CR-53209.2

Data Use and Non-Disclosure Agreement

Jail Component of "Overview of the Criminal Legal System in Michigan: Adults & Youth – Update"

THIS DATA USE AND NON-DISCLOSURE AGREEMENT ("Agreement" or "Participation Agreement") is made and entered into by and between **Wayne State University**, a public body, Sponsored Program Administration, located at 5057 Woodward Ave., 13th floor, Detroit, MI 48202 ("WSU") and **County of Washtenaw, through its Sheriff's Office, Correctional Division,** a political subdivision of the State of Michigan, located at 2201 Hogback Road, Ann Arbor, MI 48105 (hereinafter referred to as "WCSO").

Section 1: Recitals

1. The Wayne State University's Center for Behavioral Health and Justice has received funding for two new evaluations regarding individuals in jail. First, is a project funded by the Public Welfare Foundation and Michigan Justice Fund has been awarded to Wayne State University's Center for Behavioral Health and Justice (CBHJ) to update its "Overview of the Criminal Legal System in Michigan: Adults & Youth" report to support the work of the Michigan Talk Force on Jails and Pretrial Incarceration (MTFJPI). There are three components to this data-informed update, one of which involves the collection and analysis of as many as 20 counties' booking data in an effort to understand whether, and how, jail and pretrial incarceration legislative reforms enacted in March, April and October of 2021 have impacted Michigan jail populations, and whether any impacts vary across geographic regions. Second, the National Institute for Mental Health (NIMH) has funded Wayne State University, Henry Ford Health System, and Michigan State University to conduct a study that will evaluate the reliability of an algorithm applied to health record data in assessing risk for suicide in jail populations.

2. Counties participating in both components are willing to furnish booking data from March 1st, 2019 through September 30th, 2022 that will be used to examine changes in the composition of jail populations in the two years before and one year after the 2021 legislation was enacted.

3. The WSU Evaluation Team, led by Drs. Sheryl Kubiak and Matt Larson (Principal Investigators) will collect and analyze booking data, from up to 20 counties, required for the legislative reform component of the updated overview of the criminal legal system in Michigan. WCSO will collaborate with WSU to provide the booking data elements necessary to evaluate the impact of the legislative reforms on their jail population.

Now therefore the WCSO and WSU agree to enter into an Agreement under the terms specified in this Agreement and any other applicable public body, research, and confidentiality laws, including without limitation the Health Insurance Portability and Accountability Act of 1996, Public Law 104-191 (HIPAA).

Section 2: Definitions

1. "Data" for purposes of this Agreement is defined as information collected by the County from inmates booked into the WCSO from March 1st, 2019, through September 30th, 2022.

2. "HIPAA Regulations" means the regulations promulgated under HIPAA by the United States Department of Health and Human Services, including but not limited to 45 CFR Part 160, 45 CFR Part 164, and 42 CFR Part 2.

Section 3: Use of Data including Personal Identification Indicators (PII)

1. In accordance with this Agreement, the County agrees to provide Data to WSU to conduct an evaluation of: the impact of legislative reforms on the composition of its jail population and the algorithm for suicide identification.

2. WSU agrees to use the Data provided by the County in compliance with applicable research and confidentiality standards, and in accordance with the CBHJ's – Data Protocols, which is incorporated by referenced herein and attached to this Agreement as 'Exhibit A.'

3. Data from WCSO will be used by WSU to assess: 1) whether, and how, legislative reforms in 2021 affected jail populations by analyzing booking data two years before reforms were enacted to one year after and 2) if the algorithm is an appropriate method for suicide identification.

Section 4: Specific Data Elements

1. Data are requested for conducting an evaluation of jail populations before and after the enactment of legislative reforms (i.e., PA 382, PA 393, PA 395). The evaluation relies on information contained within booking data to provide an evaluation of the impact of state-level legislative reforms that were enacted in March, April, and October of 2021.

2. PII provided by WCSO will include name, date of birth, race, ethnicity, sex, veteran/military history, suicide risk, jail identification number, state identification number, booking identification number, booking date and time, offense date, arrest date, booking reason, disposition status, arresting charge (all charges, not just most serious), sentenced offense (all charges, not just most serious), offense type (i.e., misdemeanor or felony), length of sentence (if applicable), case or docket number, court code, pretrial release decision date, release date and time, release reason, bond type (if applicable), cash bound amount, cash bond posted.

