Investigations as Sampling: Implications to Program Efficacy

> "…we have learned that the number of known crimes is much higher than the number of reported crimes-often double. In other words, over the last thirty-year period, individuals are not reporting half of the crimes that are occurring." – Shima Baughman[1]

Abstract

Investigative programs may be susceptible to an organizational-version of the Dunning-Kruger effect; a well-functioning investigative program is required to determine how effective that program is at detecting worthy behavior. The process of generating investigative leads could be constructed as a non-random sampling task. In the space of all behaviors occurring within a jurisdiction at any given time, a subset of those behaviors are available to be surveilled, and a further subset are surveilled. Assuming the relationship between targeted behavioral characteristics such as "violent" or "criminal," availability of those behaviors for surveillance, and the probability of those behaviors being surveilled are not independent, some magnitude of selection bias drives-up the probability of surveilling behavior with targeted characteristics. Thus, when inferring from observed behavior back to the population of all behaviors within an organization's jurisdiction, this bias should be considered. Simulation study outlined here suggests these sampling effects strictly decrease the estimated prevalence of targeted behaviors, can carry non-trivial magnitude, and can outweigh the impacts of increasing the proportion of available behavior surveilled under certain conditions. Methods for detecting and adjusting for this bias are discussed in the context of a sampling theory of investigation, and the implications to performance assessment of investigative programs are discussed.

Theory

$$Population\ Rate \propto Observed\ Rate * \frac{Coverage}{Sensitivity\ Factor}$$

Where:

- Population Rate is defined as the rate of target-bearing behavior in the population of behaviors.
- Observed Rate is defined as the observed rate or target-bearing behavior.
- Coverage is defined as the marginal probability of a behavior being observed.
- "Sensitivity Factor" refers to the likelihood ratio of target-bearing behavior being observed to non-target-bearing behavior being observed.

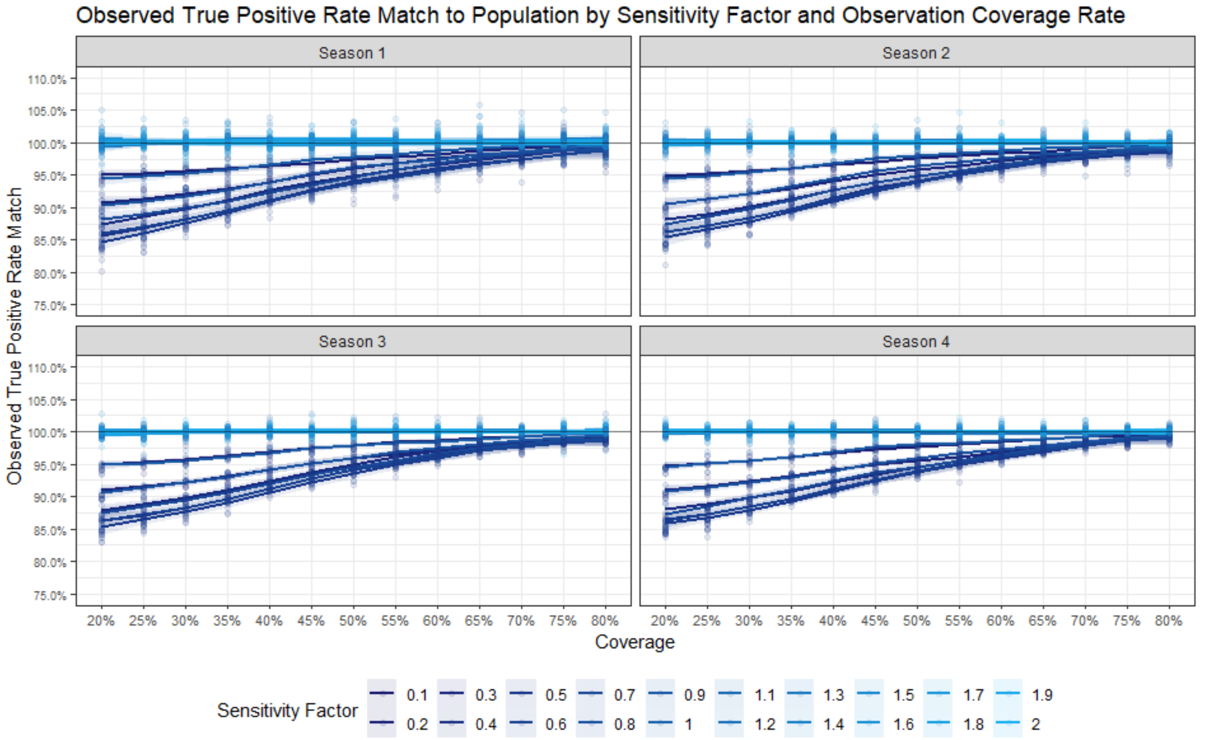

Figure 1: Comparing the influence of the "sensitivity factor" on observed positivity rate unadjusted accuracy by coverage.

Change Rate	Sensitivity Factor Standardized Linear Coefficient by Coverage Region		
	20% – 40%	40%-60%	60% – 80%
$\frac{\Delta Observed\ Match\ Proportion}{\Delta Coverage}$	*-0.69783*	*-0.71662*	*-0.74147*

Discussion

Figure 1 displays four "seasons" during which a simulated investigative organization had the opportunity to observe behaviors, from 10,000 observable behaviors in Season 1 to 40,000 in Season 4, under various coverage and sensitivity factor conditions. As coverage increases along the x-axis, the negative bias associated with lower coverage improves towards 100% of the population rate. However, increases to the sensitivity factor are associated with faster jumps towards unbiased estimation; sensitivity factors over 1 were able to reliably compensate for low coverage. This finding implies that improving the sensitivity rate of existing investigators may be more efficient at capturing the true population rate than investing in additional investigators. In addition, improving sensitivity is estimated to decrease the impact of additional coverage on estimates of the population rate, as implied by the negative standardized linear regression coefficients highlighted in the accompanying table. This finding implies that sensitivity and coverage form dueling priorities in terms of efficient deployment of investigative resources.

Sources

1. Baughman, S. B. (2020). How Effective Are Police? The Problem of Clearance Rates and Criminal Accountability. SSRN Electronic Journal, 47–112. https://doi.org/10.2139/ssrn.3566383