore the data without approaching main analysis question

- ✓ Univariate summaries of baseline variables/biomarkers (eg skewness, missingness, information, ...)
- ✓ Dependencies across included baseline variables (eg correlated variables, duplicates, ...)

IDA provides a systematic workflow for researchers to work with data responsibly

- ✓ "Ten simple rules for initial data analysis." Mark Baillie et al., PLOS Computational Biology, 2022


PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

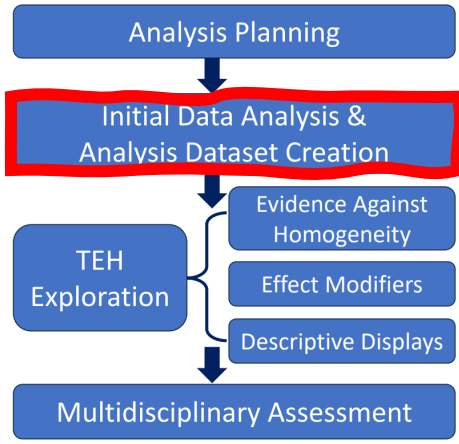
EDITORIAL

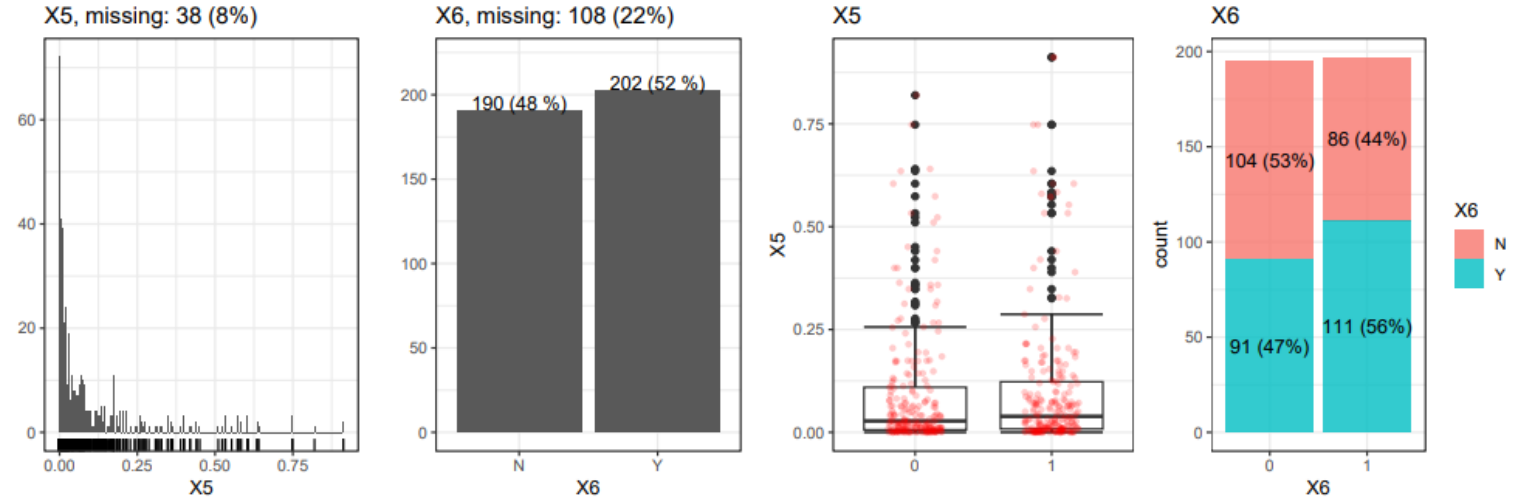
Ten simple rules for initial data analysis

Mark Baillie, Saskia le Cessie, Carsten Oliver Schmidt, Lara Lusa, Marianne Huebner,
for the Topic Group "Initial Data Analysis" of the STRATOS Initiative

