Commentary on van Lancker et al

Frank E Harrell Jr

Van Lancker et al. rightfully urge more clinical trial designs to use analysis of covariance. But they recommend that traditional analysis of covariance (ANCOVA) not be used, in favor of using more complex marginal treatment effect estimates adjusted for covariates. The authors deem traditional ANCOVA non-robust, and marginal effect estimates robust. As stated by the authors, marginal effects assume that the randomized trial participants were randomly sampled from the clinical population. This is almost never the case in clinical trials as it would imply forced participation and abandoning inclusion/exclusion criteria. The almost sure violation of the authors' key assumption renders marginal estimates biased, and this bias is very likely to be worse than any damage caused by standard ANCOVA model misspecification.

The main thrust of the paper by van Lancker, Bretz, and Dukes in this issue[1] was covered in their third reference by Benkeser et al.[2] and for which Stephen Senn and I wrote a detailed critique.[3] To give context to my comments, a gold standard for assessing treatment efficacy on a given patient is a multi-period crossover study in which a patient is compared with herself. Parallel-group studies should mimic this standard to the extent possible by following the advice of Mitchell Gail to emphasize conditional estimates. Gail was quoted as follows in Hauck et al:[4]

> For use in a clinician–patient context, there is only a single person, that patient, of interest. The subject-specific measure then best reflects the risks or benefits for that patient. Gail has noted this previously in his ENAR Presidential Invited Address in April 1990, arguing that one goal of a clinical trial ought to be to predict the direction and size of a treatment benefit for a patient with specific covariate values. In contrast, population–averaged estimates of treatment effect compare outcomes in groups of patients. The groups being compared are determined by whatever covariates are included in the model. The treatment effect is then a comparison of average outcomes, where the averaging is over all omitted covariates.

In a linear model where there is no constraint on the treatment effect (difference in means), both an unconditional (unadjusted for covariates) model and a covariate-adjusted model estimate the same quantity. The unadjusted model just estimates it inefficiently by absorbing the variance that could have been explained by covariates into the error term, reducing power and precision. Nonlinear models that have constraints on the treatment effect and that have no error term to absorb omitted covariates (e.g. the binary logistic model) behave altogether differently as detailed in my tutorial on traditional ANCOVA.[5] Covariate adjustment accounts for easily explainable outcome heterogeneity and prevents a loss of power due to such heterogeneity when covariates are omitted. The net effect is that, in all situations, covariate adjustment is beneficial to power and to getting more relevant effect estimates.

Marginal (population average) treatment effect estimates are useful only for group decision-making and do not provide estimates that are relevant for a single patient type. In nonlinear models, population average effects may be estimated by sample average effects, but only if one of two criteria is satisfied: (1) the sample is a simple random sample from the population or (2) the sample is a stratified probability sample with known sampling probabilities which are used in a weighted analysis. These criteria are rarely if ever satisfied in randomized controlled trials (RCTs) which do not coerce subjects into a "sample."

The base assumption of random sampling from the clinical population means that van Lancker et al. can estimate the marginal treatment effect easily using sample average effects. Ignoring for the moment the earlier stated disadvantages of unconditional treatment effects, the random sampling assumption has far-reaching consequences. These consequences stem from "risk magnification," that is, absolute risk reductions tend to increase as baseline risk increases, since low-risk patients have little room to move. An RCT enrolling

Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA

Corresponding author:
Frank E Harrell Jr, Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.
Email: f.harrell@vumc.org

low-risk and high-risk patients may lead to a sample average treatment effect that applies to medium-risk patients who were never enrolled. The results from an RCT enrolling primarily medium- to high-risk patients would not apply to a clinical population that included many low-risk patients. At the heart of this problem is the fact that the absolute risk reduction is a quantity that needs to be applied to individual patient types. When there are continuous prognostic covariates, there can be as many distinct risk reductions as there are patients. Presenting these risk reductions graphically can provide important information that is hidden in marginal estimation.

