

Online Appendix B: Institutional Review Board Issues

for “Data Issues in Developing Valid ROI/ROR Estimates”
Journal of Rehabilitation Administration (2019, forthcoming)

Steven Stern
Stony Brook University

Chris Clapp
University of Chicago

John Pepper
University of Virginia

Robert Schmidt
University of Richmond

November 2018

Institutional Review Boards (IRB) are administrative bodies set up in universities, some state agencies, and other places to protect the rights of human research subjects. With respect to the type of research associated with this chapter, the researcher should make a point of getting local IRB approval prior to engaging in any research including data collection/construction. The researcher needs to take a proactive role in designing the study and helping the IRB understand the procedure both to protect research subjects and to protect the legitimacy of the research process.

In this appendix, we present some basic IRB issues that are relevant for almost any data-based analysis of VR programs using administrative data. Note that this is an overview of common IRB considerations. It is limited especially due to the complexity in IRB rules and variation in interpretation by individual IRBs. The best rule of thumb is to be aware of the issues and discuss them with local IRB representatives.

B.1: Distinguishing between Research and Quality Control

IRB supervision applies to research. Data collection associated with organizational quality control are generally not considered research unless they satisfy some criteria causing them to be viewed as research. Research is meant to contribute to generalizable knowledge. Thus, a quality control or program evaluation study meant only for in-house use may not be considered research while a quality control report done with the intention to publish its results might be considered research.

One might be tempted to use "quality control" as a good mask to hide research behind. This is probably a foolish decision in that a) it severely limits how the results of the analysis can be used and b) satisfying IRB rules with respect to this kind of research is not that difficult and is good operating policy.

B.2: Exemption

A researcher (and possibly an IRB) might want to argue that a research project involving use of administrative VR data should be classified as exempt; i.e., that it need not pass an IRB review. Research can be classified as exempt if it involves only the use of existing data that are publicly available or are stored by the researcher so that research subjects cannot be identified. Sometimes, it is argued that replacing social security numbers with a meaningless identification code is enough to de-identify data (see, for example, Section 3.5 of the paper). However, the very nature of VR data information makes this a very weak argument. We think it is a useful and safe approach to request that a VR data IRB protocol not be considered for exempt status. It is good practice to ensure that the local IRB understands the data issues involved and that the researcher constructs a data safety plan consistent with having personally identifiable data.

Having data on people with mental illness or cognitive impairments implies that the researcher has to provide them with extra protection because, in the human subjects research sense, they are considered members of "vulnerable populations." This issue is separate from the issue of highly sensitive data discussed in Section 5.3 of the paper. The Belmont Report (US HEW, 1979) states that "first, that individuals should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protection." People with mental illness and/or cognitive impairments fall into the category of persons with diminished autonomy. This is another good reason to insist on getting more than an exempt level of review.

B.3: Personally Identifiable Information

Personally identifiable information (PII) is data where the identity of specific research subjects may be discovered. IRB rules for PII are more rigorous than for data that are not PII. Examples of PII relevant to VR research include date of birth, race, geographical information (such as zip or FIPS code), and health/disability information. Some PII can be adjusted so that it is no longer PII. For example, with probably no loss in research value, "date of birth" can be transformed into "year of birth" or "age." However, for most PII, transformation of this type is not an option. For example, race is an important explanatory variable for labor market outcomes (see Baldwin and Marcus, 2007; Dean et al., 2015, 2017, 2018a, 2018b). Geographical information, at least at the county (FIPS code) level is important to control for local labor market conditions (see Section 4.2 of the paper) and variation in lack of UI coverage (see Section 2.4 of the paper). Disability data are critical because disability type and severity have both large direct effects on labor market outcomes and the effect of service receipt on labor market outcomes (Dean et al., 2015, 2017, 2018a, 2018b).

Sometimes, multiple pieces of information in the VR administrative data could identify an individual when used in combination. For example, one might have data on a particular VR client from a rural county of a particular age, gender, and race with a somewhat uncommon health condition, and there might be only one person in that county with that set of characteristics.

