Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Preventive Medicine
Grant Writing Seminar Series
Session 8

Data analysis strategies and how they can enhance the proposal

Fridtjof Thomas, PhD
Professor
Division of Biostatistics, Department of Preventive Medicine
College of Medicine
The University of Tennessee Health Science Center

Dec 01, 2023

1

1

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

About the presenter

- Professor at the Division of Biostatistics, Dept. of Preventive Medicine
- At UTHSC since 2007
- NIH HEAL Data Stewardship Group that specifically works with assuring that the nationwide hundreds of HEAL-projects comply to the NIH Policy for Data Management and Sharing that has taken effect Jan. 25, 2023. (HEAL = Helping to End Addiction Long-term Initiative; heal.nih.gov).
- Design and Analysis Committee of the EARLY trials (2010-2016 – "Early Adult Reduction of weight through LifestYle intervention," a collection of seven randomized clinical trials funded by the National Heart, Lung, and Blood Institute (NHLBI) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH))
- Member of the Biostatistics Collaborative Core at the Southeast Regional Center of the NIH-NHLBI-funded Women's Health Initiative (WHI) study that has recruited over 160,000 women in over 40 clinical centers nationwide. (2010-2017)
- Grant review experience since 2012 from
  - Department of Defense's Congressionally Directed Medical Research Program (DoD CDMRP)
  - NIH Epidemiology of Chronic and Infectious Disease Study Section
  - NIH Neurological, Aging, and Musculoskeletal Epidemiology (NAME) Study Section
- 14+ years Associate Editor of *The Journal of Statistical Computation and Simulation* (JSCS; a Taylor & Francis print journal since 1972)

2

2

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Network of collaboration (Elsevier Pure "Fingerprint" 10/09/2019)

3

3

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Outline

- What makes or breaks a proposal?
- Outcomes, Missing Data, Surrogate endpoints, Adverse events
- Intend-to-treat analysis (ITT)
- Clustering/grouping of observations
- Pre-planned vs. ad hoc subgroup analyses
- Heterogeneity of treatment effects (HTE) with respect to sex/gender and race
- Data Management and Sharing plan (DMS)

4

4

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

What makes or breaks a proposal?

- Most importantly:
  - Are you using **sound science**?
  - What is your **innovation**?
- The analytical plan cannot safe your proposal, but it will sink it in a hurry if not put together thoughtfully
- Reviewers need to find "objective reasons" why they don't like a proposal: the fewer targets the analytical plan offers for that, the better
- Make sure that
  - reviewers easily find what they are looking for
  - the analytical plan is connected to your research question
  - measures (including time points) are consistent throughout
- It is difficult to get funded!
- It is not all that difficult to have a better proposal than most I have seen when reviewing grants – but that takes time and effort!

5

5

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

How strong will your derived evidence be? Levels of evidence

1a	Systematic review of high quality RCTs with similar results and effect sizes for many different RCTs.
1b	Individual high quality RCT with high precision (narrow confidence interval)
1c	All or none
2a	Systematic review of cohort studies with similar results and effect sizes.
2b	Individual cohort study or low quality RCT (e.g., <80% follow-up)
2c	"Outcomes Research" and ecological studies (based on average exposures etc. of populations of geographical or temporal units)
3a	Systematic review of case-control studies
3b	Individual case-control study
4	Case-series and poor quality cohort and case-control studies
5	Expert opinion (unless critically appraised or based on "first principles")

Source: Oxford Centre for Evidence-based Medicine https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

6

6

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

All or none: Example "Bubble Boy" disease

- Babies born without functional immune system.
- SCID-X1: 1 in 50,000-100,000 affected; caused by a mutation in a gene (IL2RG)
- Most die within first year of life. (Only about 20% have access to a suitable sibling for a bone-marrow transplant as the existing cure.)

St. Jude announced April 18, 2019: Gene therapy cure for babies with X-linked severe combined immunodeficiency

"The gene therapy, produced in the Children's GMP, LLC, manufacturing facility on the St. Jude campus, involved use of a virus to transport and insert a correct copy of a gene into the genome of patients' blood stem cells. Following the treatment, the children began producing functioning immune cells for the first time, according to St. Jude, and most patients were discharged from the hospital within one month." [All 8 babies started to produce complete sets of immune cells.]

https://www.stjude.org/inspire/news/bubble-boy-scid-x1-cure.html
https://www.stjude.org/research/news-publications/research-highlights/2019-research-highlights/st-jude-gene-therapy-holds-promise-for-treating-several-diseases.html

7

7

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Outcomes

- Outcomes: Be ambitious but realistic
  - pilot studies go a long way!
  - An R21 is not an underfunded R01
- Missing Data:
  - Missing covariate information vs. missing endpoints
  - 20% attrition in behavioral intervention studies might be acceptable, 10% is better – requires work to achieve!
  - Differential loss to follow-up needs to be addressed in the analytical plan
- Surrogate endpoints: e.g., progression free survival
- Adverse events:
  - Always occur and need to be reported/summarized (may or may not be related to the intervention)
  - How do adverse events impact on your outcome/collection of measures?

8

8

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Missing Data

- Keep it to a minimum
- Mention the word "multiple imputation", know what it means, and have a realistic approach to it
- Complete case analysis or "last observation carried forward" doesn't cut it in most cases
- Include sensitivity analyses

9

9

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Bias: What do we mean by bias?

- Statistical theory (not further covered here)
  - Sample variation
  - Estimation: if expected value = true value, estimator is unbiased
  - Asymptotically unbiased if value converges to true value if $n \to \infty$
  - Bias-variance trade-off in prediction
- Publication bias (not further covered here)
- Bias due to "distorting" true relationships: https://catalogofbias.org/ (CEBM/University of Oxford)
  - E.g., confounding incl. "confounding by indication"
  - E.g., selection bias (incl. "immortal time bias")
  - E.g., measurement bias/assessment bias
  - (…)
- Bias analysis (E-values! - VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017 Aug 15;167(4):268-274. doi: 10.7326/M16-2607. Epub 2017 Jul 11. PMID: 28693043.)

10

10

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

"Distorting" true relationships

What is a "true relationship"?

Example: What is the true odds ratio (OR) for a specific exposure and a specific event?

Operational definition: The true OR is the odds ratio that would be observed in a perfectly executed randomized clinical trial (RCT) that is large enough to make sample variation practically unimportant.

11

11

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

"Distorting" true relationships (cont.)

The following aspects have to be addressed:

- Selection bias
- Measurement bias (misclassification – outcome or covariate; exposure misclassification; systematically missing activities/episodes, e.g., in activity data; recall bias, telescoping bias, etc.)
- Blinding/masking of evaluators
- Cases and non-cases/controls need identical ways to determine covariates and outcomes!
- "Immortal time bias" in time-to-event analyses
- Differential loss-to-follow-up
- "Artefacts" due to utilizing existing data for a different purpose (e.g., billing data; medical prescription data to determine adherence to medication)

12

12

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

"Distorting" true relationships (cont.)

Related to data analysis:

- Collection/extraction of data
- Should the analyst be blinded to treatment group? ("Group A" vs "Group B" instead of explicit "Treatment" and "Control")
- Unintentional programming errors
- Undetected problems with the convergence of computational algorithms
- Validation of the data incl. approaches to unusual observations
- "fishing expeditions" (and multiplicity in testing in general)
- Any form of "P-value maximization approaches", such as
  - thresholds for continuous variables to achieve "maximal significance"
  - picking definitions for events, exposure measures, etc. based on resulting "significance"
  - etc., etc.,…

Make sure your statistical plan does not raise these concerns!

13

13

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Intend-to-treat analysis (ITT)

- Evaluation of participants as randomized
- Protects against bias due to, e.g.,
  - differential loss to follow-up
  - differential adherence to the treatment
- Breaks down if missing data is present!
- Often combined with an "per-protocol" or "as-treated" analysis (see Hulley et all ch. 11)

14

14

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Clustering/grouping of observations

- Is there any grouping structure you should account for?
- Group/cluster randomization? (Contamination of groups; not practical to individually randomize)
- Adjustment in other grouping structures (e.g., physician office even when individually randomized) that can/should be addressed in the analysis? (Often in form of mixed effects models with grouping structure being a random effect)

15

15

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Pre-planned vs. ad hoc subgroup analyses

