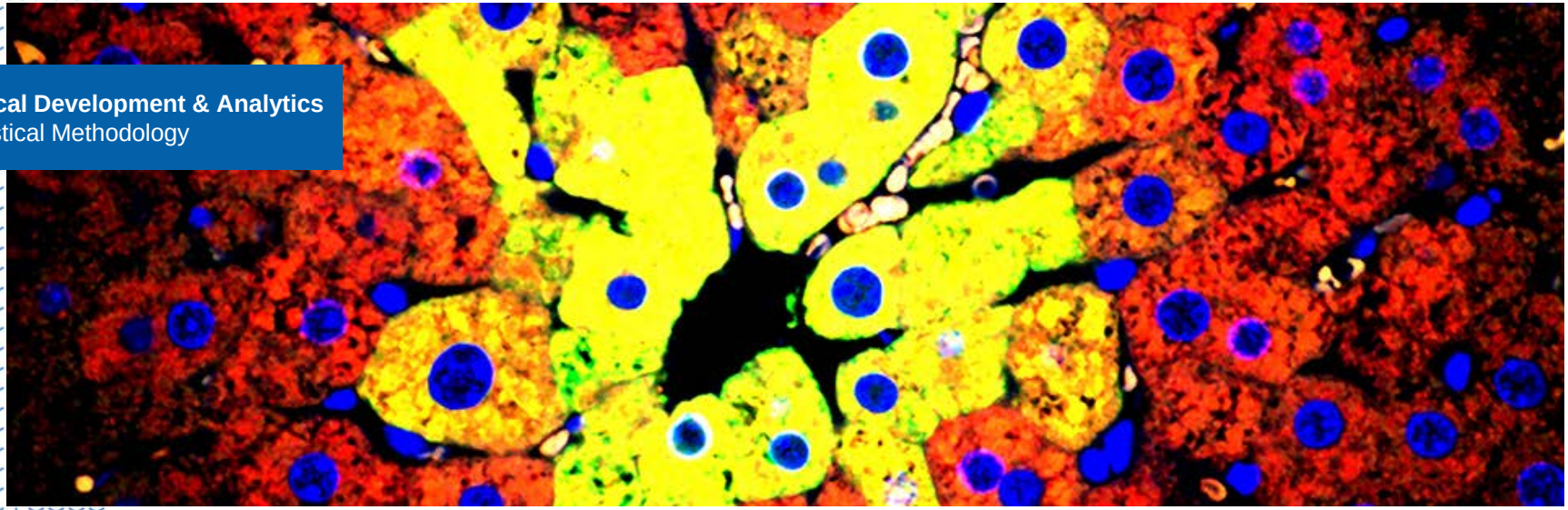
Clinical Development & Analytics
Statistical Methodology



Drug development, the ICH E9 addendum and causal inference

Björn Bornkamp

BBS Training: A gentle introduction to causal thinking

February 02, 2021

NOVARTIS | Reimagining Medicine

Disclaimer

The views and opinions expressed in this presentation and on the slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.

Acknowledgements

- Material is based on a trainings on causal inference and estimands given and prepared together with Heinz Schmidli, Dong Xi
- Dominik Heinzmann, Giusi Moffa, Marcel Wolbers, Amanda Ross and Marc Vandemeulebroecke for comments on earlier versions of this slide deck

Outline

- Causal inference & drug development (randomized clinical trials)
- Causal interpretation of ITT and per-protocol analysis
- Causal estimands, effect measures & parameter estimates in regression models
  - Standardization as an alternative to LS-means

Causal inference & drug development

NOVARTIS | Reimagining Medicine

Drug development

- Causal questions are central to clinical drug development
- Randomization facilitates causal inference
- Complex questions regarding causality arise not only in observational data analyses but also in RCTs

Causal estimands

US National Academy of Science (2010)

“The trial protocol should explicitly define

- the objective(s) of the trial;
- the associated primary outcome or outcomes;
- how, when, and on whom the outcome or outcomes will be measured;
- The measures of intervention effects, that is, the ***causal estimands*** of primary interest.

These measures should be meaningful for all study participants, and estimable with minimal assumptions.”

Causal estimands

ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials (2019)

Aligned with causal reasoning, although term “causal” not used.

“Central questions for drug development and licensing are to establish the existence, and to estimate the magnitude, of treatment effects: ***How the outcome of treatment compares to what would have happened to the same subjects under alternative treatment*** *(i.e. had they not received the treatment, or had they received a different treatment).”*

ICH E9(R1) addendum vs causal inference

- ICH E9(R1) defines intercurrent events (IE)

"... *Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. It is necessary to address intercurrent events when describing the clinical question of interest in order to precisely define the treatment effect that is to be estimated.* ..."

- **Events that prevent us from running the „idealized" randomized trial**
  - Main point in ICH E9(R1) these events need to be acknowledged on the „estimand" level (not implicitly as part of analysis)

ICH E9(R1) addendum vs causal inference

- Suggests five strategies
  - **Treatment policy**: Effect regardless of IE → IE becomes part of „treatment attribute“
  - **Hypothetical**: Effect in hypothetical scenario where IE would not occur
  - **Composite**: Effect on a composite variable, where IE is part of the variable
  - **While-on-treatment**: Effect up to IE is considered of interest (modifies variable, i.e. observation time per patient)
  - **Principal Stratum**: Effect in the subpopulation of patients where IE would not occur
- No causal notation and only limited causal language used in ICH E9(R1)
  - Lipkovich et al (2020)* translate strategies into causal language (potential outcome and causal DAGs)

* Lipkovich, I., Ratitch, B. & Mallinckrodt C.H. (2020) Causal Inference and Estimands in Clinical Trials, Statistics in Biopharmaceutical Research, 12:1, 54-67, DOI: 10.1080/19466315.2019.1697739

Causal interpretation of ITT and per-protocol analysis

ITT analysis: poll question

Double-blind randomized trial with continuous endpoint Y at week 12
→ Patients randomized to daily doses of the investigational treatment (A=1) or control (A=0).

At the end of the trial one calculates (assume no missing data)

$$\text{Mean}[Y_i \mid A_i = 1] - \text{Mean}[Y_i \mid A_i = 0],$$

i.e., the difference in means between patients randomized to A=1 and A=0, regardless of how frequently the patient takes the treatment.

Does this quantity estimate a causal effect?

- YES
- NO

ITT analysis

- Recall definition of causal effect
  - A randomized → patients with A=1 and A=0 constitute exchangeable populations (formally $Y^{a=0}, Y^{a=1} \perp A$)
  - Yes, this is estimating a causal effect
  - Estimand of the analysis is: $E[\,Y^{a=1}\,] - E[\,Y^{a=0}\,]$
    - Can be estimated by the difference in observed means

ITT analysis

But: The causal effect of what?

- Causal effect of being randomized to a treatment
  → Does this correspond to a *clinically relevant question*?
  Depends...
  - on whether post-baseline events & subsequent actions will also occur in the same way in a real-life setting, reasons for non-adherence ...
- ITT does not estimate the effect of treatment: „had everyone adhered“
  - Different question!
- Causal inference requires clear definition of what constitutes „treatment“ (SUTVA, consistency assumptions, as discussed in part 1)
  - If there are multiple versions of „treatment“ potential outcomes not well-defined

ITT analysis

- Final ICH E9 addendum: „Treatment“ is an additional estimand attribute.
  - For treatment policy (ITT) strategy, intercurrent events become part of „treatment“ attribute
  - Will (hopefully) lead to more transparency
- No longer
  - Treatment: 150mg twice daily
- Now
  - Treatment: Initiate 150mg twice daily + optional rescue medication + optional switch to another treatment if an adverse event requiring treatment discontinuation occurs.
- Clinical relevance of treatment policy strategy for dealing with intercurrent events (rescue medication, AE,...) needs to be assessed on a case-by-case basis

Analyses based on per-protocol set: poll question

* Per-protocol set: Data set generated by the subset of subjects who complied with the protocol. Compliance covers considerations as absence of protocol violations and exposure to treatment.

Let $PPS = 1$ and $PPS = 0$ be inclusion or exclusion in the per-protocol set*.
Assume we calculate

$$Mean[Y_i \mid A_i = 1, PPS_i = 1] - Mean[Y_i \mid A_i = 0, PPS_i = 1]$$

i.e., the difference in means between patients that adhered to the protocol.

Does this quantity estimate a causal effect?