3. WSU agrees to protect the confidentiality of all PII transferred to WSU by WCSO and to use the PII only for the specific purpose of conducting an evaluation of the data in accordance with the terms of this Agreement.

Section 5: Confidentiality of PII

1. Data transfer of files containing PII will occur through secure file transfer protocol (SFTP) to WSU's dedicated server. All Data will be securely stored on a password-protected and encrypted server behind the WSU firewall and in accordance with the CBHJ's Data Management Protocols. WCSO shall not be responsible for securing the confidentiality of any PII transferred by WCSO to WSU through the SFTP.

2. Unless otherwise required by law or legal process, WSU agrees to keep all Data collected from WCSO confidential. Only a limited number of specifically trained research staff will have access to identified Data for the purpose of data entry and linking Data to other

administrative data. Files will be stored in a password protected repository on an encrypted computer hard drive. Even if computers are lost, the Data will not be accessible. Once data linkage and project purpose has been achieved, all identifying information will be removed from the files and each individual will be assigned a 'case number' that is not linked to any identifiable information. For project #2, de-identified Data will be shared through SFTP with our research partners at Michigan State University and Henry Ford Health Systems.

3. WCSO shall not be responsible in any way for the security of Data within WSU's possession and control.

Section 6: De-identification of PII in Reports

1. WSU agrees to aggregate data in all of its written reports and other forms of public and academic dissemination; no individual will be identified during dissemination.

Section 7: Compliance with state and local confidentiality standards

1. To protect the individuals' confidentiality, and the integrity of the study, the WSU evaluation team has engaged in, or examined, several processes at the university, state, and federal level.

   a. __University Level.__ There are two studies connected to this data request. The study evaluating outcomes of 2021 legislative reforms has been reviewed and approved as ethical by WSU's IRB, but is deemed 'exempt' from IRB oversite. This determination is based upon the purpose of our information gathering as specific to the state and county's interest in improving programs – and not for 'generalizable knowledge production'. The 'generalizable knowledge production' (i.e., medication studies, intervention studies) has a broader – and potentially risker – implication for the individual. Individuals involved in those studies are often exposed to situations that they would not normally be involved in (i.e., taking new medications or potentially being in a placebo study). This is in direct contrast to evaluation in which assessment of individuals, or the services they engage in, are outside the control of the research team. For example, the evaluation team has no decision-making ability in terms of someone's confinement status or whether they are treated by CCSO. The WSU IRB HPR number for this study is IRB-2022-181.

   The second study connected to this data request, which will examine the reliability of an algorithm applied to health record data in assessing risk for suicide in jail populations has been approved by the IRB. The IRB number for this project is IRB-22-06-4717-B3

   b. __State Level.__ Administrators at MDHHS have reviewed WSU's Data Management Protocol and have found that the research protocol meets the most stringent standards for evaluation – even under the 42 CFR Part 2, Federal confidentiality standards (see below) that protect substance abuse treatment programs.

Section 8: Limitations of Disclosure

1. WSU agrees that access to Data will be provided only to the Principal Investigators, Drs. Kubiak and Larson and WSU evaluation team members. If records containing identifying

information will be accessed by anyone in addition to the aforementioned parties, federal 42 CFR Section 2.53 rules must be followed. Confidentiality of inmate identifying information will be maintained, and all inmate identifying information will be destroyed upon completion of the evaluation.

Section 9: Retention & Destruction of Data

1. WSU agrees that the Principal Investigators will destroy all confidential information associated with the actual records as soon as the purposes of the project have been accomplished. Once the project is complete, the PIs will: 1) destroy all hard copies containing confidential Data (e.g., shredding), 2) archive and store electronic Data containing confidential information offline in a secure place, and delete all electronic confidential Data, and 3) all other Data will be erased or maintained in a secured area. Written records which are subject to these regulations must be maintained in a secure room, locked file cabinet, safe or other similar container when not in use. Notwithstanding anything to the contrary, WSU may retain one (1) copy of the Data to the extent necessary to comply with the records retention requirements under any law, and for the purposes of research integrity and verification.

Section 10: Term and Termination

1. The term of this Agreement shall be January 1, 2023 – September 30, 2025. This Agreement may be terminated without cause by either party upon thirty (30) days written notice to the other party.

2. Should a party commit any breach or default under this Agreement, and should such breach or default not be corrected within thirty (30) days after receipt by the party of written notice from the non-breaching party specifying the breach or default, this Agreement may be terminated without further notice by the non-breaching party.