Published: February 24, 2022 • https://doi.org/10.1371/journal.pcbi.1009819

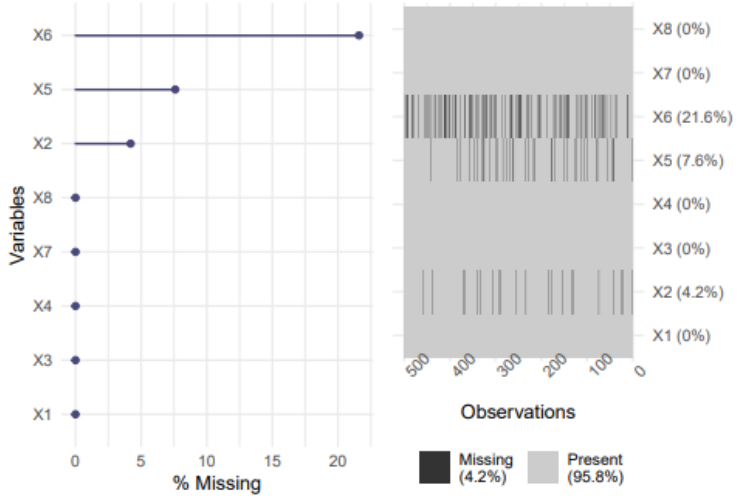
Initial data analysis (IDA)

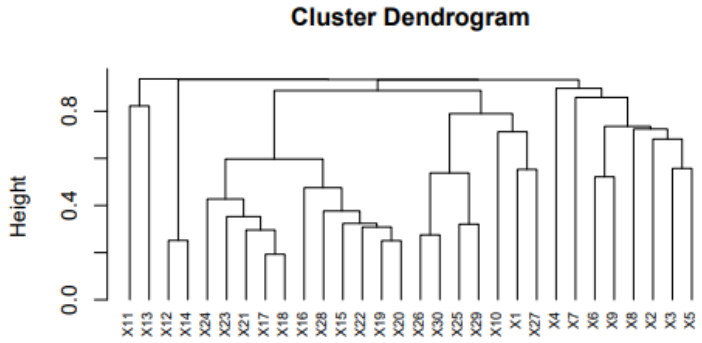




(a) Univariate summaries

(b) Variables stratified by further factors





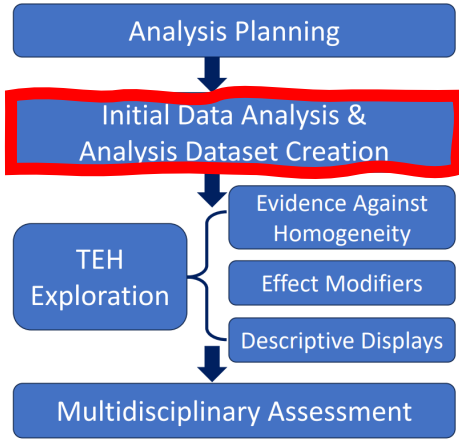
(c) Missing values

(d) Dependence analysis

Analysis Dataset Creation

Preprocess the data based on IDA

- ✓ Transform baseline variables (when there is skweness)
- ✓ Exclude baseline variables (e.g. high missingness, low information, ...)
- ✓ Merge sparse categories
- ✓ Choose between highly correlated variables (e.g. BMI or weight)
- ✓ Impute remaining missing baseline variables; for outcome variable follow estimand strategy



Explore Treatment Effect Heterogeneity (TEH)

Using stats/ML modelling to answer three questions:

- ✓ **Evidence against homogeneity**

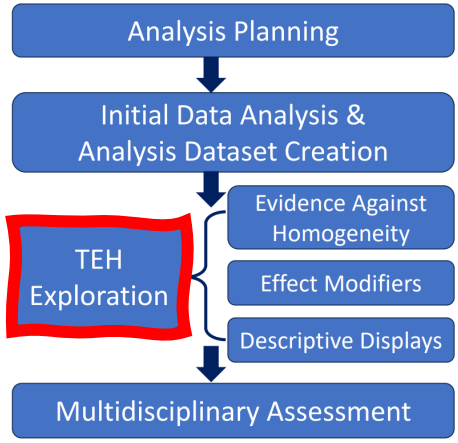
  Question 1: How strong is the overall evidence against the null hypothesis (homogeneous effect)?

- ✓ **Effect modifiers**

  Question 2: Which variables drive heterogeneity?

- ✓ **Exploratory displays**

  Question 3: How does the treatment effect change for the identified variables?



