# OBSERVATIONAL STUDIES OF RARE EVENTS:
# A SUBSET SELECTION APPROACH

## Abstract

We propose a subset selection procedure for use in observational studies of rare events. Departing from the traditional selection applications, which necessarily involve experimental settings, these problems are characterized by large, uncontrollable sample sizes and extremely low incidence rates. Although these characteristics impose increased analytical and computational burdens, they are readily apparent in many problems of practical importance and thus warrant investigation. To this end we present an approximation of a bound on the selection probability function. We empirically demonstrate its quality and conclude our study by offering methods for determining the size of the selected subset that are consistent with the observational nature of our problem context. We begin with an analysis of the behavior of a bound on the selection probability function as the number of populations from which the selection is to be made increases. A simply defined function that emerges from this asymptotic analysis can significantly reduce the computational burden associated with the selection procedure. We illustrate the performance of the approximation using data associated with a study of urban traffic hazards, for which sample sizes vary by an order of magnitude and incidence rates are virtually 0. These empirical results indicate that our asymptotically based function provides a remarkably accurate approximation of the bound even when a reasonably small number of populations are considered. These data also demonstrate that the bound may be quite conservative, but that judgments concerning the composition of the selected subset are similar to those that would be obtained by estimating the achieved selection probability.

## Full citation: