7.4.2-3: POPULATION MODEL – RARE EVENTS, FALSE POSITIVE PARADOX, SMALL SAMPLES, AND SUPERPOPULATION CONCEPT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

7.4.2-3. RARE EVENTS, THE FALSE POSITIVE PARADOX, SMALL SAMPLES, AND THE CONCEPT OF A SUPERPOPULATION

7.4.2-3.1. Study Background

Rare events and small sample sizes represent a major analytic challenge. The data and studies being assembled to support the MRTPA are a case in point, in which rare events in combination with small numbers have the potential to distort study results, particularly if they are used to measure relative differences.

7.4.2-3.2. Objective

This paper outlines a strategy for minimizing the distortion of study results resulting from the presence of rare events and small sample sizes. I first provide an overview of four distinct, but related concepts: (1) rare events; (2) the false positive paradox; (3) small samples; and (4) a super population. Following this overview, I describe linkages among these four concepts and then go on to illustrate them and their linkages with MRPTA and NSUDH data. I conclude with a brief discussion of how these ideas may prove useful for minimizing the distortion of study results resulting from the presence of rare events and small sample sizes.

7.4.2-3.3. Rare Events

Rare events do not occur in isolation. They occur with a backdrop and the backdrop is made up of experience and convention. That is, a rare event is defined relative to one's experience and conventions, with the latter largely comprised of assumptions. This is partly why there is no hard and fast definition of a rare event. However, this is not to say that the concept of a rare event has not been implemented. In fact, it provides the foundation for the classical approach to inferential hypothesis testing. If the probability of observing an outcome for a given statistical model is "low," then the model is rejected. The two deciding factors guiding the selection of $p = .05$, $p = .01$, $p = .001$, or some other "low" probability, are experience and convention.

So, while there are problems in applying quantitative models to rare events e.g., (McMorrow, 2009), which, in turn, lead to problems in trying to predict and estimate rare events, the fact that the idea of rare events in the form of probabilities has also found a long-time home in inferential statistics suggests that they can be dealt with. That is, we can use models and conceptual frameworks found in inferential statistics to extract information from data in which rare events are embedded, e.g., (Gelman, King, & Boscardin, 1998). Relevant to the discussion here are three important conceptual frameworks, the "false positive paradox," "small samples," and the idea of a "super population."

7.4.2-3.4. The False Positive Paradox

The false positive paradox is found when a rare event is analyzed. Specifically, when an event is rare, it is often found at a higher frequency than expected, given the probability of its occurrence. That is, false positive results are found when the population has a low incidence of a condition of interest. Thus, the probability of a positive test result is determined not only by the accuracy of the "test" but by the characteristics of the population. When the incidence in the population is lower than the test's false positive rate, even tests that have a very low chance of giving a false positive in an individual case will give more false than true positive results – see Brown (2011).

What constitutes a "test?" It could be a physical test, an observation, a self-administered questionnaire, or an interview, among other possibilities. How can we implement such a test? We can select a sample.

7.4.2-3.5.Small Samples

There is a great deal of theory and experience to guide sample design and selection. However, not all of the theory and experience can be put into play. This may come about because a new product, one without any data history needs to be studied. It may come about because there is no sample frame available. Or, it may come about because of a combination of these two factors and others, such as cost or time constraints. The latter may lead to the use of "small samples," which like rare events, occur with a backdrop made up of experience and convention. That is, a sample is defined relative to one's experience and conventions. One of the major conventions in small sample theory provides a definition of a small sample ($n < 30$) and guidance in regard to which inferential tests to use when a small sample is encountered (Spiegel, 1961). Together with sampling theory in general, small sample theory can provide a great deal of guidance on the size of samples needed to deal with statistical uncertainty due to sampling error. However, neither can provide much guidance on dealing with non-sampling error and, unfortunately, neither can provide much guidance when dealing with populations about which little is known and for which there are no clear or feasible sampling frames. To address these issues, at least partially, we turn to the idea of a super population.

7.4.2-3.6.Super population

The idea of a super population can be traced at least back to 1941 when W. Edwards Deming and Frederick F., Stephan (then employed by the US Bureau of the Census, and actively engaged in trying to move the Bureau into doing sample surveys in addition to decennial census counts) published a paper in the *Journal of the American Statistical Association* entitled "On the interpretation of censuses as samples" (Deming & Stephan, 1941). Basically, they argued that as a basis for scientific generalizations and decisions for action, a census is only a sample. The idea here is that when used as a basis for prediction, a census is only a sample (Deming & Stephan, 1941, p. 45). They expand on this argument in the remainder of the paper and toward the end introduce the idea of a "super population" (Deming & Stephan, 1941) in regard to what a given census represents a sample from Deming (1953) went on to elaborate on the use of a sample for analytical purposes in a paper entitled "On the distinction between enumerative an analytic surveys."

Although the concept of a super population has been refined, the definition provided by Deming and Stephen (1941: 48) is still applicable:

"….even a complete census, for scientific generalizations, describes a population that is but one of the infinity of populations that will result by chance from the same underlying social and economic cause systems. The infinity of populations may itself be thought of as a population, and might possibly be called a *super-population*. A sample inquiry is then a sample of a sample, a so-called 100 percent sample is simply a larger sample, but is still only a sample. In order to study the underlying cause systems, it is necessary to study several members of this infinity of populations…."

Not surprisingly, the idea of a superpopulation gained ground since 1941, largely due to the increased use of samples and other data for purposes of conducting analysis in order to guide decisions (make predictions). Although the concept is not without criticism (for a still useful, but somewhat dated, review see, e.g., (Graubard & Korn, 2002), it has found a home in wildlife studies and other areas of research where sampling is widely used (mainly because conducting a complete enumeration is either too costly or not possible), but without benefit of a sampling frame. Although not always explicitly stated, the concept of a superpopulation often serves as the theoretical foundation in capture/recapture (also known as dual system estimation) where samples are used to estimate the size of a finite population (Andridge & Little, 2010; J. Brown, Abbott, & Smith, 2011; Tancredi & Liseo, 2011; Wolter, 1986).

7.4.2-3.7. Rare Events, False Positives, Small Samples, and the Concept of a Superpopulation

Rare events lead to false positive outcomes, the consequences of which are exacerbated by the absence of sample frames and the use of small samples, which may violate the assumptions underlying the use of tools of statistical inference, leading, in turn, to analytic difficulties. However, understanding these factors and the analytic difficulties they can pose is a step toward clearing a path to effectively dealing with them. When coupled with the idea of a Superpopulation, the path ahead becomes clearer still.

The basic argument here is that knowing that rare events lead to false positives, we can examine rare events and provide empirical data that can be used in dealing with them. When the rare events of interest also represent "new" phenomena, we know that frames from which samples to study the phenomena are not available and that the event of interest may need to be grounded in a superpopulation. In the rest of this paper, these issues are explored within the context of the MRTPA.

7.4.2-3.8.Claim Comprehension and Intentions Study

ALCS conducted a Claim Comprehension and Intentions Study in 2017 to determine the effect of the proposed modified-risk claims on the likelihood of use of the MRTP candidate among users and nonusers of tobacco. The study involved samples. Without going into all the details here, we instead focus on one group in the study, namely those 'Tobacco users and nonusers who, after adopting the proposed product, switch to or switch back to other tobacco products that may present higher levels of individual health risk'. This group is identified in Appendix 7.4.2-1; Table 1 as: "2. Former MST User" (whose behavior is to) "initiate the MRTP" either in the presence of a product message (Base with MRTP claim) or in its absence (Base without MRTP claim). The sample without the MRTP claim had 31 subjects, while the sample involving the MRTP Claim had 28 subjects. Among the former sample, none of the 31 subjects intend to initiate the MRTP post ad exposure, while in the latter sample only one of the 28 subjects intends to initiate the MRTP post ad exposure. If we consider these two samples together, we have 1 adult respondent out of 59 switching (0.0169).