Biometrical Journal

RESEARCH ARTICLE | Full Access

Comparing algorithms for characterizing treatment effect heterogeneity in randomized trials

Sophie Sun, Konstantinos Sechidis, Yao Chen, Jiarui Lu, Chong Ma, Ardalan Mirshani, David Ohlssen, Marc Vandemeulebroecke, Björn Bornkamp

First published: 27 November 2022 | https://doi.org/10.1002/bimj.202100337

Explore Treatment Effect Heterogeneity (TEH)

Using stats/ML modelling to answer three questions:

- ✓ **Evidence against homogeneity**

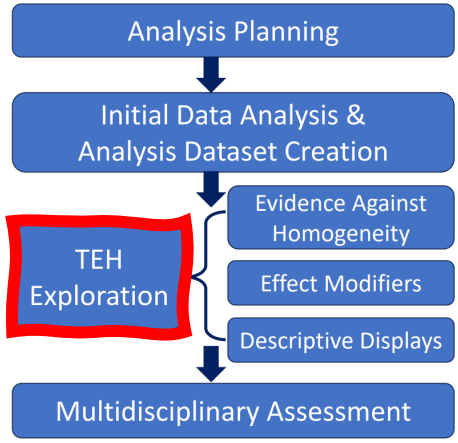
  Question 1: How strong is the overall evidence against the null hypothesis (homogeneous effect)?

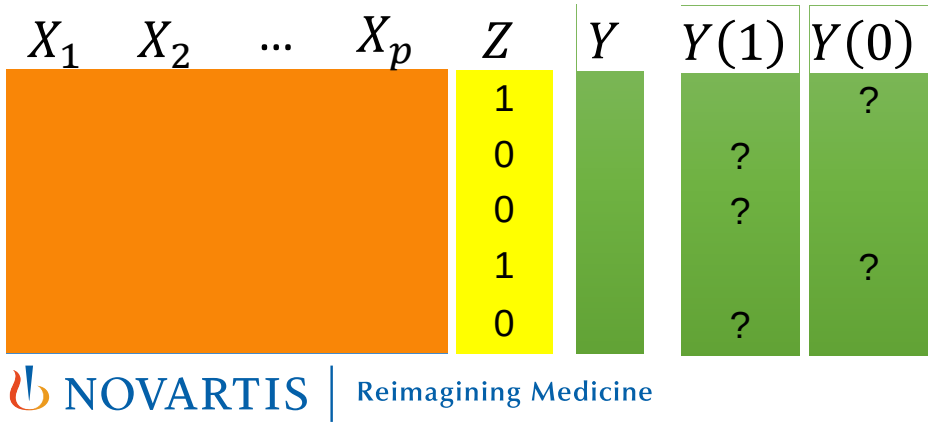
- ✓ **Effect modifiers**

  Question 2: Which variables drive heterogeneity?

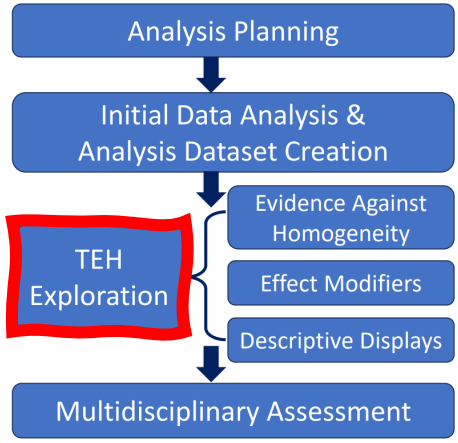
- ✓ **Exploratory displays**

  Question 3: How does the treatment effect change for the identified variables?





Explore Treatment Effect Heterogeneity (TEH)



Using stats/ML modelling to answer three questions:

- ✓ **Evidence against homogeneity**

  Question 1: How strong is the overall evidence against the null hypothesis (homogeneous effect)?

- ✓ **Effect modifiers**

  Question 2: Which variables drive heterogeneity?

- ✓ **Exploratory displays**

  Question 3: How does the treatment effect change for the identified variables?

X_1	X_2	...	X_p	Z	Y	$Y(1)$	$Y(0)$	$Y(1)-Y(0)$
				1			?	?
				0		?		?
				0		?		?
				1			?	?
				0		?		?

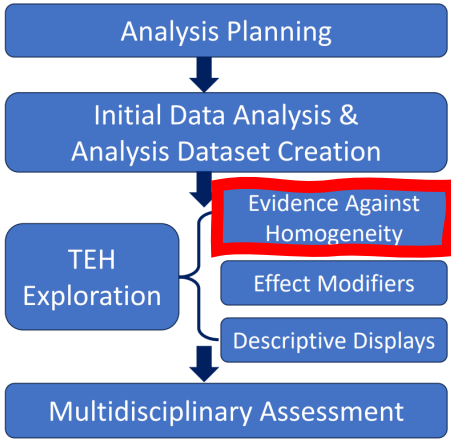
Our approach: create a standard prognostic modelling problem by using a doubly robust estimator to obtain pseudo-observations for the treatment effect of each patient.

How strong is the evidence against homogeneity?

- Clinical trials are planned in a specific population, and there is prior expectation of consistent effect across that population (homogeneity)
- We use a global test for testing the null hypothesis of homogeneity:

$$\boldsymbol{X} \perp \left(\boldsymbol{Y}(\mathbf{1}) - \boldsymbol{Y}(\mathbf{0})\right)$$

- Interpret p-value on a continuous scale not as a binary decision rule
- p-value as measure of surprise!



JOURNAL ARTICLE EDITOR'S CHOICE

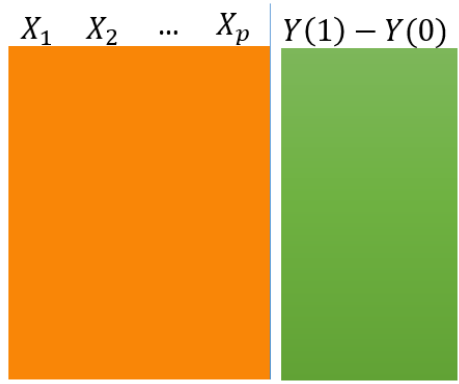
Surprise! FREE

Stephen R Cole ✉, Jessie K Edwards, Sander Greenland

American Journal of Epidemiology, Volume 190, Issue 2, February 2021, Pages 191–193, https://doi.org/10.1093/aje/kwaa136