- YES
- NO

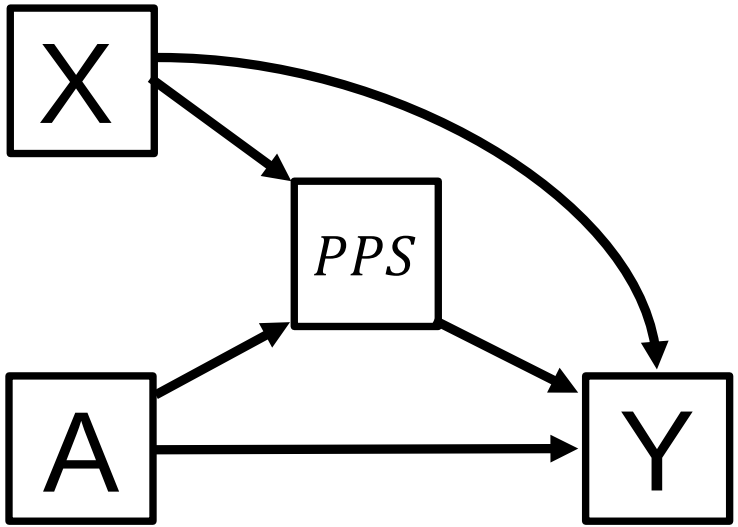
NOVARTIS | Reimagining Medicine

Analyses based on per-protocol set

- Let $PPS^{a=1}$ and $PPS^{a=0}$ be the potential outcomes of protocol adherence
- In potential outcomes we would compare
  - $Mean[Y_i^{a=1} | A_i = 1, PPS_i^{a=1} = 1] - Mean[Y_i^{a=0} | A_i = 0, PPS_i^{a=0} = 1]$
- Population with $PPS^{a=1} = 1$ (protocol adherers under trt) and $PPS^{a=0} = 1$ (protocol adherers under placebo) can be different
  - **Not a causal effect!**
  - Per-protocol analyses are discouraged in the final ICH E9 addendum.

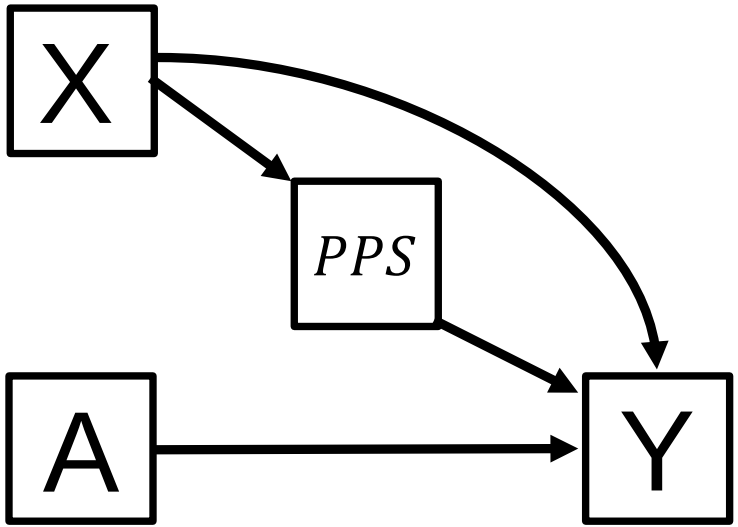
Analyses based on per-protocol set

A – treatment
Y – outcome
PPS – protocol adherence
X – patient covariates



Analyses based on per-protocol set

A – treatment
Y – outcome
PPS – protocol adherence
X – patient covariates



If this DAG would be true, the per-protocol analysis would target a causal effect as PPS is unaffected by A, so that

$$PPS^{a=1} = PPS^{a=0}$$

ICH E9(R1) on per-protocol set analyses

- ICH E9 (1998) suggests analyses based on Full Analysis Set (FAS) and per-protocol set (PPS)
  - Consistent results → increases confidence in trial results
  - Mentions results on PPS might be subject to severe bias
- ICH E9(R1) (2019) (Section A.5.3 (Supplementary Analysis))
  - "... May not be possible to construct a relevant estimand to which analysis of the PPS is aligned ..."
  - "... PPS does not achieve the goal of estimating the effect in any principal stratum, for example, in those subjects able to tolerate and continue to take the test treatment, because it may not compare similar subjects on different treatments. ..."