Section 11: Indemnification

1. To the extent allowed by law and except to the extent arising from the gross negligence or willful misconduct of WCSO, WSU shall indemnify and defend WCSO, its officials, officers, agents, employees, and assigns, from and against all loss, damage or injury, and reasonable costs and expenses arising out of the acts or omissions of WSU in connection with the representations, duties and obligations of WSU under this Agreement. The parties' respective rights and obligations under this Section shall survive the termination of this Agreement.

2. To the extent permitted by Michigan law, WCSO shall indemnify and defend WSU, its officials, officers, agents, employees, and assigns, from and against all loss, damage or injury, and reasonable costs and expenses arising out of the acts or omissions of WCSO in connection with the representations, duties and obligations of WCSO under this Agreement. The parties' respective rights and obligations under this Section shall survive the termination of this Agreement.

Section 12: Governmental Immunity

1. WCSO does not waive any immunity by entering into this Agreement, and fully retains all immunities and defenses provided by law with respect to any action based upon or occurring because of this Agreement.

Section 13: Notice

1. All notices, demands or other writings permitted or required by the terms of this Agreement shall be deemed to have been fully given, made or sent when made in writing and deposited in the United States Mail, registered and postage prepaid, and addressed to the Contract Administrators as follows:

WCSO: Commander Kurt Schiappacasse
Washtenaw County Sheriff's Office – Correctional Division
2201 Hogback Road
Ann Arbor, MI 48105
Email: Sschiappacassek@washtenaw.org

WSU: Sheryl Kubiak, PhD
Principal Investigator
Dean, School of Social Work
Wayne State University
5447 Woodward Avenue
Detroit, Michigan 48202
Phone: 313-577-2240
Email: spk@wayne.edu

Matthew Larson, PhD
Principal Investigator
Director of Implementation, CBHJ
Associate Professor, School of Social Work
Wayne State University
5447 Woodward Avenue
Detroit, Michigan 48202
Phone: 313-577-4014
Email: mattjlarson@wayne.edu

With a copy to:
Wayne State University
Sponsored Program Administration
5057 Woodward Avenue, 13th floor
Detroit, Michigan 48202

The address to which any notice, demand or other writing may be given or sent to any party may be changed by written notice given to the other party.

Section 14: Entire Agreement

1. This Agreement together with any affixed schedules, exhibits or addenda referred to herein, shall constitute the entire agreement between the parties. Any prior understanding,

representation or negotiation of any kind preceding the date of this Agreement shall not be binding upon either party except to the extent incorporated in this Agreement.

Section 15: Modification

1. Any modification of this Agreement or additional obligation assumed by either party in connection with this Agreement shall be binding only if evidenced in a writing signed by each party or its authorized representative.

Section 16: Assignment

1. The rights and obligations of each party under this Agreement are personal to that party and may not be assigned or transferred to any other person, firm, corporation or other entity without the prior written consent of the other party. In the event of a proper assignment, this Agreement shall be binding upon and inure to the benefit of the parties' successors and assigns.

Section 17: Consent to Personal Jurisdiction

1. WSU acknowledges that this Agreement shall be deemed to have been executed in the State of Michigan, and hereby consents to the exercise of general personal jurisdiction over it by the appropriate courts in the State of Michigan.

Section 18: Attorney Review

1. The parties represent that they have carefully read this Agreement and have had the opportunity to review it with an attorney. The parties affirmatively state that they understand the contents of this Agreement and sign it as their free act and deed.

In witness whereof, each party to this Agreement has caused it to be executed on the dates indicated below.

Wayne State University:

By: Patty Yul Kieleszewski
Patty Yuhas Kieleszewski
Associate Director, Contract Admin, SPA

Date: 04/06/2023

County of Washtenaw:

By: Gregory Dill 05/19/2023
Gregory Dill
County Administrator

Date:

Read and understood by Principal Investigator:

By: Sheryl Kubiak
Sheryl Kubiak, Ph.D.
Dean, School of Social Work

By: Matthew Larson
Matthew Larson, Ph.D.
Associate Professor, School of Social Work

Date:_________________________________

Date: 03/12/23

Attested to:

By: Lawrence Kestenbaum 05/22/2023

Lawrence Kestenbaum
County Clerk/Register

Date: 05/22/2023

Approved as to Content:

By:_________________________________

Jerry Clayton
County Sheriff

Date: 3.27.23

Approved as to Form:

By: Michael L. Ritter 05/17/2023
Office of Corporation Counsel

Michelle K Billard
Office of Corporation Council

Date:_________________________________

EXHIBIT A

CBHJ Evaluation- Data Protocols

'Data' definition- any information or observations that are associated with projects within the Center for Behavioral Health and Justice (CBHJ), including: administrative data, interviews, surveys, phone call transcripts, email communication, and any other related product.

