When submitting research papers to *Frontiers in Epidemiology*, authors should clearly address in their cover letter in which domain or tasks of health data sciences **the aim** of their paper falls. The three domains or tasks are description, prediction, or causal inference (including counterfactual prediction modeling). All three domains are important and require specific method applications and reporting. For an overview of the three tasks and how to identify in which category your research question falls, see Hernán et al. (1). Sometimes classification may not be straightforward (for example because within one study multiple tasks are addressed), in which case this should be explained in the cover letter. In general, we recommend following reporting guidelines from the EQUATOR network (2).

The aim of descriptive tasks is broadly to summarize a phenomenon in a compact way. Analytic tools in this task range from the calculation of proportions or means to more complex techniques, such as unsupervised learning algorithms (to identify clusters in large data sets). Modern data visualization can also be used (1). Ecologic studies also fall in this category. Assessment of association can be, in some particular instances, justified.

The aim of prediction tasks is to forecast values of a target variable, by relying on the association with other variables. The analytics used for the prediction studies range from calculating simple correlations to sophisticated prediction models to identifying risk groups for an outcome of interest (random forest, penalized regression, neural networks, survival models, and others). Reporting of prediction models should follow current guidelines (3). We warn against interpreting selected predictors as causal factors (i.e., risk factors). Studies that aim to develop, validate, and update prediction models fall in this category.

Causal inference (or counterfactual prediction) addresses questions of cause and effect. The type of studies generally aims at quantifying causal relationships. Proper causal inference methods should be applied when analyzing data in the causal inference domain (4). Observational studies (e.g. case-control studies, cohort studies) in which the aim is to assess the association between an exposure and an outcome adjusting for confounding fall in this category. In these types of studies, we encourage the use of causal language when defining the objective and the causal estimand (5). We encourage the submission of directed acyclic graphs if they were used to inform analytical strategies. The analytics employed for causal inference range from elementary calculations in randomized experiments to regression models and complex implementations of g-methods in observational studies (1).

If research papers deviate from these three domains or mix them, the authors should give clear reasoning for doing so.

References


1. Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *Chance* (2019) **32**:42–49.
2. The EQUATOR Network. https://www.equator-network.org/ [Accessed February 17, 2022]
3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. https://www.equator-network.org/reporting-guidelines/tripod-statement/ [Accessed February 16, 2022]
4. Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC (2020).
5. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am J Public Health* (2018) **108**:616–619.