ICH E9(R1) on per-protocol set analyses

- Intercurrent events vs protocol deviations
  - A protocol deviation might not be an intercurrent event (e.g., visit outside time window)
  - An intercurrent event might not be a protocol deviation (e.g., patient death)
- Overall approach towards deriving estimand same as for any other situation
  - i.e. need to define intercurrent events & handling strategies etc
  - To obtain a „PPS“ estimand some intercurrent events might be handled differently (e.g. treatment adherence)
- “... Estimands might be constructed, with aligned method of analysis, that better address the objective usually associated with the analysis of the PPS. If so, analysis of the PPS might not add additional insights. ...”

What is the objective usually associated with the analysis of the PPS?

- ICH E9 (1998)

The 'per protocol' set of subjects, sometimes described as the 'valid cases', the 'efficacy' sample or the 'evaluable subjects' sample, defines a subset of the subjects in the full analysis set who are more compliant with the protocol and is characterised by criteria such as the following:

i) the completion of a certain pre-specified minimal exposure to the treatment regimen;

ii) the availability of measurements of the primary variable(s);

iii) the absence of any major protocol violations including the violation of entry criteria.

The use of the per protocol set may maximise the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol. However, the corresponding test of the hypothesis and estimate of the treatment

- Currently no established estimand emerged addressing these aims
  - hypothetical or principal stratum may be relevant strategies to approach treatment non-adherence?

Pre-Baseline protocol deviations

- Not meeting a trial entry criterion
  - This event cannot be caused by the administered treatment
  - In addition will be balanced (on average) across treatment arm (→ randomization)
  - No issue from a causal inference perspective to exclude these patients
- When removing patients that violate the inclusion criteria, estimand/analysis will target a slightly different population → closer to population in protocol
- Estimand/analysis including all patients may still be considered more generalizable to a broader population

Causal estimands, effect measures & parameter estimates in regression models

Treatment effects in ICH E9(R1)

- ICH E9(R1) defines treatment effects as

*Central questions for drug development and licensing are to establish the existence, and to estimate the magnitude, of treatment effects: how the outcome of treatment compares to what would have happened to the **same** subjects under alternative treatment (i.e. had they not received the treatment, or had they received a different treatment).*

- Is this a recommendation on how to define the treatment effect measure?
  - If yes, suggests to calculate the population summary measure under one treatment and compare to population summary measure under alternative treatment...
  - If yes, may challenge some of the treatment effect measures traditionally used in drug development

Treatment effects in ICH E9(R1)

- Examples ($E(.)$ is the expectation wrt the population of interest)
  1) Continuous data: $E[Y^{a=1}] - E[Y^{a=0}]$
  2) Binary data (odds ratio): $\frac{E[Y^{a=1}]}{1-E[Y^{a=1}]} / \frac{E[Y^{a=0}]}{1-E[Y^{a=0}]}$
  3) Count data (rate ratio): $E[Y^{a=1}]/E[Y^{a=0}]$
- All contrasts of marginal (population average) quantities
  - First calculate a population average summary per treatment group (averaging over population = „marginalization“) then calculate contrast to other treatment group
- Do standard regression-methods target these quantities when adjusting for baseline covariates X?

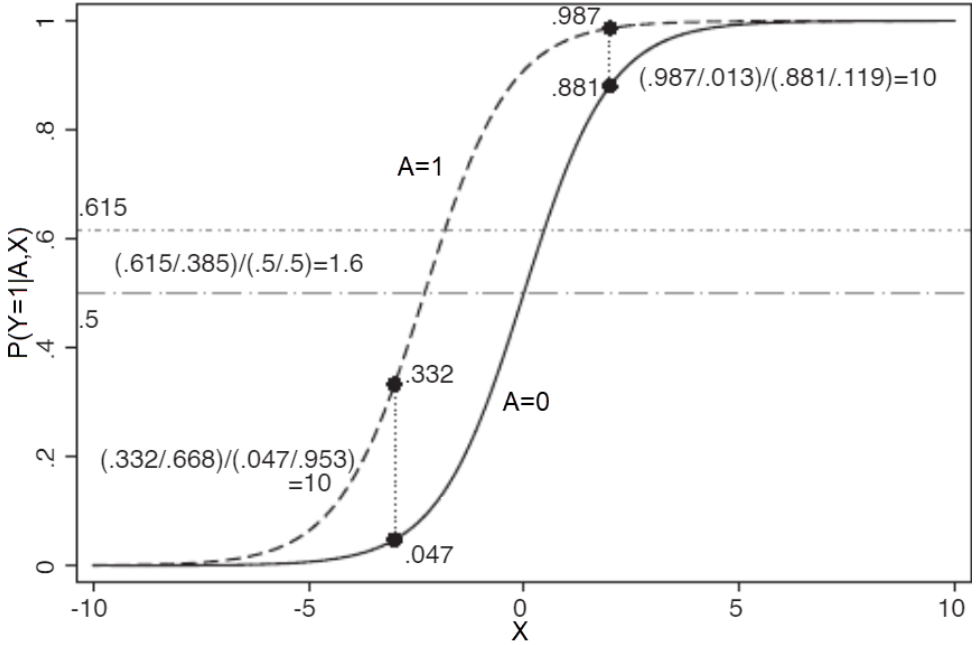
Estimands and estimation

- Given generalized linear model (GLM) model fit with link g and adjusted for covariates X: $\hat{E}[Y|A,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 X\}$
- Is the regression coefficient $\hat{\beta}_1$ an estimator of the marginal estimand from the previous slide?
  - $\hat{\beta}_1$ is usually called the conditional effect, as the analysis conditioned on X
  - Steingrimsson et al. (2017)

1. Linear regression (identity link): **yes**
2. Binomial regression (logistic link): **no** (targets „conditional odds ratio (OR)“)
3. Negative binomial regression (log link): **yes** (targets marginal risk ratio)

- For detailed calculations see Appendix

Illustration: Marginal versus conditional odds-ratio



Source: Daniel et al. (2021), Biometrical Journal

- Assume true model
  - logit(P(Y=1 | A , X) = log(10)A + X
  - X uniformly distributed on [-10,10]
  - For every X: true conditional OR=10
- Averaging response probabilities over X → marginal OR = 1.6
  - Difference due to averaging on probability scale → No major difference between trt and control on probability scale for X in [-10,-6] and [6,10] → ~40% of patients

Marginal versus conditional odds-ratio

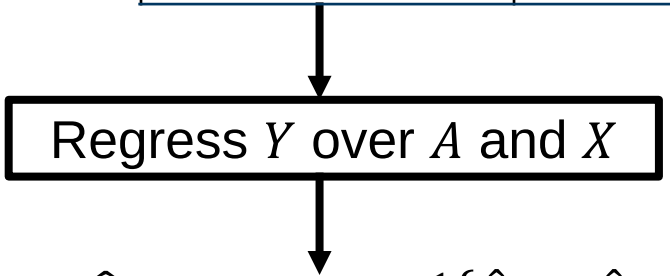
- Marginal and conditional odds-ratio different quantities
  - For every situation there is a „true“ marginal and conditional OR → separate quantities
  - One may prefer one over the other for reasons of interpretability (marginal OR may be more aligned with the definition in ICH E9(R1))
  - For both quantities covariate adjustment can be used (more later)
- How different are conditional and marginal OR?
  - Depends on impact of covariate on outcome and magnitude of treatment effect
  - But: Conditional OR generally larger than marginal ORs
- OR is not „collapsible“ as an effect measure → In general *identity* and *log-link* the only link functions that result in collapsible effect measures (see Appendix)

Standardization

- Treatment coefficient in logistic regression targets conditional and not marginal OR: Should we abandon logistic regression?
  - No: Fitted logistic regression model can be used to estimate a marginal OR using standardization (postprocessing the output of the model fit)
- Standardization
  1. Model fitting
  2. Predict outcome of *all* patients under *both* treatments
  3. Averaging
- Could be seen as a direct implementation of the definition of treatment effect in ICH E9(R1) → Predict the outcome of *all* patients on *all* treatments

Step 1 in standardization: Fit a regression model (e.g., GLM)

Treatment (A)	Covariates (X)	Response (Y)
1	x_1	y_1
0	x_2	y_2
⋮	⋮	⋮



Model fit: $\hat{E}[Y|A,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 X\}$