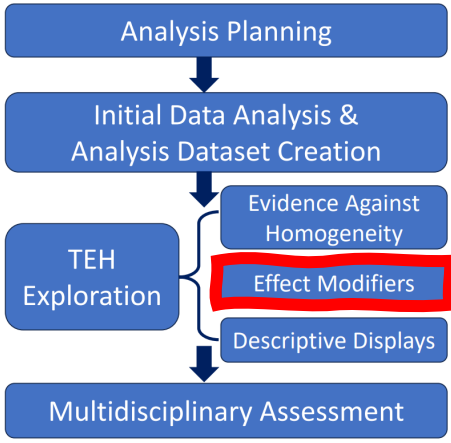
p-value = 0.063 ⇒  x4

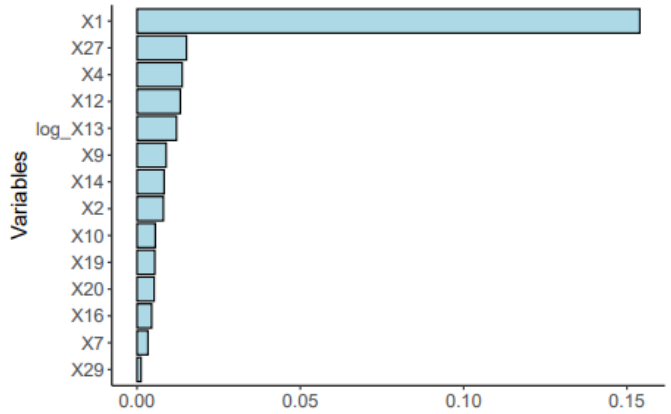
p-value = 0.001 ⇒  x10



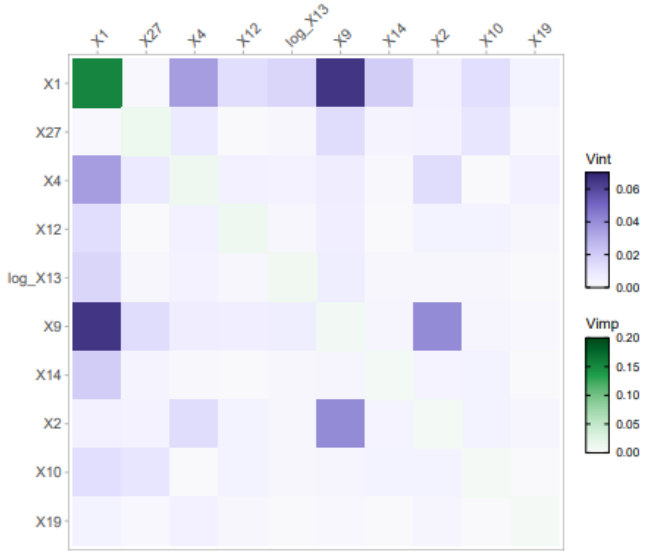
Which variables drive heterogeneity?

- Identifying effect modifiers is critical towards assessing heterogeneity.
- They define subpopulations of heterogeneous treatment effect.
- We provide variable importance scores that captures how strongly each variable modifies the treatment effect.

