Afterword

Afterword: Sample Design Is the Key

In a Methods Corner article in a recent issue of *Health Services Research*, Dowd, Greene, and Norton (2014) proposed a formula for the asymptotic variance of an estimator of the mean prediction (or difference of predictions) over a population from a nonlinear model. Such mean estimands are widely used in health services research when predicting average treatment effects from logistic, Poisson, and other nonlinear regression models. In these two-step calculations, the researcher first estimates the model parameters (“estimation step”) and then uses them to predict outcomes or treatment-related outcome differences (“prediction step”), which are then averaged over a sample.

In this issue, we present a commentary by Terza on Dowd et al. and a rejoinder by the original authors. Terza (2016) argues that the formula of Dowd, Greene, and Norton (2014) omits an important second term and therefore systematically underestimates the relevant variances. His two-term formula applies when the predictor and outcome variables are derived from a common sample, almost certainly the most common, “default” research design in health services research. In their rejoinder, Dowd et al. defend the relevance of their original formula under a different design assumption.

The controversy initially took the form of a disagreement over the correct formulation of asymptotics, that is, the limiting behavior of the estimators as the sample size n increases. An exchange of views among the authors and an editor revealed its roots in assumptions about sample design.

Heuristically interpreting the two terms of Terza’s formula (6), the first term, equivalent to the sole term of Dowd et al.’s formula, multiplies the variance of the model parameter estimator by derivatives of the predicted outcome effect with respect to those parameters, thus representing the variance of prediction due to parameter estimation. The second term, omitted from Dowd et al.’s formula, is the variance of mean predictions under sampling of the

cases providing the covariate (“ X ”) values, fixing the parameters at a point estimate.

Discussion of asymptotics suggests consideration of the sample or samples used in estimation as n increases. Most typically, in observational studies, a single sample is drawn from a larger population to support estimation of both model parameters and predicted means in the same population. In this case, n is the same for estimation of the variance of the parameter estimates used in the first term as for the summations in both terms of Terza’s formula, which assumes this common design.

Nonetheless, in the two-step calculation considered here, the samples involved in the two steps could differ. In their rejoinder, Greene, Dowd, and Norton (2016) explicitly assume that the target estimands are predictions for a fixed finite population that is entirely observed (possibly a single unit), while the predictive model is estimated from repeated observations on the same population of units (or replications with exactly the same distribution of the predictors X) with the outcome subject to independent random variation over repetitions. Under such a design, they argue that the distribution of X is fixed and the second term vanishes.

While the design described by Dowd et al. in their rejoinder is unusual in practice, other research designs may likewise entail different samples for parameter estimation and prediction. These include designs in which outcomes Y are measured only in a study sample, but the covariates X are routinely available from claims or other administrative systems for the entire population of interest, which may be larger than or even entirely distinct from the study sample. Only the study sample can be used for parameter estimation, but the entire population can be used in prediction. For example, an intervention might have a heterogeneous effect modeled in a sample as a function of age, sex, and comorbidity status. Potential benefits of implementation in another population are projected using parameters and their covariances from the original study, applied under a model to the distribution of covariates from files covering the entire target population or a sample from it, yielding a point prediction and measure of uncertainty. Similarly, a two-step procedure might be used for standardization of outcomes in several groups to “recycled predictions” for a preselected reference distribution of X , for example, in quality reporting (comparing health care units) or disparities calculations. Modifi-

Address correspondence to Alan M. Zaslavsky, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115; e-mail: zaslavsk@hcp.med.harvard.edu.

cations of the variance formulae required for these and other cases are beyond the scope of this commentary.

In conclusion, this exchange illustrates the importance of correct technical methods interpreted specifically to the design and analysis at hand. For the “default” case, in which parameter estimation and prediction use the same sample, we recommend substituting parameter estimates and associated covariance estimates based on that sample into Terza’s equation (6), taking summations over the same sample.

Alan M. Zaslavsky, Senior Associate Editor

REFERENCES

- Dowd, B. E., W. H. Greene, and E. C. Norton. 2014. “Computation of Standard Errors.” *Health Services Research* 49 (2): 731–50.
- Greene, W. H., B. E. Dowd, and E. C. Norton. 2016. “Response to ‘Inference Using Sample Means of Parametric Nonlinear Data Transformations’.” *Health Services Research* 51 (3 Pt 1): 1114–6.
- Terza, J. V. 2016. “Inference Using Sample Means of Parametric Nonlinear Data Transformations.” *Health Services Research* 51 (3 Pt 1): 1109–13.