Step 2 in standardization: Predict potential outcomes

All patients under $a = 0$

Treatment $a = 0$	Covariate (X)
0	x_1
0	x_2
⋮	⋮

Model fit: $\hat{E}[Y|A, X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 X\}$

Predict

Potential response under $a = 0$
$\hat{E}[Y^{a=0}|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y^{a=0}|X = x_2]$
⋮

Step 2 in standardization: Predict potential outcomes

All patients under $a = 0$

Treatment $a = 0$	Covariate (X)
0	x_1
0	x_2
⋮	⋮

Model fit: $\hat{E}[Y|A, X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 X\}$

Predict

All patient under $a = 1$

Treatment $a = 1$	Covariate (X)
1	x_1
1	x_2
⋮	⋮

Potential response under $a = 0$
$\hat{E}[Y^{a=0}|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y^{a=0}|X = x_2]$
⋮

Potential response under $a = 1$
$\hat{E}[Y^{a=1}|X = x_1] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_1\}$
$\hat{E}[Y^{a=1}|X = x_2]$
⋮

Step 3 in standardization: Average over individual predictions

Averaging (marginalizing over covariates)

$$\hat{E}[Y^{a=0}] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(0)|X = x_i]$$

Averaging (marginalizing over covariates)

$$\hat{E}[Y^{a=1}] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y^{a=1}|X = x_i]$$

Estimated population average causal treatment effect:
A function of $\hat{E}[Y^{a=1}]$ and $\hat{E}[Y^{a=0}]$ (e.g. odds ratio)

Implementation of standardization

- SAS macro “Margins” fits the GLM or GEE model and estimates marginal mean and population average treatment effects (i.e., difference in means)
  - Compatible with GENMOD
  - Use the delta method for confidence intervals, p-values
  - https://support.sas.com/kb/63/038.html
- A general approach using bootstrap
  - Create bootstrap datasets using SURVEYSELECT in SAS or boot in R
  - Within each dataset, complete steps 1 (model fitting), 2 (predicting), 3 (averaging)
  - Summarize over the bootstrap datasets for confidence intervals, p-values

Regulatory feedback

- A PhIII clinical trial comparing treatment against control
- Primary estimand uses the marginal (population average) odds ratio $\frac{p_1}{1-p_1}/\frac{p_0}{1-p_0}$
  - p_1 and p_0 are population average response rates in treatment and control arms
- Primary analysis uses the logistic regression with covariates
  - Regression coefficient as the estimate of the primary estimand

FDA: Estimand uses the marginal odds ratio but the logistic regression uses the conditional odds ratio, which does not align with the estimand

LS-Means

- Often there is interest in presenting a summary measure per treatment arm
- Given the model fit: $\hat{E}[Y|A,X] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 X\}$, it is natural to use the least squares mean as an estimate of the population average/marginal mean $E[Y^a]$
  - LS mean plugs in the average of covariates $\hat{E}[Y|a, X = \bar{x}] = g^{-1}\{\hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 \bar{x}\}$
  - LS mean estimates the effect of a “special” patient with average values of covariates

LS-Means for different models

- Do following models provide(s) an LS mean $\hat{E}[Y|a, X = \bar{x}]$ that coincides with the marginal mean $\hat{E}[Y^a]$?
  - Linear regression (identity link): **Yes**
  - Logistic regression (logit link): **No**
  - Poisson / Negative binomial regression (log link): **No**
- Whenever the link function is non-linear these two quantities are different
- Estimated marginal mean $E[Y^a]$ can easily be extracted if standardization has been performed, provides an adequate population summary of the outcomes on the different treatments

Regulatory feedback on marginal mean

- SIROCCO, a PhIII trial to compare benralizumab to placebo for severe asthma
- Primary analysis uses standardization from negative binomial regression of the number of exacerbations with covariates

> *FDA: The study SAP proposed the marginal standardization method in calculating mean annual exacerbation rates.*

> *FDA: We agree with the applicant's proposal in that, in the negative binomial regression setting, the marginal method more closely aligns with the crude annual exacerbation rate, and as such, provides a more appropriate covariate-adjusted summary within treatment groups.*

https://www.fda.gov/media/110333/download

Properties of standardization

- Standardization derives population averages on the outcome scale
  - Coincides with the linear model estimator
  - More interpretable for discrete outcomes
  - Incorporates covariates for efficiency (→ more efficient estimate of marginal quantities compared to unadjusted estimates)
- Standardization is more robust (than a regression model) to model misspecification under randomization & provides a consistent estimator even when the GLM is misspecified (e.g., wrong choice of covariates), see Rosenblum and van der Laan (2010) for GLM
- Standardization provides flexible estimators for different effect measures (difference, ratio, odds ratio etc.)