Data Ownership

The data is owned by Wayne State University and the steward of the data is the lead Principal Investigator (PI) on each project. The PI controls the course and publication of the data and is subject to institutional review as well as review and feedback from the collaborating agencies. Any use of the data external to the CBHJ or for publication and dissemination (i.e., presentation) must be approved by the PI.

Research Ethics Training

All CBHJ staff, collaborators, students, or volunteers must successfully complete research ethics training through Wayne State University, or a university associated with the Collaborative Institutional Training Initiative (CITI).

Directions for completing the required basic training courses for every person associated with the CBHJ:

- Go to this website http://www.citiprogram.org.
- Register or login
  - MAKE SURE you select Wayne State University and not Wayne College when registering.
- Four modules must be completed:
  - Human Subjects Research: Basic Course in Human Subjects Research Curriculum Select Social/Behavioral Researchers
  - Responsible Conduct of Research Select Social/Behavioral Researchers
  - CITI Health Information Privacy and Security- HIPS Select Students & Instructors
  - Internet Research SBR
- Each module will consist of an article webpage that must be read and there are associated multiple choice questions to answer at the end of the module. Take notes as you read the modules. The passing score is usually 75-80% and you must score this comprehensively to pass the modules. If you do not pass a module, you can take it again.
- Once you have successfully completed the modules you will be able to print off a sheet saying you passed the modules along with the date and score(s) you received- please keep for your records.
- When you have completed all the modules, please email Heidi Bisson (heidi.bisson@wayne.edu) a copy of your completion certificate that has the expiration date.
- It is incumbent upon everyone at the CBHJ to maintain continuity of ethics certification. CITI will send email reminders a few months prior to the basic course expiration. Please be sure to update CBHJ when new certificates are received.

Additional trainings may be required, dependent upon the needs of the project.

Institutional Review

It is the responsibility of the PI and/or Data Director to maintain Institutional Review Board (IRB) approval for ongoing and new projects in the CBHJ. Any project that includes identifiable data must be reviewed by WSU's IRB.

Staff members who will be accessing data must be included or subsequently added to the appropriate IRB protocol(s) as key personnel. Access to data will only be provided after approval of key personnel by the IRB.

Data Use Agreements (DUA)/Memorandum of Understanding (MOU)

Stakeholders/funders may require data use agreements. Team members work with the Coordinator for Research in the School of Social Work (neva.nahan@wayne.edu) to create these documents that must be reviewed through Wayne State University's legal team and the stakeholders/funders that require this documentation. Once a draft is received by the CBHJ, it can take one to three months for legal team review. Once a signed version of the DUA is ready for execution, PI and team members are held to the requirements of the agreement. Team members should periodically review DUAs and renew when they expire (if applicable).

Data Team Documents

All data sources will have eight associated files. These include the: 1) raw data file, 2) cleaned data file, 3) the codebook, 4) the data file template, 5) the logbook, 6) the 'Read Me' document, 7) cleaning syntax file, and 8) analysis syntax file. (See Appendices for examples).

Raw Data: The raw data file contains the data exactly as it was given to / collected by / input by the CBJH team. This file is uncleaned to preserve the original information. This can be useful in case data are cleaned incorrectly, or if we need to compare information in the "cleaned" data file to the original information we received. The raw data file should never be manipulated or edited. This file is typically either an excel file or an SPSS file (occasionally, both). For data collected and analyzed on an ongoing basis (often in support of continuous quality improvement in program evaluation), data manipulation may occur in the raw data file prior to the data cleaning phase. Changes must be documented in the logbook and on the original data collection form.

Manipulated Data: The manipulated data file is the "living" data used for analyses. These data have been fully cleaned in accordance with standards and rules set out by the data team. This file should contain all value labels, variable names, and missing information. Further, if major edits are made to this file, the person editing the file renames it after manipulating to change the initials / date. For example, "datafile_VLN_5-20-2019." This allows us to track who created the most recent changes. If you are making changes that may result in the LOSS OF DATA, please copy/paste the file, and create a separate file. This way, if data are inadvertently lost, the primary data file / master data file everyone is working from will not be edited. Previous versions should be saved in an 'Archive' folder created inside the data folder.