- Ad hoc analyses will be made and should be acknowledged (e.g., when safety issues arise during the trial)
- Pre-plan what is important to you and adjust for the multiplicity in testing accordingly
- Think about details, e.g.:
  - ANOVA with post-hoc analysis using Tukey's honest significant difference
  - Dunnett's test to adjust for a single comparison/control group but several active intervention groups

16

16

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Fishing expeditions

"Throw everything in the kitchen sink to the wall and see what sticks"

Here is what can stick:

Facts about Abraham Lincoln and John Fitzgerald Kennedy and their assassinations in 1865 and 1963, respectively. Jones and Muirhead (2012) found the following:

1. They were elected 100 years apart (1860 and 1960)
2. They were both assassinated on a Friday in the presence of their wives
3. Lincoln was shot in Ford's Theatre; Kennedy was shot in a Ford car.
4. Both assassins were known by three names – John Wilkes Booth and Lee Harvey Oswald, with 15 letters in each complete name.
5. Booth shot Lincoln in a theatre and fled to a warehouse; Oswald shot Kennedy from a warehouse and fled to a theatre.
6. Both succeeding vice-presidents named Johnson (Andrew and Lyndon), with 13 letters in their names and born 100 years apart (1808 and 1908).

Can this be coincidence?
See Jones and Muirhead (2012):
https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2012.00545.x

17

17

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

You can easily do better!

Follow this principle to avoid the worst pitfalls:

"Draw your assumptions before your conclusions."

Miguel Hernán, Prof. of Biostatistics and Epidemiology, Harvard CAUSALab

18

18

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Heterogeneity of Treatment Effects (HTE)

- Race and gender! (Unless you have an acceptable reason for not considering these variables).
- NIH: "The purpose of this Notice is to inform the research community that NIH is amending its NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research to include a requirement that recipients conducting applicable NIH-defined Phase III clinical trials ensure results of valid analyses by sex/gender, race, and/or ethnicity are submitted to Clinicaltrials.gov." https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html
- FDA: Evaluation of Sex-Specific Data in Medical Device Clinical Studies https://www.fda.gov/files/medical%20devices/published/Evaluation-of-Sex-Specific-Data-in-Medical-Device-Clinical-Studies---Guidance-for-Industry-and-Food-and-Drug-Administration-Staff.pdf
- Patient-Centered Outcomes Research, especially PCORI
- PCORI Methodology Standards
- HTE:
  - How are HTEs assessed?
  - Should there be adjustments for multiplicity in testing?
  - Which type of subgrouping variables?
  - Should a HTE-analysis be reported even when the overall treatment effect is not statistically significant?
  - When reported, what should be reported and why?

19

19

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Heterogeneity of Treatment Effects (HTE)

"William Osler c1912" by Unknown - [1]. Licensed under CC BY 4.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:William_Osler_c1912.jpg#mediaviewer/File:William_Osler_c1912.jpg

"Variability is the law of life, and as no two faces are the same, so no two bodies are alike, and no two individuals react alike and behave alike under the abnormal conditions which we know as disease"

Sir William Osler
1849 – 1919; Canadian physician and "Father of modern Medicine"

20

20

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Heterogeneity of Treatment Effects (HTE)

FDA: Evaluation of Sex-Specific Data in Medical Device Clinical Studies https://www.fda.gov/files/medical%20devices/published/Evaluation-of-Sex-Specific-Data-in-Medical-Device-Clinical-Studies---Guidance-for-Industry-and-Food-and-Drug-Administration-Staff.pdf

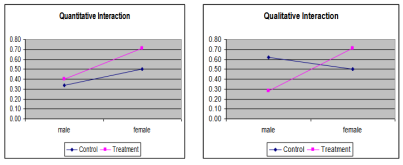


Figure 1. Illustrations of quantitative (left graph) and qualitative (right graph) interactions.

Statistical hypothesis tests of treatment by sex interaction have been widely utilized to detect treatment effect heterogeneity across sex. Most of the tests of interaction in common use have as their null hypotheses the absence of treatment by sex interaction. As statistical tests, their significance levels should be pre-specified in the investigational plan. Note, however, that the power of such tests may be unspecified. Therefore, lack of statistical significance for a test of treatment by sex interaction may not convincingly evidence the absence of clinically relevant interaction. By the same token, moderate statistical significance may not convincingly evidence the presence of clinically relevant interaction. While statistically significant interactions will be investigated for their clinical meaningfulness, interactions without associated statistical significance may also be examined for clinical reasons specific to the design and endpoint.