Hazard ratio

- Hazard ratio estimated by Cox model also not collapsible
  - Not immediately obvious as a log-link is used (issue here: conditions on past survival)
- Causal interpretation difficult
  - If covariate & treatment have an effect, hazard ratio typically compares „different populations“ after baseline (hazard rate at time t conditions on being event-free up to time t)
- For an approach towards standardization see Daniel et al. (2021)
- A lot of very recent literature on this; triggered by Hernán (2010)
  - Aalen et al. (2015), Sjölander et al. (2016), Martinussen et al. (2018), ...

Consequences?

- Issues well-known at least since the early 1980s (see references in Gail, 1984)
- Cox proportional hazards model and log-rank test provide valid tests of the null hypothesis of no treatment effect
  - May consider separating testing and estimation and use a different effect measure?
  - Alternative effect measures have been discussed (e.g. restricted mean survival time); so far not one particular alternative approach emerging
  - Plot of Kaplan-Meier curves (or cumulative incidence curves) usually provide a good overview of treatment effect (potentially stratified by important prognostic covariates)

Why do we need causal inference in drug development?

- Provides a language to discuss causal effects (potential outcomes & DAGs)
  - applies in observational and randomized data situations
- Understand properties of estimators for a better alignment with estimands
  - What estimand is the chosen estimator targeting?
  - What are the assumptions underlying the estimator and how plausible are they?
- Sheds new light on the understanding of some standard statistical practices
  - LS means, interpretability of treatment effect parameters
- Will help implementing the ICH E9 addendum
  - Adopts counterfactual viewpoint to define treatment effects
  - Causal thinking & techniques apply to all intercurrent event strategies (Lipkovich et al (2020))

References

- ICH E9 Addendum and Causal inference
  - National Research Council. (2010) The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: The National Academies Press. https://doi.org/10.17226/12955
  - ICH (2019) Addendum on estimands and sensitivity analysis in clinical trials E9(R1) E9-R1_Step4_Guideline_2019_1203.pdf (ich.org)
  - Lipkovich, I. et al (2020) Causal Inference and Estimands in Clinical Trials, Statistics in Biopharmaceutical Research, 12:1, 54-67, DOI: 10.1080/19466315.2019.1697739
- Collapsibility
  - Daniel, R. et al (2021) Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets, to appear Biometrical Journal, https://doi.org/10.1002/bimj.201900297
- Standardization
  - Rosenblum, M., & Van Der Laan, M. J. (2010). Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *The international journal of biostatistics*, *6*(1).
  - Steingrimsson, J. A., Hanley, D. F., & Rosenblum, M. (2017). Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemporary clinical trials*, *54*, 18-24.
- Hazard ratio
  - Aalen, O. O., Cook, R. and Røysland, K. (2015) Does Cox analysis of a randomized survival study yield a causal treatment effect? Lifetime Data Analysis 21, 579-593
  - Gail, M. H., Wieand, S., & Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika, 71, 431-444.
  - Hernán, M. A. (2010) The Hazards of Hazard Ratios, Epidemiology, 21, 13-15.
  - Martinussen, T., Vansteelandt, S. & Andersen P.K (2018) Subtleties in the interpretation of hazard ratios, arXiv:1810.09192
  - Sjölander, A., Dahlqwist, E. and Zetterqvist, J. (2016) A Note on the Noncollapsibility of Rate Differences and Rate Ratios, Epidemiology, 27, 356-359.