Codebook: The code book will contain all information about the data after it has been fully cleaned / manipulated. The codebook will include the variable name, variable label, value labels, and missing values for all variables within a data file. This document is created at the beginning of data collection (prior to requesting data from stakeholders) and minor edits can be made as the project advances.

Data Template: The data template is housed in the same excel workbook as the codebook. The template will simply include all variable names in the order they are found in the manipulated data file.

Logbook: Logbooks are used to track data decisions as data are cleaned, manipulated, and analyzed. Logbooks identify difficulty / problematic cases, as well as how these cases were handled, who decided on how to handle these cases, and the dates that these problems were both identified and resolved. *We encourage team members to use logbooks judiciously!* These files (along with the Read Me files) are the "breadcrumbs" that allow us to follow the changes that were made from the raw data file to create the manipulated data file.

Read Me: The Read Me file contains extensive information about each dataset. Information may include where data came from, what format the data came in, who we requested data from, dates data were requested from, steps to clean the data, coding schema, primary variables required to clean data, and primary variables for analysis. This file will also include the name(s) of those who have cleaned the data previously to facilitate questions.

Cleaning Syntax File: the cleaning syntax file will be a step-by-step process on how data in their raw form were manipulated to create the manipulated data file. This syntax, when run on the original raw file, should create the manipulated file through SPSS commands. *NOTE: Because some manipulated / data cleaning must be done by hand, it may not always be possible to simply RUN the file and be done. However, syntax files will all make a note when a particular variable is created by hand or has been hand manipulated.

Analysis Syntax File: The analysis syntax file includes all analyses that are found in the table / report / document for which they are associated. It is possible to have multiple analysis syntax files for one data set, depending on the types of products that are being created. Analysis files are "living" files, in that they may change as the team adds, removes, or changes the analyses for the product. Running this file should give you the statistical output that is required for a particular product.

Data Collection

Diligent and systematic record keeping is essential to ensuring the integrity of research data. Team members should document when and how data was requested and received. This information should be kept in a "Read Me" file (read below) for each data source. Read Me files include specific details about the data that would be useful for other team members to know about the data: date requested, date received, variables requested, how specific variables in this data may differ from other sites in the study, etc. The Read Me file is most useful when training new team members- so they can acclimate themselves to the data/projects.

Data is received in many forms from stakeholders: pdf files, paper copies that are mailed, paper copies that are picked up at research sites. All hard copy data should be locked in a cabinet in a locked office as soon as the data is received by the team.

Electronic files have also been received through encrypted emails and by sharing a folder on our cloud system. Data received by email should be immediately saved to the cloud and the email should be deleted. Keep passwords in the "read me" me file. When a stakeholder wants to send data using the cloud system, the team member must create a new folder in the largest 'tree' of dropbox, give permissions to see the folder to the stakeholder and those team members that need to see the data. Once all data is received, the data should be transferred to data team folders and the shared folder should be deleted.

Data Storage & Transfer

Hard Copy Data

Any hard copy data that comes in print outs or in handwritten notes must be:

2. Logged into a logbook upon retrieval from each site
3. Stored in a locked cabinet when not in use

Logbook entries should note the date and time when hard copy data is transitioned from one team member to another. Once hard copy data has been entered into an electronic format and verified by another team member, hard copies must remain in a locked cabinet in a locked office until IRB requires destruction of the hard copies (within 1 to 5 years after the project is completed).

Electronic Data

All electronic data, and hard copy data transferred to electronic format, must be stored on computers that are password protected. Project computers should have anti-virus protection and use intrusion detection software. Team members should do everything in their power to keep the computer safe from theft or damage. All computers used at the CBHJ should only use networks that ensure no outside wireless devices will have access to the files.

Original and finalized data will be stored on a cloud system (i.e., dropbox). It is important that team members upload data often, to back-up the data. All team members will have controlled access to the cloud system where data is stored, limiting the access to identifiable data to only those team members that need access. Individual team members will have the ability to upload and download data that is pertinent to their role. Accessing and staying connected to the cloud service is imperative.

Transferring Data

Email: In instances where identifiable data is emailed with stakeholders or amongst team members the files must be encrypted, and password protected. Passwords should be emailed separately to the email receiver.

