Strategic considerations and value of covariate adjustment

Courtney Schiffman, Principal Statistician
Genentech, a member of the Roche group

Defining the Scope

Covariate Adjustment: Using prognostic baseline variables to improve precision and power when estimating marginal Average Treatment Effects (ATE)

We focus on the **standardized estimator, where treatment arm means are estimated as the average of predicted outcomes from a working regression model** (glm) that adjusts for baseline covariates

- Applicable to continuous, binary and count outcomes
- Clearly defined estimand: **marginal Average Treatment Effect (ATE)**, a contrast of the treatment arm means, e.g. marginal ATE = Active Arm Mean - Control Arm Mean
- Same estimand as an unadjusted analysis, not a conditional analysis to enable personalization!
- **Asymptotically valid even if the working regression model is misspecified**

Model-assisted, not model-dependent!

Historical Context of Covariate Adjustment

William G Cochran, "*Sampling Techniques*," John Wiley and Sons, **1977**.

Li Yang & Anastasios A Tsiatis, "*Efficiency Study of Estimators for a Treatment Effect in a Pretest–Posttest Trial*," The American Statistician, **2001**.

David A. Freedman, "*On regression adjustments to experimental data*", Advances in Applied Mathematics, **2008**.

Moore KL, van der Laan MJ, "*Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation,*" Stat Med. **2009**.

Rosenblum, Michael, and Mark J van der Laan. "*Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables*." The international journal of biostatistics, **2010**.

Winston Lin, "*Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique*." Ann. Appl. Stat., **2013**.

Trower, Antonia, Felix Balazard, and Radha Patel. "*Leveraging Machine Learning to Optimize Clinical Trials: The Power of Covariate Adjustment*." Owkin https://www.owkin.com/newsfeed/leveraging-machine-learning-to-optimize-clinical-trials **2020**

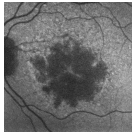
Bingkai Wang et al. "*Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment*", Journal of the American Statistical Association, 118:542, 1152-1163, **2021**

Strategies for Determining Covariates

Maximize the Value of Data

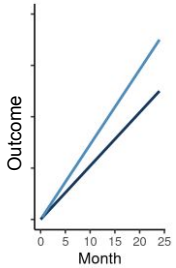
Leverage **independent data** (e.g. historical trial data, observational data) to find the most promising "super covariates" for covariate adjustment

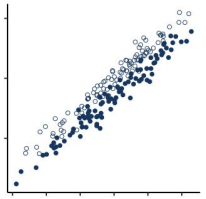
Develop and select a prognostic model *using independent datasets*



Pooled and harmonized data from multiple previous clinical trials and observational studies

Apply pre-specified model *to patients in a* ***new trial*** *and adjust for their predicted outcomes* ***as a "super covariate"***





"**Super Covariate**": A multivariate function of covariates

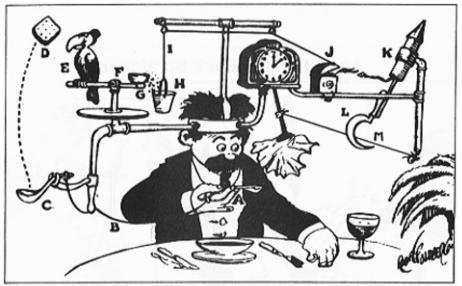
- Should be built using data independent of the trial
- Ideally a prediction of the trial outcome

Modeling Goal: Minimum complexity with maximum performance

Always develop a comparison benchmark model with minimum complexity

Factors characterizing complexity:

- **Number of features**: Cleaning and using many variables creates a lot of work for study teams
- **Type of features**: E.g. variables that are difficult, invasive or costly to collect (e.g. CSF biomarker, PET scan)
- **Type of models**: Black-box models, E2E deep learning, additive models
    - Ease of model deployment
    - Interpretable models are generally preferred and may be more robust and generalizable



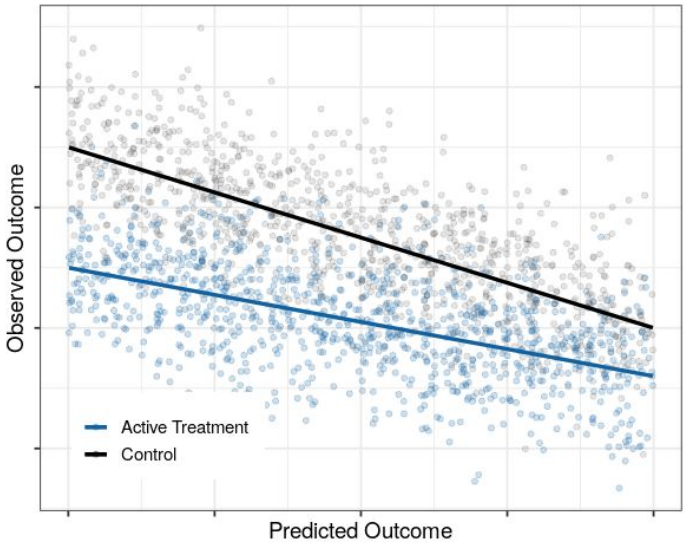
Don’t build a Rube Goldberg model

> 5. “Simplicity is a great virtue but it requires hard work to achieve it and education to appreciate it. And to make matters worse: complexity sells better.”
>
> — Edsger W. Dijkstra

Use Intuitive Metrics to Quantify Expected Benefits of Adjustment

Effective Sample Size Increase (ESSI) for Continuous Outcomes

Example: ESSI = 40% means precision and power gain from covariate adjustment is equivalent to **running a 40% larger trial!**



The ESSI only depends on **two parameters**:

$$ESSI = \left(\frac{1}{1 - \left(\frac{r_{control} + r_{active}}{2} \right)^2} - 1 \right) * 100\%$$

$\mathbf{r_{control}}$ = correlation between covariate and outcome in control arm

$\mathbf{r_{active}}$ = correlation between covariate and outcome in active trt arm

*ESSI shown here assumes 1:1 randomization and equal variances. More general formulas available!
https://www.stats4datascience.com/posts/covariate_adjustment/#essi-formulas

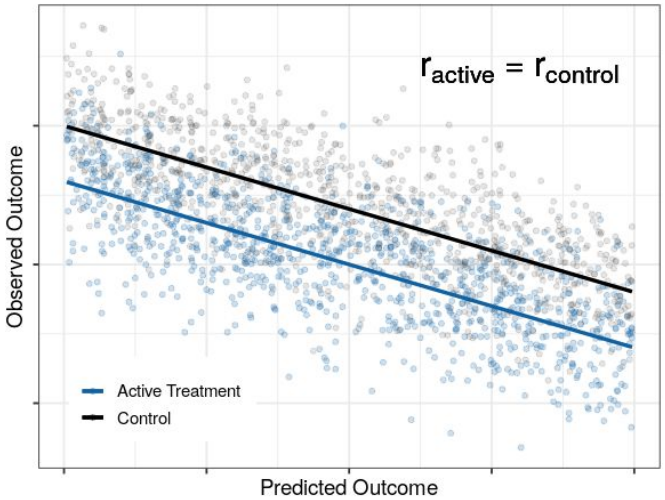
How to estimate ESSI

- $r_{control}$ can be estimated from prior, independent data
- **r_{active} depends on the "treatment effect scenario"**

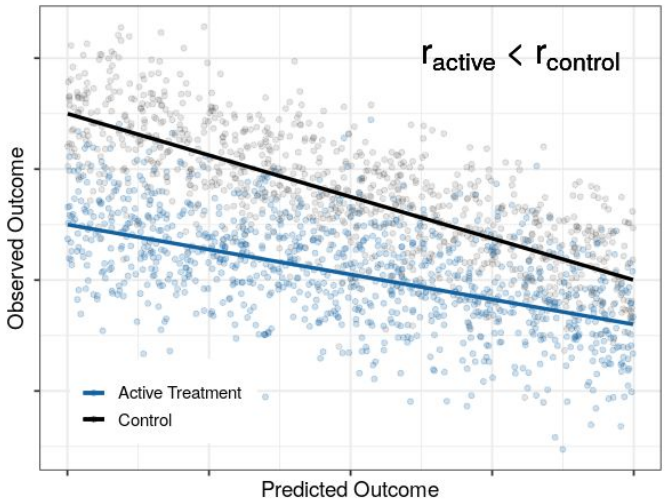
$$ESSI = \left(\frac{1}{1 - \left(\frac{r_{control} + r_{active}}{2} \right)^2} - 1 \right) * 100\%$$

Consider two plausible "treatment effect scenarios":

1. Constant Absolute Treatment Effect



2. Constant Proportional Treatment Effect



Treatment effect heterogeneity reduces benefits of adjustment!!

Use caution when considering reducing sample size

Phase 2 Recommendations

improve decision making

- **ESSI depends on treatment effect heterogeneity**. In a worst case, ESSI = 0%
- Due to the risk of underpowering, if no prior data is available on treatment effect heterogeneity, we **DO NOT GENERALLY RECOMMEND reducing sample size in Phase 2**
- If sample size is reduced, be sure to communicate the risks to the team

Phase 3 Recommendations

Improve trial efficiency

- Supplement ESSI estimates from independent datasets with estimates from Phase 2
- Could consider reducing sample size, if other trial requirements allow for it (e.g. safety and subgroup analysis). However be aware of risks and tailor sample size reductions accordingly: (1) noise in ESSI estimates and (2) study-to-study effects

Strategies for Determining the Form of the Working Regression Model

Working Regression Model Budget (for Continuous Outcomes)

Adjusting for too many covariates in your working model relative to your sample size may invalidate the analysis

Model Budget (i.e number of allowable model terms excluding the overall intercept) should probably not exceed **5-7% of the total sample size** (*borrowed from prognostic modeling practices* - more work needed*)

Spend your model budget wisely

- Prioritize covariates with the **most prognostic value** (use ESSI formulas!)
- Utilize "super covariates" when model budget is tight
    - If it is a simple linear model, and the model budget allows, you can include the individual covariates in your working regression model!
- Make **additive models** (as opposed to models with treatment by covariate interactions) the default choice
- **Avoid dichotomania** forced by stratification to increase precision

*Harrell, Frank E. 2011. *Regression Modeling Strategies*. New York: Springer.

Use covariate adjustment to account for stratified randomization

Roche

Covariate adjustment assuming simple random sampling is **consistent** and **inference is at worst conservative**.

To remove conservatism, could add indicators for all combinations of strata factors + interactions with treatment*. **This is an inefficient way to capture prognostic value and leads to too many model terms!**

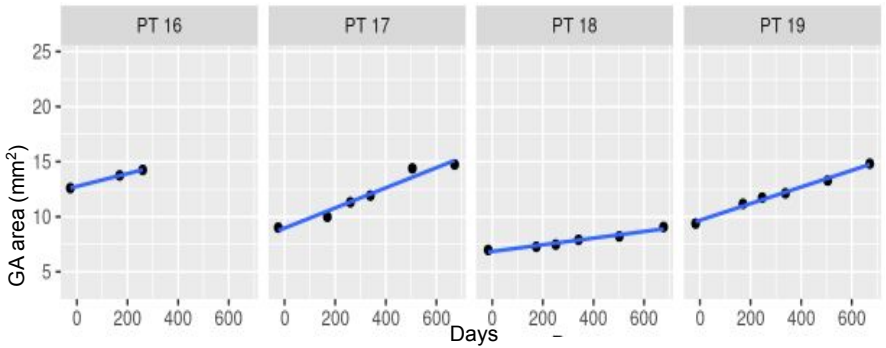
Instead, use covariate adjustment and remember to spend your model budget wisely!

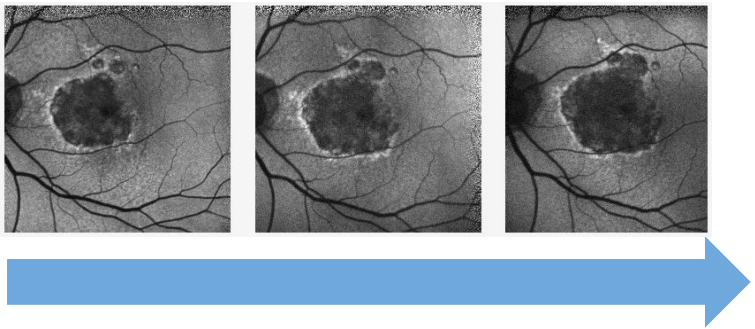
This may be a controversial recommendation, but it is important to **avoid adding unnecessary complexity** to analysis and **exceeding the working regression model budget**!