Q & A

Estimand of linear regression coefficient

- Linear regression: $g^{-1}\{\cdot\} = \cdot$

$$\hat{E}[Y(1)] - \hat{E}[Y(0)] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(1)|X = x_i] - \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(0)|X = x_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_2 x_i\right)$$

$$= \hat{\beta}_1 \text{ (Yes)}$$

- Targets marginal quantity

Estimand of logistic regression coefficient

- Logistic regression: $g^{-1}\{\cdot\} = \frac{\exp(\cdot)}{\exp(\cdot)+1} \equiv \text{expit}(\cdot)$

$$\frac{\hat{E}[Y(1)]}{1-\hat{E}[Y(1)]} \Big/ \frac{\hat{E}[Y(0)]}{1-\hat{E}[Y(0)]} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X=x_i]}{1-\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X=x_i]} \Big/ \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X=x_i]}{1-\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X=x_i]}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2 x_i)}{1-\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2 x_i)} \Big/ \frac{\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_2 x_i)}{1-\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_2 x_i)}$$

$$\neq \exp\hat{\beta}_1 \text{ (No)}$$

- Note that if we plug in the mean of covariate X

$$\frac{\hat{E}[Y(1)|X=\bar{x}]}{1-\hat{E}[Y(1)|X=\bar{x}]} \Big/ \frac{\hat{E}[Y(0)|X=\bar{x}]}{1-\hat{E}[Y(0)|X=\bar{x}]} = \frac{\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2\bar{x})}{1-\text{expit}(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2\bar{x})} \Big/ \frac{\text{expit}(\hat{\beta}_0+\hat{\beta}_2\bar{x})}{1-\text{expit}(\hat{\beta}_0+\hat{\beta}_2\bar{x})}$$

$$= \frac{\exp(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2\bar{x})}{\exp(\hat{\beta}_0+\hat{\beta}_2\bar{x})}$$

$$= \exp\hat{\beta}_1 \text{ (i.e., conditional effect on the mean of covariate)}$$

Estimands of Poisson/negative binomial coefficient

- Poisson / Negative binomial regression: $g^{-1}\{\cdot\} = \exp(\cdot)$

$$\frac{\hat{E}[Y(1)]}{\hat{E}[Y(0)]} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(1)|X=x_i]}{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(0)|X=x_i]}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\exp(\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2 x_i)}{\frac{1}{n}\sum_{i=1}^{n}\exp(\hat{\beta}_0+\hat{\beta}_2 x_i)}$$

$$= \exp\hat{\beta}_1 \text{ (Yes, under the following assumptions)}$$

- Note that $\frac{\hat{E}[Y(1)]}{\hat{E}[Y(0)]}$ is the ratio of rates assuming every patient would have the same exposure (or offset)
- Also assume no Z by X interactions

Estimand of linear regression LS mean

- Linear regression: $g^{-1}\{\cdot\} = \cdot$

$$\hat{E}[Y(z)] = \frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(z)|X = x_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i\right)$$

$$= \hat{\beta}_0 + \hat{\beta}_1 z + \hat{\beta}_2 \bar{x}$$

$$= \hat{E}[Y|z, X = \bar{x}] \text{ (Yes)}$$

Estimand of logistic regression LS mean

- Logistic regression: $g^{-1}\{\cdot\} = \frac{\exp(\cdot)}{\exp(\cdot)+1} \equiv \text{expit}(\cdot)$

$$\hat{E}[Y(z)] = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(z)|X=x_i]}{1-\frac{1}{n}\sum_{i=1}^{n}\hat{E}[Y(z)|X=x_i]}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1 z+\hat{\beta}_2 x_i)}{1-\frac{1}{n}\sum_{i=1}^{n}\text{expit}(\hat{\beta}_0+\hat{\beta}_1 z+\hat{\beta}_2 x_i)}$$

$$\neq \text{expit}(\hat{\beta}_0+\hat{\beta}_1 z+\hat{\beta}_2 \bar{x})$$

$$= \hat{E}[Y|z, X=\bar{x}] \text{ (No)}$$

- Interpretation: LS mean estimates the effect of a “special” patient with average values of covariates

Estimands of Poisson/negative binomial regression LS mean

- Poisson / Negative binomial regression: $g^{-1}\{\cdot\} = \exp(\cdot)$

$$\hat{E}[Y(z)] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[Y(1)|X = x_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \exp\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x_i\right)$$

$$\neq \exp\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 \bar{x}\right)$$

$$= \hat{E}[Y|z, X = \bar{x}] \text{ (No)}$$

- Interpretation: LS mean estimates the effect of a “special” patient with average values of covariates

Thank you

NOVARTIS | Reimagining Medicine