
Olga A. Vsevolozhskaya
*University of Kentucky, vsevolozhskaya@uky.edu*

Fernando A. Wagner
*Morgan State University*

James C. Anthony
*Michigan State University*

Click here to let us know how access to this document benefits you.

Follow this and additional works at: [https://uknowledge.uky.edu/biostatistics_present](https://uknowledge.uky.edu/biostatistics_present)

Part of the *Biostatistics Commons*

Repository Citation


[https://uknowledge.uky.edu/biostatistics_present/5](https://uknowledge.uky.edu/biostatistics_present/5)

This Presentation is brought to you for free and open access by the Biostatistics at UKnowledge. It has been accepted for inclusion in Biostatistics Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Abstract

Aims: At CPDD 2015, we applied parametric Hill functions to estimate the probability of drug dependence in relation to the duration of drug-taking experience. A problem we and others have encountered in the estimation of risk of becoming a drug-dependence case is an observed point estimate of zero – the so-called “zero-numerator problem.” This problem can be easily observed in certain low risk subgroups even when the sample is large (e.g., the incidence of heroin dependence among 62-year-old newly incident heroin users) or with small subgroup sample sizes. In these instances, an observed zero point estimate does not necessarily imply zero risk of developing dependence for the subgroup. Here, our aim is to describe our approach to a potential solution to the zero-numerator problem based on a Bayesian model in conjunction with parametric Hill functions.

Methods: The traditional frequentist statistical approach can provide an estimate for the 95% upper bound of an incidence rate even when there are observed zeros in the numerator. A Bayesian approach is required if estimation of the incidence rate itself is of interest. The Bayesian approach demands specification of a prior distribution for the risk parameter. In this work, we employ the sensitivity of the Hill function parameter estimates to the choice of a particular informative prior distribution across a range of estimated chances of developing drug dependence very soon after onset of drug use.

Conclusions: Where we frame our work in relation to risk of developing drug dependence syndromes, the zero-numerator problem is often faced in other contexts (e.g., pharmacokinetics, toxicology). Our approach, combining Bayesian statistics in conjunction with Hill functions, is expected to provide a useful solution to these zero-numerator problems.

The Zero-Numerator Problem

Hill Function Parameters

Consider data from United States (US) National Surveys on Drug Use and Health (NSDUH) 2004 – 2013, over n = 1,513 (unweighted) subjects with smoking onset within 5-months of assessment, who had smoked at least once during the past 30 days.

Different Choices for Informative Priors

Often researchers want the data ‘to dominate’ and thus assign a prior probability of an event that is ‘uninformative’ or vague in some sense. However, if one puts vague prior distributions on the parameter values, it is easier to see why an estimate is zero-like after X smoking days past month – an unlikely scenario in the zero-numerator setting. Additionally, with a correctly specified informative prior, Bayesian inference is not susceptible to selection bias, e.g., how many smoking days past month is associated with the highest rate of nicotine dependence? or to multiple comparisons. Next, we will look at the role of different informative priors on the results in zero-numerator problems.

Using the above algorithm, the posterior expectations of nicotine dependence for different W’s are illustrated in Figure 1. Since the posterior expectation is a weighted average between the prior and the posterior means, the width of the ‘rolling window’ affects the results. If the window contributing to the prior knowledge of nicotine dependence is too wide (W = 30), the posterior expectation is dominated by the prior overall mean (flat line in the left plot of Figure 1). If the window is narrow, e.g., W = 1 or W = 2, the posterior probabilities are sensitive to day-to-day variability in the empirical chances of dependence.

Regardless of the choice of W, the zero numerator problem is completely eliminated. So which value of W should one use in practice? The answer can be obtained via leave-one-out cross-validation, which in our case finds W = 2 to be the optimal value. The right plot of Figure 1 illustrates the posterior expectations of nicotine dependence (with the corresponding 95% credible intervals) and the weighted empirical estimates from NSDUH. Note the overlap in the 95% credible intervals and 95% confidence intervals.

Hill Function Parameters

We propose Beta(a, b) priors with a and b chosen to reflect prior knowledge about p – the probability of dependence after X smoking days. To capture this knowledge, we consider a ‘rolling window’ across W and X + W days. The parameters a and b are obtained as follows:

- Assume a uniform Beta(1, 1) distribution of dependence probability over the X and W smoking days window.
- Note, the information at X smoking days is excluded from the likelihood formation.
- The posterior probability of dependence over the X and W smoking days follows Beta(a, b) distribution with:
  \[ a = (W - 1) \text{ of dependent cases after } X - 1, X, X + 1, \ldots, X + W \text{ smoking days} \]
  \[ b = (W - 1) \text{ of subjects with dependence over the same window} + 1 \]
- Under the assumption of common p – the probability of dependence over the X and W smoking days window, the posterior Beta(a, b) becomes prior probability of dependence after X smoking days past month.

Funding Source and Disclosure

Disclosure

There are no relevant financial interests to disclose. Funding Source

Supported via a National Institute on Drug Abuse Senior Scientist JCA: K05DA015799.