Cloud/Dropbox: When sharing a Dropbox folder with a stakeholder or consultant external to the team that will house individual identifying information, team members should:

4. Protections on the folder should be set so that no link can be created for the folder. This link could then be shared, leaving the identifiable data in the folder open to no secure sources.
5. Share the folder with only those stakeholders that have access to this same identifiable data in their position of employment.
6. The Data Director and Deputy Director must also be added to the folder, along with any other pertinent team members.
7. Once all data has been transferred to this folder, the files should be moved into a permanent folder and this externally shared folder must be deleted.

Version Control

Because team members are working on project documents from different locations, often (nearly) simultaneously, ensuring version control is imperative. When a new file is created:
The original team member should give the file an appropriate and descriptive file name that includes version, date, and initials. (Example: K6Assessments2017Cohort_KentCounty_V1_3-12-17_EC.).
When a second person makes **MINOR** changes, the file should be **RENAMED** as version two, with initials and dates updated (example: K6Assessments2017Cohort_KentCounty_V2_3-30-17_VLN.). This process should be replicated as each person creates new versions / as new individuals edit the file.

1. Typically, we request that individuals NOT save their changes as a separate file, as this creates many versions of the same information. The *exception* to this rule is if SUBSTANTIAL changes are being made, and information from the previous version that is being edited or replaced may be needed / references in a later version. This is a gray process that requires individual discretion; however, we ask individuals to do this sparingly.
2. Before downloading / editing a file from the cloud, verify that no one else is working on the file. If two people are working on the same file at the same time, changes may be lost, or multiple versions with different information created. The cloud service will tell you if someone else is currently using the data, and who that person is. If you must make immediate changes, please contact the person in the data file to make arrangements for them to save their edits and exit the file.
3. If multiple versions of a file are created, create an "Archive" folder within the data folder. As new versions are created, move old versions to the "Archive" folder to keep information organized and clean, and to reduce the chances of any one person using an old and outdated version of the file.

Data Cleaning & Verification

As team members are cleaning data, we ask that they keep a record of data decisions they make / questions they have about data in the logbook. Logbooks should include enough information that a person reviewing the logbook is able to go into the file, find the issue / problematic case, and either identify the resolution or recommend a resolution. All logbook entries should include:

1. The person's initial who is logging the issue
2. The date it is being logged
3. The unique identifier of the case/variable where there is an issue
4. A description of the problem
5. The initials and date of the individual who is resolving the issue (only an Analyst or Data Director can make resolutions)
6. A description of the resolution/decision.

When you make an entry in the logbook, you should notify the Data Director/Data Analyst that there are items for their review. They will review the entry and will decide on how to proceed. Once the Director/Analyst responds in the logbook, they will email you to let you know that the issue has been resolved/what course of action is appropriate.

Logbooks have three purposes. First, they allow us to communicate with one another and to collaboratively make data decisions that make sense to the project. Second, they help us stay consistent across sites/projects/past years. Third, they serve as a reference going forward. All logbooks from previous years were saved, and include information on the type of problem, and the resolution for the problem. As such, if you encounter an issue as you are entering/cleaning data, we ask that you find the logbook for that specific data source that were used in previous years, and to scan through it to see if there have been similar issues. Even if the issue was found in previous years, still log the issue in the current year logbook and not that you found your resolution from past years logbook. Data personnel may not remember exactly which decisions were made in the past, and this allows us to make decisions that are consistent with prior decisions.

In instances where the team has worked with the same data source in the past, it would also be useful to review the 'data description' documentation written by other team members who have worked with the data (i.e. jails, MDOC OTIS, SCAO, Specialty Court Database, Medicaid billing, etc.). There may have been issues with entry/coding in the past that could explain why there may be discrepancies in the data. Once the original data have been entered into an electronic format and meet the criteria within the codebook, a different team member will double check the accuracy of the final database against the original files, by auditing 10% of each county's participants. Below are the steps of the verification process:

1. Identify cases for the audit. These should be randomly selected; this random sample should include 10% of all the cases in the entire file.
2. The case ID for each of these cases should be entered into the logbook, each on their own line.
3. Make a note for each of the cases about any potential discrepancies. If there are no discrepancies, leave the rest of the row for that case blank and write 'no discrepancies. If there are discrepancies between the hard copy and the electronic copy/the raw data and the manipulated data, identify the variable(s) that do not match, what the manipulated data included, and what the raw data included.
4. *If more than five discrepancies exist, the entire file needs to be verified.*

Once individual files have been entered and verified, data files are often merged. This process should only be done by the Data Director or a Data Analyst. Syntax files for "Cleaning" must be kept as a separate record for each project. The syntax for merging is documented at the beginning of the "Cleaning" syntax file. Once merging is complete, another team member must verify that the merge was done accurately by selecting the single source files for 10 cases and verifying variable matching.