Table 1: Affirmative Responses and Percent Change by Group

Status	Intended Behavior	Base without MRTP claim N (Sample Size)	Pre-Ad	Post-Ad	Base with MRTP claim N (Sample Size)	Pre-Ad	Post-Ad	% Change	Total N
1. Never-user of tobacco	Initiate MRTP	167	5	4	168	8	6	-5%	335
2. Former MST user dipped at least 20 times	Re-initiate MRTP	31	1	0	28	2	1	0%	59
3. Current MRTP user	Switch to cigarette								
4. Adult Smokers (consisting of % ASPQ and % ASNPQ)	Switch to MRTP	209	37	33	207	29	31	21%	
-Adult Smokers Planning to Quit		211	44	34	214	41	39	23%	425
-Adult Smokers Not Planning to Quit		209	36	33	206	27	30	21%	415
5. Adult Smokers (consisting of 1% ASPQ and 8% ASNPQ)	Switch to Dual (MRTP+CIG)		50	42		37	38	24%	
-Adult Smokers Planning to Quit		211	45	33	214	35	30	18%	425
-Adult Smokers Not Planning to Quit		209	51	43	206	37	39	25%	415
6. Dual users (MST+CIG)	Switch to MRTP	403	138	131	407	144	145	6%	282
7. Dual users (MRTP+CIG)	Switch to cigarette								

When we combine the two samples the event of interest (initiating the MRTP) is “rare” in that the probability of initiating the MRTP from Former MST status is only 0.0169. That is, what if the probability of a former MST user initiating the MRTP in the population from which this sample was drawn is lower than the 0.0169 found in the sample. If this were the case, then an argument could be made that the “test” result of finding one “positive” subject among 59 subjects is, in fact, a false positive? As a means of empirically examining this question, we employ 2014-2015 of the Population Assessment of Tobacco and Health Study (PATH) adult public use files (Wave 1 and Wave 2) and do so with the idea that this large sample is a superpopulation, one from the sample of interest shown in Table 1 were drawn.

7.4.2-3.9.False Positive Test

The PATH data set shows that among 9,179,576 (n = 1,156) adult Former MST users at Wave 1who dipped 20+ times in their lifetime and currently did not use smokeless tobacco at all, 45,318 (n = 3) re-initiated smokeless tobacco with Copenhagen Snuff by Wave 2. This transition suggests that the probability of switching to the MRTP from Former MST is 0.0049, which is far lower than the probability found in the sample (0.0169). The probability of the latter in fact is approximately 3.4 times larger than the former. Using the probability from the “superpopulation” (0.0049), we would expect to find 0.2891 “switchers” in the sample of 59. That is, we would expect to find zero switchers. Therefore, there is a strong argument that the finding of one “switcher” among the 59 is a false positive.

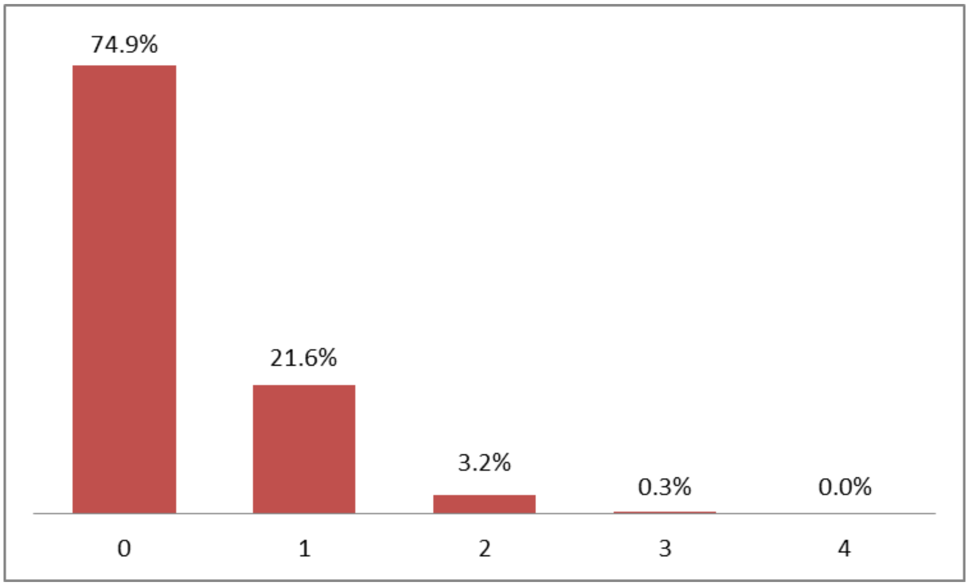
The exercise is designed to test the False Positive Paradox, whereby a rare event occurs in a given population but in a sample drawn from the population, the realized value is higher than the expected value. We can illustrate this phenomenon with data related to this report. In so doing, we stress that although the data are real, the exercise is designed to point out that this phenomenon occurs. Given this, we take an actual sample generated to support the MRTPA. In the sample, 1 person switched from FMST to MRTP, which yields p = 0.0169 (where 0.0169 = 1/59). This p is higher than would be expected from the PATH "superpopulation" where the p of switching from Former MST to tobacco (operationalized as Copenhagen Snuff) is 0.0049.