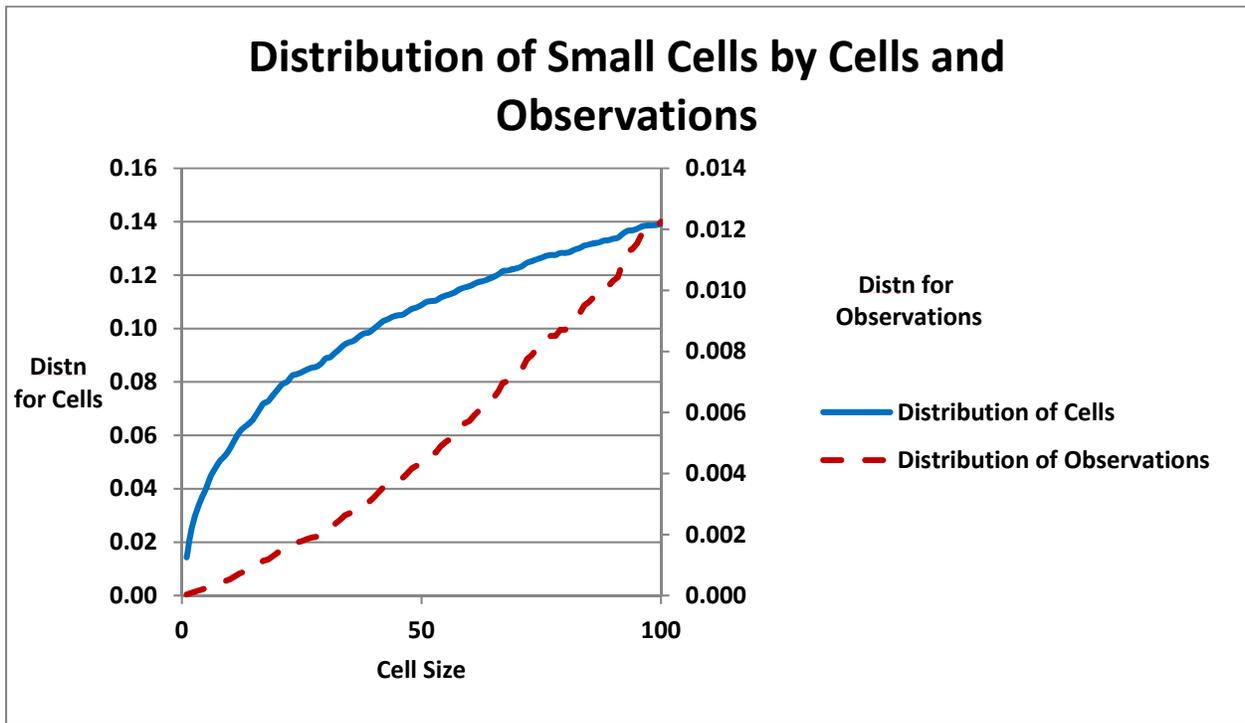


Figure 1 shows the distribution of cell sizes for (arbitrarily chosen) Florida where a cell is defined as a group of observations (in this case, from Bureau of Census, 2018) from a particular Florida county with the same age, race, and gender.¹ Two curves are presented in the figure. The solid curve, labeled "distribution of cells," shows the distribution of small cells (cell size ≤ 100) in Florida. For example, the height of the curve at "cell size" = 20 is 0.0772; this means that 7.72% of county/race/gender/age cells have size of 20 or less. This would be reduced from 20 or less to 4 or less if one were to add a disability descriptor to the mix, for example intellectual disability with population prevalence of 4.5% (Kraus, 2017).

One might reasonably argue that using "cell" as a unit of observation over-inflates the reported frequency of such events. Instead, one could use observations. The dashed curve in Figure 1 shows the distribution of observations in small cells using observations as the unit of observation. In this case, the frequency of observations with identifying information captured by cell definition variables is quite small (1.2%). But, in a data set of size 2000 (see Section 2.3 of the paper) there should be 24 observations on average from cells with cell size no more than 100.

¹ Race is limited to black or white. We limit observations to be in the age range, 18 - 64. The table in Bureau of Census (2018) is aggregated into age groups of length either 3 years or 5 years. Using a uniform distribution assumption within each Census table cell for age, we divide the reported number of observations in each Census cell by the number of integer ages included in that cell.

The probability of having no such spells is $(1-.012)^{2000}=3.27\times 10^{-11}$. Thus, there are almost surely some observations with PII.

Some of the data elements available in VR administrative data, such as information about mental illness and cognitive impairments, are considered "highly sensitive data." These are considered highly sensitive because disclosure would result in possible significant harm to the individual with respect to for example employment or health insurance eligibility. Because the data may be highly sensitive, the expectation for care and security associated with the data may be higher.

B.4: Data Security

When using VR data of the type discussed above, a researcher should have a Restricted Use Agreement (sometimes called a Data Sharing Agreement) with the owner of the data. Such an agreement is a legal contract that may place restrictions on how the data are used, by whom it can be used, and when (and how) it must be destroyed. Also, usually it specifies steps to be taken to protect the data from unwanted access and theft. These usually include specifications for the security of the machine, network, or other devices where the data are stored, from where the data can be accessed, as well as any data encryption requirements.

Such rules are usually straightforward and not difficult to follow. However, some data owners (for example, the Department of Education) require many of their data sets to be used only on computers disconnected from the rest of the world. This should be avoided if possible as it severely limits the productivity of the machine used to store the data and it sometimes limits the software and hardware available to analyze the data. Other rules that can cause some bureaucratic problems are those that require destruction of the data prior to publication of the research relying on the data. Especially in economics, where there are long lags between submission of research and publication, this is potentially problematic. However, most data owners are quite flexible about delaying the data destruction date. At the same time, it is becoming more common for journals to require access to data used in research to be published. Fortunately, most also provide an exception for data subject to restricted use agreements.

References

- Baldwin, M. & Steven Marcus (2007). Labor Market Outcomes of Persons with Mental Disorders" *Industrial Relations*, 46(3), 481-510.
- Bureau of Census (2018). Population and Demographic Data - Census Data. Retrieved from <http://edr.state.fl.us/Content/population-demographics/data/>.
- Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern S. N. (2015). The effects of vocational rehabilitation for people with cognitive impairments. *International Economic Review*, 56(2), 399-426.
- Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2017). The effects of vocational rehabilitation for people with mental illness. *Journal of Human Resources*, 52(3), 826-858.
- Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2018). The effects of vocational rehabilitation for people with physical impairments. *Journal of Human Capital*, 12(1), 1-37.
- Dean, D. H., Pepper, J. V., Schmidt, R. M., & Stern, S. N. (2019). The effects of youth transition programs on labor market outcomes. *Economics of Education Review*, 68 (February), 68-88.
- Kraus, L. (2017). *2016 Disability Statistics Annual Report*. Durham, NH: University of New Hampshire.