Verification of Identifiers

Projects that require follow-up and linkage to other administrative data sources requires verification of identifiers throughout the data collection process. For example, if identifiers are first collected on screening instruments from individual participants, and these identifiers conflict with what is recorded in booking reports (received later), the most valid form of data is the booking report. In this case, the original file with the screening instruments would stay the same as was entered and discrepancies get recorded in the file's logbook. The master file must include the most valid identifiers and would be merged into the master file from the booking report or other more valid sources for identifiers. However, it is the position of the CBHJ that the only self-report measure that is more valid than another source (i.e., booking report) is race/ethnicity. Due to constraints in data collection in the criminal/legal system around race/ethnicity, the individual's self-identification takes precedent.

Syntax

All syntax files should include a title, who they were created by, when they were created, who they were last edited by, when they were last edited, a short description of the purpose of the syntax file, the data file and/or template product file(s) associated with the syntax, and a log of changes over time. Every time you make a substantial change to a syntax file, in addition to update the "edited by" and the "date edited on," you should also include the date and what you are changing in the syntax file.

There are two primary types of syntax files that must be kept: (1) the cleaning syntax and (2) the analysis syntax. It is possible that a particular data set may have multiple analysis syntax files if there are multiple products (for example: if a county is getting both a PPT report and tables, there will likely be two analysis files: one that produces each of these products).

The cleaning file should be a step-by-step process that shows how the raw data were manipulated and transformed to create the manipulated data. A person with a reasonable amount of SPSS experience should be able to open a syntax file and follow the processes taken to clean the data. Please include all code for variable renames, creating value labels, creating dummy variables, variable creation, auto recodes, assigning missing's, etc. If you are still learning SPSS syntax, it can be useful to use the "point-click" option, and then click the "paste" function. This will paste the syntax into a syntax file. This is useful for several reasons: (1) it will give correct syntax (2) you can copy / paste the syntax and edit the syntax as needed to cut down on the time it takes to write the syntax and (3) it will get you used to reading / writing SPSS syntax. Below each command, use the SPSS syntax note code to explain what each chunk of syntax accomplishes, how the resulting variables should be used, etc. Example: the following code shows a created variable, and how to use (*) to create notes. Anything that follows (*) on a line in a syntax file will not run when running the file.

```
COMPUTE K6Answered=Q1aD+Q1bD+Q1cD+Q1dD+Q1eD+Q1fD.
EXECUTE.
        *This creates a count variable of how many questions on the K6 screen an individual answers.
        Values range from 0-6.
        *All of these with 6 are good to go.
        *Any of these with 4 or 5 need to have the variable IMPUTED MEAN.
        *Any of these with 3 or fewer will be DROPPED.
```

In addition to basic cleaning, sometimes files must be restructured (long to flat, for example) or merged with other files to be usable. First, if you restructure / merge data, DO NOT OVERWRITE EXISTING FILES. Save your "working" file(s) with one name, and after you complete the restructure / merge, write a save-as code in the syntax to save the newly created file SEPARATELY. Next, it is *imperative* that this code is also included (and noted) in the syntax file. Given the complexity of data restructuring and data merging, it is not uncommon for this to be done incorrectly. Having the syntax that created the flattened or merged file can identify the issues with the restructure / merge and having the files from JUST PRIOR to the restructure / merge allows you to easily re-do these tasks without having to start over from the raw data.

Next, analysis files should include all numbers presented in a particular product. There are several components of code that are vitally important for analysis. First, make sure that if you are using any filters that this code is written into the analysis. We often switch between different subsamples (for example: those identified with SMI only, or minority individuals only). Including the filters in the syntax ensures that all people use *identical* filters to run an analysis. Include a note below the filter syntax to explain what sample you are selecting and, if necessary, why you chose that sample. Next, analysis files should contain all the syntax required to get every number presented in a report / in a table. When possible, these should be to the product. Including notes that (1) explain what the output will show ("This is a crosstab comparing the rate of SMI by race") and (2) where this information can be located ("This is on slide 12 of the template / XXX County report.").