To conduct the test, we resample the dataset 10,000 times, and find that the realized p is equal to 0.00498 (where 0.00498 = 2940/10000/59) while the expected p = .0049. Since the realized p is higher (0.00498) than the expected p (.0049) found in the "superpopulation” from which the sample (n = 59) is assumed to be drawn - these results provide empirical evidence that the False Positive Paradox is distorting sample results and needs to be dealt with. Specifically, in the case of the sample in question, the single person reporting a switch from Former MST to MRTP can be taken as a false positive and, in fact, the "switch" should be recoded as "no switching."

7.4.2-3.10. Sampling With Replacement

An extension of this experiment is looking at when a sample of 59 is drawn from the superpopulation, what percent of the time would that sample contain no positive respondents, 1 respondent (as was found in the in the Claim Comprehension and Intentions Study), 2 respondents and 3 respondents.

Figure 1: Percent of the Time a Sample of 59 Respondents Would Yield 0, 1, 2 and 3 Positive Responses



These findings suggest that 74.9% of the time, the ALCS 2017 Claim Comprehension and Intentions Study would not have captured one positive respondent. 21.6% of the time, the study *would* have captured one positive respondent, 3.2% of the time the study would have captured 2 respondents and 0.3% of the time the study would have captured 3 respondents.

7.4.2-3.11. Summary

In looking at Table 1, it is clear the group we examined is not the only one that is highly likely to be subject to the false positive paradox. Given the lack of clear sample frames for them and their small sizes, it is probably useful to use the PATH data as representative of the superpopulation from which these samples have been drawn. Together, they point to a path forward that provides a means of dealing with rare events and small numbers that have the potential to distort study results, particularly if they are used to measure relative differences. One outcome of this work may be that where rare events and small samples together signal a substantial relative difference, there may be a justification for a finding of either no relative difference or that the findings are inconclusive. This could be examined on a case-by-case basis using the ideas and concepts described here along with an empirical "superpopulation" referent such as the PATH data set.

7.4.2-3.12. Literature Cited

Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *Int Stat Rev*, *78*(1), 40-64. doi:10.1111/j.1751-5823.2010.00103.x

Brown, J., Abbott, O., & Smith, P. A. (2011). Design of the 2011 census coverage surveys in England and Wales. *Journal of the Royal Statistical Society: Series A, 174*, 881-906.

Brown, S. (2011). Medical False Positives and False Negatives (Conditional Probability). Retrieved from https://brownmath.com/stat/falsepos.htm

Deming, W. E. (1953). On the distinction between enumerative and analytic surveys. *Journal of the American Statistical Association*, *44*(262), 244-255.

Deming, W. E., & Stephan, F. F. (1941). On the interpretation of censuses as samples. 36, 45–49. Durbin, J. (1967). . *Journal of the American Statistical Association*, *36*(213), 45-49.

Gelman, A., King, G., & Boscardin, J. (1998). Estimating the probability of events that have never occurred: when is your vote decisive? *Journal of the American Statistical Association*, *93*(441), 1-9. doi:doi:10.2307/2669597

Graubard, B. I., & Korn, E. L. (2002). Inference for Superpopulation Parameters Using Sample Surveys. *Statistical Science*, *17*(1), 73-96.

McMorrow, D. (2009). *Rare Events* (JSR-09-108). Retrieved from McLean, VA: https://fas.org/irp/agency/dod/jason/rare.pdf

Spiegel, M. R. (1961). *Schaum's Outline of Theory and Problems of Statistics*. New York: Schaum Publishing.

Tancredi, A., & Liseo, B. (2011). A hierarchichal Bayesian approach to record linkage and population size problems. *The Annals of Applied Statistics*, *5*(2B), 1553-1585. doi:DOI: 10.1214/10-AOAS447

Wolter, K. M. (1986). Regression models for adjusting the 1980 census: comment. *Statist. Sci.*, *1*(1), 24-28. doi:10.1214/ss/1177013808