21

21

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

PCORI's HTE guidelines

Varadhan R, Stuart EA, Louis TA, Segal JB, Weiss CO. **Review of Guidance Documents for Selected Methods in Patient Centered Outcomes Research: Standards in Addressing Heterogeneity of Treatment Effectiveness in Observational and Experimental Patient Centered Outcomes Research.**

Source: http://www.pcori.org/assets/Standards-in-Addressing-Heterogeneity-of-Treatment-Effectiveness-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf: PCORI; 2012. Accessed 01/23/2015.

22

22

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

5: Standards for Heterogeneity of Treatment Effects

HT-1 State the goals of HTE analyses

State the inferential goal of each HTE analysis, specifying how it is related to the topic of the research, translate this into an analytic approach, and highlight the linkages between the two. Identify analyses as hypothesis driven (sometimes denoted confirmatory), or hypothesis generating (sometime denoted exploratory).

HT-2 For all HTE analyses, pre-specify the analysis plan; for hypothesis-driven HTE analyses, pre-specify hypotheses and supporting evidence base

The study protocol should unambiguously pre-specify planned HTE analyses. Pre-specification of hypothesis-driven HTE analyses should include a clear statement of the hypotheses the study will evaluate, including how groups will be defined (e.g., by multivariate score or stratification) and outcome measures, and the direction of the expected treatment effects. The pre-specified hypotheses should be based on prior evidence, which should be described clearly in the study protocol and published paper.

Source: http://www.pcori.org/assets/Standards-in-Addressing-Heterogeneity-of-Treatment-Effectiveness-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf: PCORI; 2012. Accessed 01/23/2015.

23

23

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

HT-3 All HTE claims must be based on appropriate statistical contrasts among groups being compared, such as interaction tests or estimates of differences in treatment effect

A common error in HTE analyses is to claim differences in treatment effect when one group shows a statistically significant treatment effect and another does not. To claim differences in treatment effect among subgroups, appropriate statistical methods must be used to directly contrast them. Such contrasts include, but are not limited to, interaction tests, differences in treatment effect estimates with standard errors, or a variety of approaches to adjusting the estimated subgroup effect, such as Bayesian shrinkage estimates. Within each subgroup level, studies should present the treatment effect estimates and measures of variability.

HT-4 For any HTE analysis, report all pre-specified analyses and, at minimum, the number of post hoc analyses, including all subgroups and outcomes analyzed

Protocols and study reports must report the exact procedures used to explore HTE, including data mining or any automatic regression approaches. HTE analyses should clearly report the procedures by which subgroups were defined (e.g., by categorical predictors or continuous risk scores) and the effective number of subgroups and outcomes examined. If a non-prespecified stratum or subgroup is claimed to show a treatment effect that is different from others, methods should be used that account for the number of contrasts examined. These methods include, but are not limited to, p-value adjustment, false discovery rates, Bayesian shrinkage estimates, adjusted confidence intervals, and validation methods (internal or external).

Source: http://www.pcori.org/assets/Standards-in-Addressing-Heterogeneity-of-Treatment-Effectiveness-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf: PCORI; 2012. Accessed 01/23/2015.

24

24

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

HTE

- Definition HTE:
  "the non-random, explainable variability in the direction and magnitude of treatment effect"
- Beneficial and adverse responses of interest
- Assessment of HTE is **essential** in patient-centered outcomes research
- Stakeholders such as patients, clinicians, and policy makers have to understand HTE to make informed decisions

25

25

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

HTE (cont.)

Two main goals:

1. Estimate treatment effect for clinically relevant subgroups (→ subgroup analysis)
2. Predict whether a specific individual might benefit from treatment (→ predictive learning; not covered here)

26

26

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

HTE (cont.)

- Subgroup analysis (SGA) is the most common analytic approach for examining HTE
- They propose **minimum** standards for SGA:
  - Prior planning
  - Careful analysis
  - Responsible reporting

27

27

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

HTE (cont.)

Two types of SGA:

1. estimating treatment effect separately within levels of baseline covariates
2. modeling the interaction between the treatment and covariates

28

28

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

How are HTEs assessed?