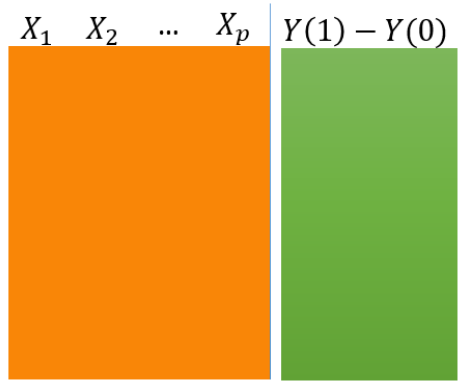




(a) Permutation Variable importance

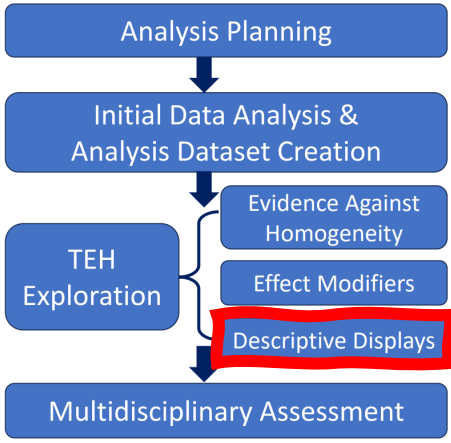


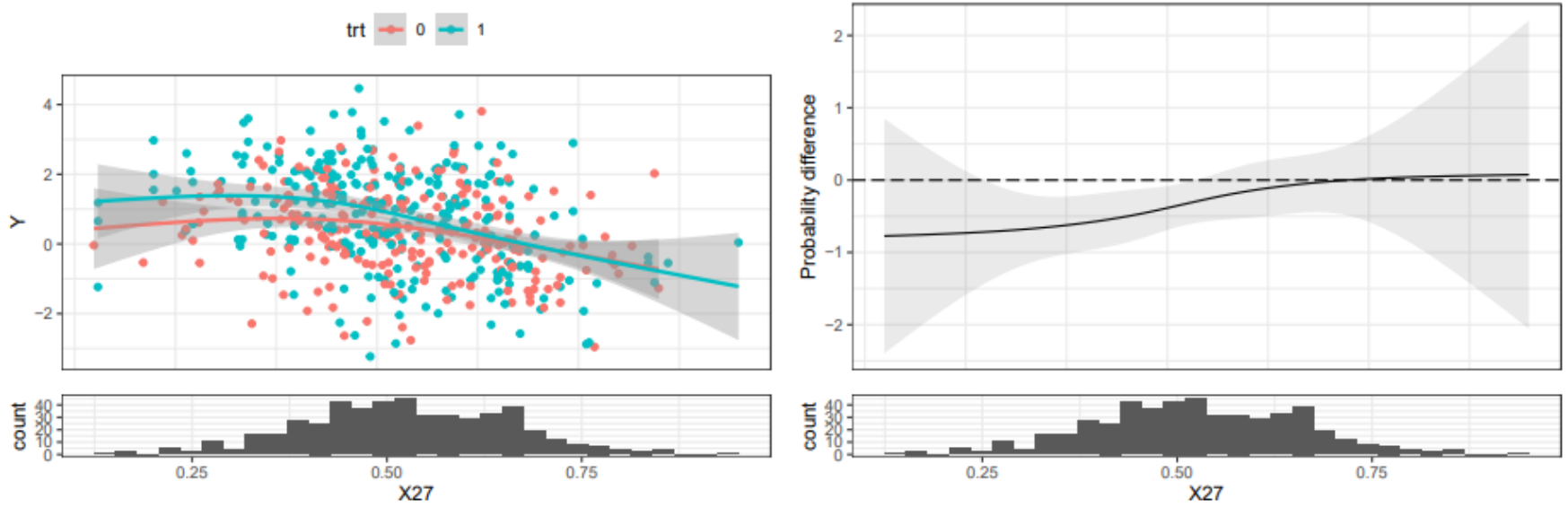
(b) Interaction variable importance



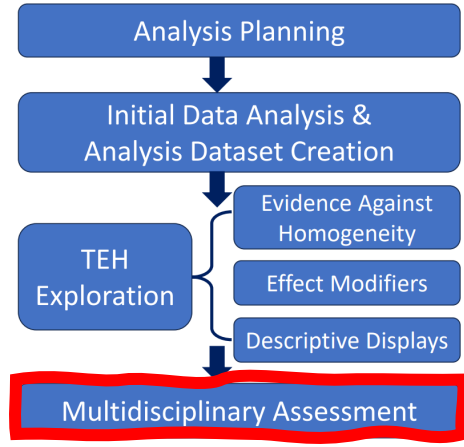
How the treatment effect changes ... for the identified variables?

- Present visualizations of how the treatment effect changes with respect to the identified effect modifiers.
- Starting point for discussions about heterogeneity with the clinical team.
- Methods to determine cut-off subgroups could also be used (not in the first step, only after initial discussion).





Multidisciplinary Assessment



- Background information is important to assess credibility of findings
- Present outputs to cross-functional team to assess findings credibility
- Avoid strong confirmatory language (and discussion around treatment effects etc); emphasize need for interpretation taking into account a-priori or external evidence
- What could be next steps?
  - None
  - Analyses to explain specific unexpected results or findings
  - Update analysis with additional variables; additional endpoints
  - Team may be interested in a subgroup
    (need to utilize corrected estimates of efficacy in identified subgroups)

Conclusions


arXiv > stat > arXiv:2405.00859

Search... Help |

Statistics > Applications

[Submitted on 1 May 2024]

WATCH: A Workflow to Assess Treatment Effect Heterogeneity in Drug Development for Clinical Trial Sponsors

Konstantinos Sechidis, Sophie Sun, Yao Chen, Jiarui Lu, Cong Zang, Mark Baillie, David Ohlssen, Marc Vandemeulebroecke, Rob Hemmings, Stephen Ruberg, Björn Bornkamp

This paper proposes a Workflow for Assessing Treatment effeCt Heterogeneity (WATCH) in clinical drug development targeted at clinical trial sponsors. The workflow is designed to address the challenges of investigating treatment effect heterogeneity (TEH) in randomized clinical trials, where sample size and multiplicity limit the reliability of findings. The proposed workflow includes four steps: Analysis Planning, Initial Data Analysis and Analysis Dataset Creation, TEH Exploration, and Multidisciplinary Assessment. The workflow aims to provide a systematic approach to explore treatment effect heterogeneity in the exploratory setting, taking into account external evidence and best scientific understanding.

- ✓ Understanding how treatment effect varies across patients may influence important sponsor decisions
- ✓ **WATCH** provides a systematic approach to explore TEH considering external evidence and best scientific understanding.
- ✓ Ongoing internal work in various project, to answer various questions, such as:
  - ✓ identify populations with *differential* treatment effect
  - ✓ identify populations that response better to one dose versus the other
  - ✓ identify populations that manifest specific adverse events
  - ✓ explain multiregional differences (see E17 ICH guideline)

Thank you

Kostas Sechidis

kostas.sechidis@novartis.com

NOVARTIS | Reimagining Medicine