Analysis files are also frequently edited / updated. As previously stated, please log all changes made to an analysis syntax file. However, in addition, if you are changing syntax (for example, changing a filter, changing an analysis, etc.), removing syntax (for example: we were originally including a comparison between SMI and non-SMI, but decided not to use it), or replacing syntax (example: we originally percentages a crosstab on gender, but now would like to percentage the other way on SMI), please do the following: (1) Comment out the ORIGINAL SYNTAX using (*) at the beginning of each line of syntax for that particular analysis; (2) below the syntax, write a note that explains what is being changed / why it is being changed that INCLUDES THE DATE ("5/20/2019: We are removing this syntax because we are no longer including this analysis" or "we have chosen to use a different filter because…"); and (3) insert the new syntax below the old syntax. It is not uncommon as we go through versions / iterations of products for us to run one analysis, remove it, and decide in a later version that we do want to include it. Often, this happens after considerable time has passed since we wrote the syntax file, and it can be difficult to remember filters, the exact variables we used, etc. This allows us to simply go back, remove the (*) from the syntax, and re-run the code to get the exact values we had previously.

Finally, whereas cleaning files are connected to the data file they clean, please note that analysis files are connected to the *product they create*. As such, a particular data file may have multiple analysis files (for example: county XXX K6 file may have both a report syntax file, and a tables syntax file), or it may not have an analysis file at all (for example: in the JDW data, none of the tables are analyzed independently of one another; rather, the analyses are all completed on the final merged product, and the analysis file corresponds to the completed merged data file).

File Deidentification

After all verification processes have been cleared, unique identifiers should be created for everyone (e.g., "WSUID" for Jail Diversion data). These identifiers have codes, so be sure to discuss how to create them if you have questions. For example, Wayne County identifiers will ALL begin with "82," regardless of the data file / project. This way, in all merged data, we can easily tell which cases are from Wayne County (all begin with an 82) as compared to Barry (all begin with 08). Second, projects have certain ranges for their data. The 2017 K6 data collection cohort, for example, all have identifiers that are less than 500 (example: 82331 could be an identifier from the 2017 K6 data collection cohort from Wayne Count), whereas the 2019 K6 data collection cohort all have identifiers that are greater than 500 (example: 08732 could be an identifier for the 2019 K6 data collection from Barry County).

Once all cases / individuals have a unique WSUID identifier, an "identification key" should be created. This is an excel document where the columns represent a variable that is an identifiable piece of information, and where rows are individual cases. Columns should include all identifiable information, as well as the WSUID. For example, an "identification key" will likely contain the following columns: WSUID, First Name, Last Name, Date of Birth, Booking Number, Inmate Number, SSN, Race, Gender.

Once all cases have an identifier, and after the "identification key" has been created, all identifying information should be removed from the data file (this does not include the race/gender information, although these variables should be included in the identification key). This creates a de-identified data set that can be shared more widely than those data sets that are still identified.

Data Retention

The PI and Data Director will destroy all confidential information associated with the actual records as soon as the purposes of the project have been accomplished, or the date set by the funder/stakeholders. Once the project is complete, the data director will: 1) destroy all hard copies containing confidential data (e.g., shredding), 2) archive and store electronic data containing confidential information offline in a secure place and delete all on-line electronic confidential data, and 3) all other data will be erased or maintained in a secured area.

Files Retained for Data Repository

Eight documents (seven files) are kept for each data source and each merged data file (all have been discussed above):

1. Raw Data File
2. Cleaned Data File
3. Codebook
4. Data File Template (a separate sheet inside the codebook)
5. Logbook
6. Read Me file
7. Syntax: Cleaning
8. Syntax: Analysis (potentially multiple)

Data Reporting

The PI generally has pre-established agreements with funders (ex. The Michigan Department of Health and Human Services) about dissemination of the data. It is the practice of the CBHJ to report findings of projects to the funders and community stakeholders prior to academic or public dissemination. In some cases, there may be timelines when the data can be made available for academic or public dissemination. The PI and Directors, along with the other team members, will collaborate on the interpretation and dissemination of all data related to the project.

Plan for Addressing Research Misconduct and Data Mismanagement

All research team members are responsible for letting the PI or Directors know if they suspect data fraud, manipulation, or other misconduct. Team members will communicate any breach or compromise of the data security, confidentiality, or integrity where personal information of an individual was, or is reasonably believed to have been, acquired and/or accessed by an unauthorized person. The PI will then communicate this information to the appropriate authorities.