Common approach - example:

Outcome Y is binary and logistic regression is estimated. Treatment effect is log-odds ratio of

Y = 1 when treatment indicator 1 vs. 0 ("averaged" over other covariates)

29

29

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

How are HTEs assessed? (Cont.)

Separate models are estimated within strata of covariates, say, X = 1 and X = 2, resulting in θ_1 and θ_2.

HTE is inferred when association in group 1 is statistically significant and in group 2 is not, or vice-versa.

Observation: This is incorrect for inferring HTE (but it is correct for estimating stratified treatment effects)

30

30

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

How are HTEs assessed? (Cont.)

At a minimum: Test for HTE needs to test whether the difference between the two stratified treatment effects is zero using a Wald test. (Default in SAS PROC FREQ RISKDIFF option is METHOD=WALD)

Better: Model the interaction between treatment and X

$$g(E[Y \mid X, A]) = b_0 + b_A A + b_X X + b_{A.X} A * X$$

31

31

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

How are HTEs assessed? (Cont.)

$$g(E[Y \mid X, A]) = b_0 + b_A A + b_X X + b_{A.X} A * X$$

Stratified effects : $b_A = \theta_1$ and $b_A + b_{A.X} = \theta_2$.

Interaction term : $b_{A.X} = \theta_2 - \theta_1$

Test for significance of interaction term!
Easy to implement even when X has more than two levels or is continuous.

32

32

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Should there be adjustments for multiplicity in testing?

PCORI complains about "the myopic view of SGA as a hypothesis testing problem rather than as an estimation problem" and the "dichotomization of SGA into confirmatory (hypothesis-testing) and exploratory (hypothesis-generating) analyses"

Suggestion: **descriptive** HTE analysis

33

33

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Which type of subgrouping variables?

- Demographics (e.g., **age and sex – always?**)
- Behavior (e.g., smoking)
- pathophysiology (e.g., measures of disease severity)
- genetic markers
- comorbid conditions (e.g., diabetes status in cardiovascular disease trials)
- "Studies might use descriptive HTE for sub-populations for which limited evidence is available in the literature, such as the priority populations specified by the Agency for Healthcare Research Quality, including women, children, minorities, elderly, individuals with disabilities, and rural populations" (p. 3)

Source: http://www.pcori.org/assets/Standards-in-Addressing-Heterogeneity-of-Treatment-Effectiveness-in-Observational-and-Experimental-Patient-Centered-Outcomes-Research.pdf: PCORI; 2012. Accessed 01/23/2015.

34

34

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Should there be adjustments for multiplicity in testing? (cont.)

"In descriptive HTE analysis, the focus is on estimation of treatment effects in pre-specified subgroups. Each study reports these effect estimates and their **standard errors** in order to **facilitate future meta-analysis**."

→ No, **descriptive** HTE analysis does not need to be adjusted for multiplicity in testing, but confidence intervals have to be provided.

35

35

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Should an HTE-analysis be reported even when the overall treatment effect is not statistically significant?

Yes!

Argument: There is no theoretical argument that suggests that, e.g., males and females cannot have effects in opposite direction effectively cancelling when "averaging" (especially when comparison is not placebo).

36

36

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

What should be reported and why?

- Point estimates and standard errors to facilitate future meta-analyses
- All HRT tests even if not statistically significant because statistical significance is not necessarily expected when study is not powered adequately (but descriptive HTE still adequate)

37

37

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Table 1. The essential characteristics of the three different types of HTE analyses.

Properties	*Confirmatory HTE Analysis*	*Descriptive HTE Analysis*	*Exploratory HTE Analysis*
Inferential Goal	To test hypotheses related to subgroup effects	To report treatment effects for future synthesis	To generate hypotheses for further study
Number of subgroups analyzed	A small number, typically, one or two	Moderate to large	Not made explicit, but may be large
Scientific rationale and prior evidence for hypotheses	Strong	Immaterial	Weak or none
Pre-specification of data analytic strategy	Fully pre-specified	Fully pre-specified	Not pre-specified

38

38

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Table 1. The essential characteristics of the three different types of HTE analyses.