The presence of distinct patient-specific treatment effects is much easier to handle than it appears at first glance because with standard regression modeling, the number of free parameters is much less than the sample size, consisting of the treatment parameter and a parameter for each covariate term if there are no treatment interactions in the model. For a binary outcome, a conditional relative treatment effect, the odds ratio, may be a single number (in the absence of interactions). But on the highly useful absolute risk scale, there are multiple estimates, and results should be presented that way.[6]

Conditional regression models, with all the flexibilities that have been developed over the past 40 years (flexible nonlinear covariate effects, penalization, etc.), provide low mean squared error estimates of the patient-type–specific absolute treatment effects (average risk reduction in logistic models; differences in cumulative incidence in Cox models). They also provide the clearest way to assess differential treatment effects through interactions. For estimating interactions, regression modeling provides two clear modern approaches: fitting Bayesian models with pre-specified priors on interaction effects or fitting penalized frequentist models where several covariate main effects are adjusted without shrinkage, but all the treatment-by-covariate interaction parameters are penalized. In a case study where the RCT sample size (30,000) is large enough to allow for meaningful interaction analysis, the optimum penalty for all the treatment interaction terms was infinity, indicating no evidence for interaction on the log odds scale.[7] This finding was supported by computing the Akaike information criterion for a model allowing all interactions versus a model with only main effects. The criterion worsened when allowing for interactions, indicating that estimation of patient-type–specific treatment effects with interactions was less accurate than just using a single odds ratio.

Van Lancker et al. repeatedly cast doubt on the validity of traditional covariate adjustment due to possibly violated model assumptions. This is misleading. Tukey showed that even highly suboptimal covariate adjustment is much better than unadjusted treatment effect estimation.[8] Covariate adjustment need not be perfect. Anything is better than unadjusted analyses, for example, in studies that pretend that young patients die as soon as elderly patients, as many COVID-19 studies did. Covariate adjustment also frequently makes other model assumptions such as proportional hazards more likely to hold. Methods advocated by van Lancker et al. and others have a more serious problem than model misspecification: systematic bias when the sample is not representative of the population (as it is not intended to be in an RCT). I have examined this issue in detail, including simple simulations studying the effect of misspecification of covariate form on treatment effect, power, and type I assertion probability in a binary logistic model.[9]

I have also provided an example where an ordinal outcome is analyzed with covariate adjustment for a baseline variable that operates in a highly non-proportional odds fashion, the result being that the covariate adjustment effectively ignored that covariate when assessing the treatment effect.[10] Thus, contrary to the claims of van Lancker et al., the regression model does not need to be approximately correct for its conditional effect estimates to transport well to broader populations. The better the model captures important covariates and models them correctly, the better estimation of course. But any covariate adjustment is better than none and is better than using marginal estimates with false RCT representativeness assumptions. Similarly, the authors' claim that "an unadjusted estimator without covariates…yields an unbiased point estimate" is easily demonstrated to be incorrect, due to the non-random sampling used in all RCTs.

ORCID iD

Frank E Harrell Jr https://orcid.org/0000-0002-8271-5493

References

1. van Lancker K, Bretz F and Dukes O. Covariate adjustment in randomized controlled trials: general concepts and practical considerations. *Clin Trials* 2024; 21(4): 399–411.
2. Benkeser D, Dıaz I, Luedtke A, et al. Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics* 2020; 77: 1467–1481.
3. Harrell F and Senn S. Commentary on improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes, fharrell.com/post/ipp (2021, accessed 11 February 2024).
4. Hauck WW, Anderson S and Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clin Trials* 1998; 19(3): 249–256.
5. Harrell FE Jr. Analysis of covariance in randomized studies. In: *Biostatistics for biomedical research*, hbiostat.org/bbr/ancova (2024, accessed 11 February 2024).
6. Harrell F. Avoiding one-number summaries of treatment effects for RCTs with binary outcomes, fharrell.com/post/rdist (2021, accessed 11 February 2024).
7. Harrell F. Assessing heterogeneity of treatment effect, estimating patient-specific efficacy, and studying variation in odds ratios, risk ratios, and risk differences, fharrell.com/post/varyor (2019, accessed 11 February 2024).
8. Tukey JW. Tightening the clinical trial. *Control Clin Trials* 1993; 14: 266–285.
9. Harrell F. Incorrect covariate adjustment may be more correct than adjusted marginal estimates, fharrell.com/post/robcov (2021, accessed 11 February 2024).
10. Harrell F. Assessing the proportional odds assumption and its impact, fharrell.com/post/impactpo (2022, accessed 11 February 2024).