*Bingkai Wang et al. "Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment", Journal of the American Statistical Association, 118:542, 1152-1163, 2021

An example from ophthalmology

Geographic Atrophy (GA) Progression Modeling Objective



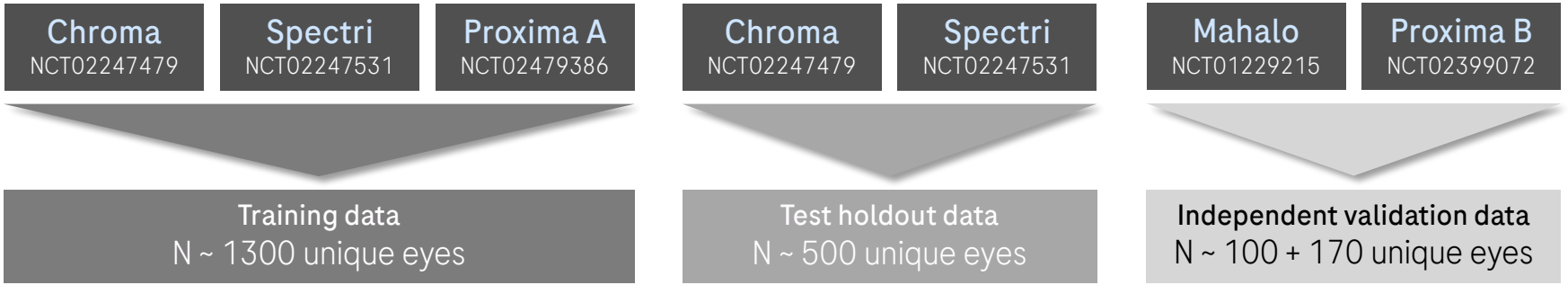


GA Progression is assessed by lesion growth rate over time in **Fundus Autofluorescence (FAF)** images as the primary endpoint in a clinical trial

- **Endpoint in GA trials:** GA progression as defined by a **growth rate in GA area**
- **Objective:** Develop model(s) that **predict the primary endpoint** using baseline information. Use the model(s) to improve trial power via covariate adjustment

Rigorous data strategy for model development

Roche datasets from previous clinical trials



Two models for creating a “super covariate”

Roche

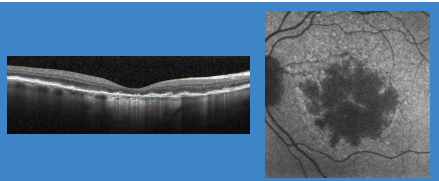
Benchmark Feature-Based Model

Known prognostic factors, lesion characteristics from reading center and functional outcomes

Benchmark model:
Most predictive **baseline** features from a pre-specified list of standard variables:

A **simple linear regression with 5 features** (more complex models didn’t improve performance)

E2E Deep Learning Model*



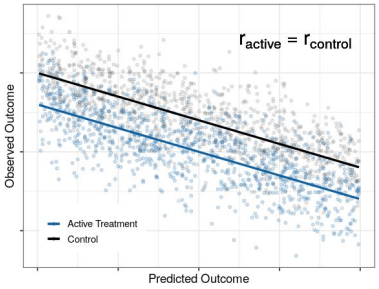
FAF Model:
A deep learning model predicting GA progression directly from the raw FAF baseline image

*Anegondi, Neha et al. 2022. “Deep Learning to Predict Geographic Atrophy Area and Growth Rate from Multi-Modal Imaging.” *Ophthalmology Retina*. https://doi.org/10.1016/j.oret.2022.08.018.

Model Performance

Metric	Model	Holdout	MAHALO	Proxima B
$r^2_{control}$	Benchmark	0.16	0.33	0.25
	DL FAF	0.48	0.63	0.48
ESSI* vs Unadjusted	Benchmark	19%	50%	34%
	DL FAF	92%	174%	91%
ESSI* vs Bench	DL FAF	61%	82%	43%

*Assuming a constant proportional treatment effect



$$ESSI = \left(\frac{1}{1 - \left(\frac{r_{control} + r_{active}}{2} \right)^2} - 1 \right) * 100\%$$

Impact: DL model significantly increases the effective sample size

- **Doubles** size of trials using unadjusted analyses
- **Increases by 50%** trials using benchmark analyses
- In this example, expected ESSI is worth the model complexity!