Properties	*Confirmatory HTE Analysis*	*Descriptive HTE Analysis*	*Exploratory HTE Analysis*
Control of family-wise type I error probability	Should be done	Not needed	Difficult, since it is not obvious how many related tests were performed
Characterization of sampling properties of the statistical estimator (e.g., standard errors, type-I error rate)	Easy to achieve	Possible	Difficult
Power for testing hypothesis	Ideally, study designed to have sufficient power	Likely to be inadequately powered, but this is immaterial	Typically, inadequate power to examine several hypotheses

39

39

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Proposed Minimum Standards for HTE Analysis in PCOR

State the Goals of HTE Analyses
For Confirmatory HTE Analyses, Prespecify a Few Subgroup Hypotheses
For Confirmatory HTE Analyses, Involve Stakeholders in the Selection of Subgroups and Outcomes
For Confirmatory HTE Analyses, Report *a priori* Statistical Power
For Confirmatory and Descriptive HTE Analyses, Report Sufficient Information on Treatment Effect Estimates
For Descriptive HTE Analyses, Prespecify Subgroups and Outcomes
For Descriptive HTE Analyses, Involve Stakeholders in the Selection of Subgroups and Outcomes
For Exploratory HTE Analysis, Document the Number of Subgroups and Outcomes Analyzed
For Exploratory HTE Analyses, Discuss Findings in the Context of Study Design and Prior Evidence
For Any HTE Analysis in Observational Data, Explicitly Assess Study Quality for Making Causal Inference
For Any HTE Analysis, Describe the Analytical Methods in Detail
For Any HTE Analysis, Perform an Interaction Test
For Any HTE Analysis, Report All Prespecified Analyses
For Any HTE Analysis, Use Appropriate Methods for Post-treatment Subgroups

(Box 1, p. 8)

40

40

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Data Management and Sharing Plan (DMS)

https://sharing.nih.gov/

41

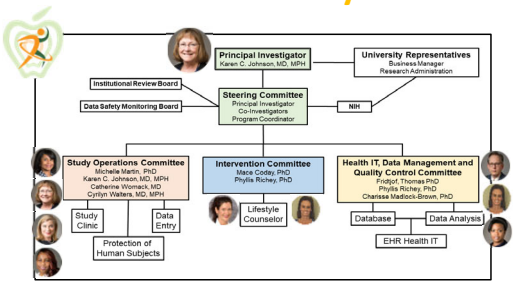
41

Department Name Here

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Example from CHAMPS study

Structure of Study Team



42

42

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Example from CHAMPS study

MOP

Chapter 13 CHAMPS Data Analysis, Sharing, and Monitoring Plan

CHAMPS Data Analysis, Sharing and Monitoring Plan

1. **Power and Sample Size Estimate**
2. **Data Management Plans**
   - 2.1. Data Confidentiality
3. **Analysis Plans**
   - 3.1. Primary Analysis
   - 3.2. Missing and Incomplete Data
   - 3.3. Censoring
4. **Data Sharing (NOT-OD-21-013)**
   - 4.1. Data Type
   - 4.2. Related Tools, Software and/or Code
   - 4.3. Standards
   - 4.4. Data Preservation, Access, and Associated Timelines
   - 4.5. Access, Distribution, or Reuse Considerations
   - 4.6. Oversight of Data Management and Sharing

2-page DMP merged into the MOP

CHAMPS
CHOOSING HEALTHY ACTIVITIES AND LIFESTYLE MANAGEMENT THROUGH PORTAL SUPPORT

43

43

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4. DATA SHARING (NOT-OD-21-013)

The CHAMPS study will prepare and distribute an electronic dataset with scientific data, incl. electronic versions of any paper forms that might have been used in data collection. Confidentiality of individual participants will be maintained with all releases of data. The final CHAMPS study analytical database will be processed according to HIPAA definitions for public data sharing. During this process, the participant data will be de-identified by using processes which include but are not limited to: removal of identifiers, translation of dates to delta time values, and assignment of random study identifiers. Out of this process will be a series of de-identified data files representing the final analytical data set. These data files will be provided in a standard format that is readable across a variety of applications and operating system platforms. Documentation that will be provided along with the data sharing file that may include but is not limited to: data dictionary, data code book, valid variable ranges, the protocol, MOP, and intervention manual or programs. The data release documentation will provide detailed, organized documentation of study variables and clear instructions on how to install and access the data. CHAMPS intents to make all data available as outlined in the Final NIH Policy for Data Management (NOT-OD-21-013). However, CHAMPS has been funded before Jan 25, 2023, and is therefore technically not covered by the new NIH data management and sharing policy.

44

44

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4.1 Data Type

The scientific data to be generated is from 250 randomized participants and originates from questionnaires (socio-demographics, medical history, medications, diet/DHQ III, physical activities/GPAQ and sedentary behavior, alcohol and smoking, neighborhood environment, quality of life/PROMIS, social determinants of health, technology use), clinic measures (vital signs, weight, height, waist and hip circumference), and others (adverse events, intervention participation/video viewing through webpage analytics).

Data that will be preserved and shared consists of questionnaire and clinic visit data on the individual level, as well as processed data summarizing intervention exposure/web-viewing on the individual level.

Metadata is publicly available at the U.S. National Library of Medicine in the ClinicalTrials.gov database (Identifier: NCT05410353; https://www.clinicaltrials.gov/ct2/show/NCT05410353) and will be updated throughout the conduct of the study. In addition, structured metadata will be created as required by the respective repositories that will be used.

DHQ III = Diet History Questionnaire III
GPAQ = Global Physical Activity Questionnaire
PROMIS = Patient-Reported Outcomes Measurement Information System (NIH)

45

45

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4.2 Related Tools, Software and/or Code

All data will be shared as comma-separated values (CSV) files with UTF-8 encoding that do not require specific software for reading/encoding. Data sets will be of rectangular arrangement with rows corresponding to participants.

46

46

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4.3 Standards

Mostly standardized questionnaires are used such as the DHQ III, GPAQ, or PROMIS. A detailed account is available in the Manual of Procedures (MOP), Chapter 11: Data Collection. Unique participant identifiers will link entries in the various data sets. Data dictionaries will be provided for the available data sets.

DHQ III = Diet History Questionnaire III
GPAQ = Global Physical Activity Questionnaire
PROMIS = Patient-Reported Outcomes Measurement Information System (NIH)

47

47

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

List of all nutrients and food groups available from DHQ III

Expand All | Collapse All

- + Carbohydrate Constituents
- + Carotenoids and Tocopherols
- + Dietary Constituents from Supplements (reported separately in output files)
- + Fats, Fatty Acids, & Cholesterol
- + Food Pyramid Equivalents
- + Macronutrients & Energy
- + Minerals
- + Other
- + Protein Constituents
- + Vitamins

- − Minerals
  - Calcium
  - Phosphorus
  - Magnesium
  - Iron
  - Zinc
  - Copper
  - Selenium
  - Sodium
  - Potassium
  - Manganese

- − Fats, Fatty Acids, & Cholesterol
  - Cholesterol
  - Total saturated fatty acids
  - Total monounsaturated fatty acids
  - Total polyunsaturated fatty acids
  - Solid Fat
  - SFA 4:0 (Butanoic)
  - SFA 6:0 (Hexanoic)
  - SFA 8:0 (Octanoic)
  - SFA 10:0 (Decanoic)
  - SFA 12:0 (Dodecanoic)
  - SFA 14:0 (Tetradecanoic)
  - SFA 16:0 (Hexadecanoic)
  - SFA 18:0 (Octadecanoic)
  - MFA 16:1 (Hexadecenoic)
  - MFA 18:1 (Octadecenoic)
  - MFA 20:1 (Eicosenoic)
  - MFA 22:1 (Docosenoic)
  - PFA 18:2 (Octadecadienoic)
  - PFA 18:3 (Octadecatrienoic)
  - PFA 18:3N3 (Alphalinolenic)
  - PFA 18:4 (Octadecatetraenoic)
  - PFA 20:4 (Eicosatetraenoic)

48

48

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4.4 Data Preservation, Access, and Associated Timelines

Metadata available at ClinicalTrials.gov (Identifier: NCT05410353) will be archived by the U.S. National Library of Medicine. The collected human subject data is identifiable protected health information (PHI) under The Health Insurance Portability and Accountability Act of 1996 (HIPAA).

To the extent possible, data that we will make available will be de-identified of PHI using the Safe Harbor de-identification method as defined in §164.514(a) of the HIPAA Privacy Rule. We foresee to be able to furnish such limited data sets at the point in time of publications and these will contain the information needed to reproduce the main analyses in the respective publication. Because limited data sets remain protected health information under HIPAA, access to the data will need to be controlled by data use agreements (DUAs) and disclosure will only be for the purpose of research, public health or health care operations. These data sets are intended to be posted in a repository "close" to the publication (e.g., in PubMed) or in a repository explicitly named in the publication.

