University of Richmond Institutional Review Board

Multisite Research: Data Safety Procedures

for NIDILRR Grant: Estimating Return on Investment on State Vocational Rehabilitation Programs

This document details safety procedures used within this project to maintain individual confidentiality while allowing researchers to perform analysis on individual-level data. Our approach works at three levels and we discuss each in turn.

1. Removal of Identifiers from Data Used by Researchers

The United States Department of Labor (DOL) defines Personal Identifiable Information (PII) as follows.

> "Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means. Further, PII is defined as information: (i) that directly identifies an individual (e.g., name, address, social security number or other identifying number or code, telephone number, email address, etc.) or (ii) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification. (These data elements may include a combination of gender, race, birth date, geographic indicator, and other descriptors). Additionally, information permitting the physical or online contacting of a specific individual is the same as personally identifiable information. This information can be maintained in either paper, electronic or other media." (*https://www.dol.gov/general/ppii*)

Similarly, the United States Department of Education (USED) defines PII as "information that can be used to distinguish or trace an individual's identity either directly or indirectly through linkages with other information" (*http://ptac.ed.gov/glossary/personally-identifiable-information-pii*). We follow different procedures for the two means of identification that the DOL and USED describe.

i. *Information that directly identifies an individual*. Prior to transmitting individual-level data to Dr. Robert Schmidt (co-PI, University of Richmond), participating agencies or their agents replace social security numbers with PINS and remove names, addresses, phone numbers, and similar such direct PII.

ii. *Information that can be used in combination to infer the identity of an individual indirectly*. Nevertheless, the researchers do require such data elements as gender, race, birth date, marital status, disabling condition, and locale (e.g., city, county, state, zip code). However, Dr. Schmidt "masks" both the birth date and locale information before transmitting data to the other researchers for their use in statistical analysis.[1] This process is described more fully in the next paragraph.

[1] This step is being implemented for the data collected under the NIDILRR Grant; "Estimating Return on Investment on State Vocational Rehabilitation Programs" funded from October 1, 2014 – September 30, 2019. As of this date, we have not applied this step retrospectively to data collected under earlier grants.

Following the procedure described in the next section, participating agencies or their agents securely transmit data files to Dr. Schmidt. Using birth date and date of application for VR services, Dr. Schmidt calculates the individual's age in quarters as of the application date. Although age is an important control variable within the model, we do not need to retain precise birth dates. Similarly, locale identifiers are not used within the model. Rather, they enable us to include important control variables that influence the likelihood of employment as well as earnings level if employed. Thus, Dr. Schmidt matches the locale identifiers with federal databases to determine three items available at the county level.

- Quarterly employment rates (an indicator of the local economic environment in which an individual seeks employment).
- Proportion employed out-of-state. This is useful because Unemployment Insurance (UI) data does not include earnings in other states.
- Proportion employed by the federal government, This is useful because UI data does not include earnings by federal employees.

Dr. Schmidt then replaces birth date, city, county, and zip code with these variables before sharing the data files with the researchers. Thus, Dr. Schmidt is the only researcher who retains access to birth date and locale identifiers in case they are needed again to incorporate updated data from federal databases. However, he does not store them on any computer or hard drive used in analysis. Rather, he stores the original files on an external hard drive that requires a strong password for access and is protected by Windows 10 Bitlocker (XTS-AES 128-bit device encryption).

We believe that elimination of birth date, city, county, and zip code significantly reduces the ability to infer individual identities indirectly in the unlikely event of a breach by one of the researchers.

Procedures for Securely Transmitting Data Files

Dr. Robert Schmidt initially receives data sets without direct PII from either the partner agency or its agent through secure means. Passwords for access to these files are always transmitted by phone and never by email. Here are examples of how files have been transmitted to Dr. Schmidt.

- Agency provides Dr. Schmidt with limited and secure access to a directory on an agency file server.
- Agency uploads the files to a limited-access and secure directory created by Dr. Schmidt in his Box account provided by the University of Richmond. Dr. Schmidt deletes these files from his Box account after he verifies their validity.
- 32-bit encrypted Zip files that meet Social Security Administration security standards. In the case of Virginia, Dr. Schmidt has used an encrypted thumb drive to pick them from the agency.

Robert Schmidt masks birth date and locale information as described above before transmitting files sets to other researchers via 32-bit encrypted Zip files that meet Social Security Administration security standards.

Storage Procedures for Individuals with Access to the Data

1. Dr. Robert M. Schmidt (PI, University of Richmond) stores these data on three mobile devices (a laptop and two external backup hard drives) as well as a more powerful workstation. All devices are encrypted using BitLocker and require strong passwords for access. Dr. Schmidt uses the external hard drives solely to back up the laptop and keeps the two hard drives in separate locations. Dr. Schmidt does not store or backup this information on any remote site. The University of Richmond uses SSL as its data security protocol.

2. [**NOTE:** Comparable information should be included for any other researchers who access these data.]