Thoughts about using Super Covariates in a regulated setting

- Covariate adjustment can be classified as “low risk” application
- Models should be built on data external to the trial, should be clearly pre-specified (in SAP) and locked prior to unblinding
- Appropriate documentation on data pre-processing and model deployment

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

13 July 2023
EMA/CHMP/CVMP/83833/2023
Committee for Medicinal Products for Human Use (CHMP)
Committee for Medicinal Products for Veterinary Use (CVMP)

Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle
Draft

https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf

2.2.3.3. Data analysis and inference

When AI/ML models are used for transformation or analysis of data within a clinical trial of a medicinal product, they are considered a part of the statistical analysis and should follow applicable guidelines on statistical principles for clinical trials (see Section 5) and include analysis of the impact on downstream statistical inference. In late-stage clinical development, this requires a detailed description of a pre-specified data curation pipeline and a fully frozen set of models used for inference, within the statistical analysis plan.

Prior to the opening of any dataset used for hypothesis testing, the data pre-processing pipeline and all models should be locked and documented in a traceable manner in the statistical analysis plan. Once a dataset has been opened, any non-prespecified modifications to data processing or models implies that analysis results are considered *post hoc* and hence not suited for confirmatory evidence generation.

If possible, it is encouraged that models are published in an open repository prior to their deployment in a pivotal clinical trial.

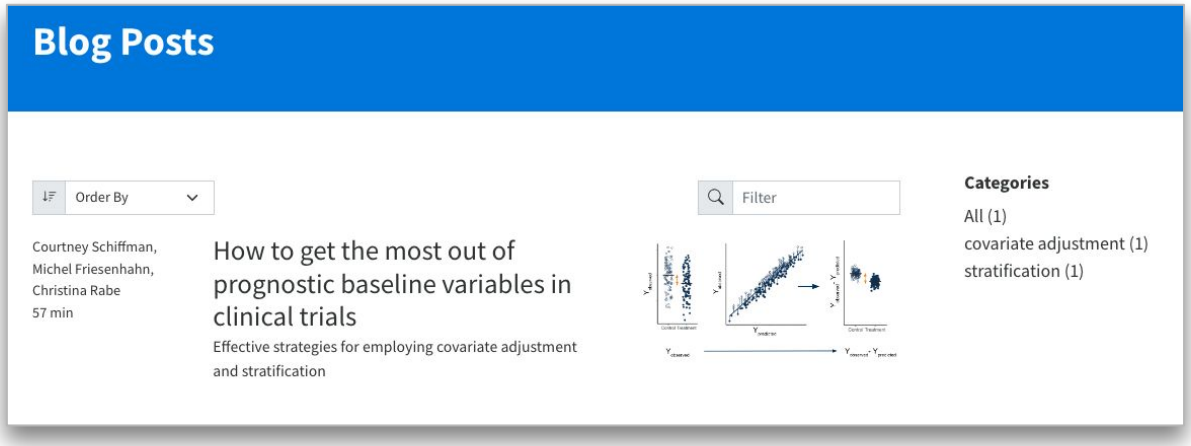
Some concluding remarks

- Covariate Adjustment is not synonymous with “Digital Twins”
  - Covariate adjustment is all about predicting point estimates of outcomes, not distributions of outcomes
  - Digital Twins involves generating distributions of outcomes
- If mixed models are used for longitudinal data, be careful with model specifications to ensure optimal power gains from covariate adjustment
  - E.g. for mmrm make sure to include appropriate covariate-by-time interactions*
- In light of recent methodological developments, should the 2015 EMA guidance document for covariate adjustment be updated?

*Schuler A. Mixed Models for Repeated Measures Should Include Time-by-Covariate Interactions to Assure Power Gains and Robustness Against Dropout Bias Relative to Complete-Case ANCOVA. Ther Innov Regul Sci. 2022 Jan;56(1):145-154. doi: 10.1007/s43441-021-00348-y.

Check out our blog post on covariate adjustment

stats4datascience.com



Special thanks to my colleagues **Christina Rabe** and **Mike Friesenhahn**!!