A comprehensive scientific data set will be posted in a repository by the end of the project performance period. That comprehensive scientific data will include electronic versions of any paper forms that might have been used as well as scientific data not (yet) used in the study publications: e.g., the DHQ-III data output with comprehensive information about carbohydrates, vitamins, minerals, protein constituents, etc. will be released as available in this study. To protect the privacy or our participants, we will remove "facial" identifiers such as names, medical record numbers, etc., but will possibly not create a data set formally classified as limited data under HIPAA, and access will have to be controlled.

We anticipate that the produced data is of value for the larger research community, institutions, and/or the broader public, and therefore prefer repository-solutions with anticipated "indefinite" storage such as ClinicalTrial.gov, PubMed, or other NIH-recommended repositories. The scientific data will be findable and identifiable by the references in the respective publications as well as through the metadata provided to the repositories (incl. ClinicalTrials.gov; Identifier: NCT05410353; https://www.clinicaltrials.gov/ct2/show/NCT05410353)

Supporting material that will also be made available includes the Study Protocol/Manual of Operations (MOP), Statistical Analysis Plan (SAP), Informed Consent Form (ICF), and the analytic code for main analyses.

49

49

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4.5 Access, Distribution, or Reuse Considerations

The IRB approved informed consent allows for usage of the collected scientific data for other medical research (secondary research). Individual details obtained from participants can be provided in publications or presentations, but they cannot be discussed in a way that would allow to identify the participant.

All our released data will remain protected health information under HIPAA, and access to the data will need to be controlled by a data use agreement (DUA; for limited data sets), an IRB approval (for data not de-identified by the Safe Harbor method as defined in §164.514(a) of the HIPAA Privacy Rule), and disclosure will only be for the purpose of research, public health or health care operations.

FAIR = Findable, Accessible, Interoperable, and Reusable
https://fair-research.org/

Is your research reproducible?
Is your data processing and are your statistical analyses reproducible?

50

50

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

4.6 Oversight of Data Management and Sharing

Compliance with this DMS plan is monitored by the PI of the study, Dr. Karen Johnson, during the project performance period. The HIPAA Privacy Rule requires appropriate safeguards to protect the privacy of protected health information and sets limits and conditions on the uses and disclosures that may be made of such information without an individual's authorization. This liability is transferred to the respective repository when data sets are surrendered to these repositories.

51

51

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

Health Sciences Library Research Support

https://www.uthsc.edu/library/research.php
https://libguides.uthsc.edu/data/uthsc-data

Jess Newman, MSIS
Research Data and Scholarly Communications Lead at the Health Sciences Library

General Support Services, email library@uthsc.edu
Literature searches, systematic reviews, citation managers

Data Support Services, email jnewman@uthsc.edu
Assessing & improving research impact (h-index, citation metrics, etc)
Writing Data Management Plans (DMPs)
Selecting appropriate data repositories for storage and sharing
Basic data wrangling in Excel and OpenRefine
Enhancing data visualizations
Locating data for re-use
Identifying & evaluating journals for publication
Open Access publishing
Copyright and licensing

Slide content provided by the UTHSC Library/Jess Newman

52

52

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER

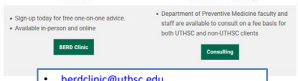
Michelle Martin, PhD
Co-Director TN-CTSI

UTHSC
TENNESSEE CLINICAL AND TRANSLATIONAL SCIENCE INSTITUTE

https://tnctsi.uthsc.edu/

Karen Johnson, MD, MPH
Co-Director TN-CTSI

Biostatistics, Epidemiology, and Research Design (BERD) Unit
https://tnctsi.uthsc.edu/consultation-and-services/biostatistical-support/



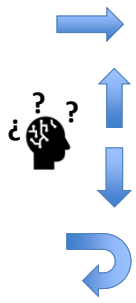
- berdclinic@uthsc.edu
- BERD Consulting Manager – Tristan Hayes thayes@uthsc.edu

53

53

Preventive Medicine

THE UNIVERSITY of TENNESSEE HEALTH SCIENCE CENTER



Thank you!

Questions?

54

54