National Survey of
Early Care & Education

# National Survey of Early Care and Education (NSECE) Level-2 Restricted-use (L2) Data Application and Information Packet

**OPRE Report #2022-53**
March 2022, Updated May 2022

OPRE

# National Survey of Early Care and Education (NSECE) Level-2 Restricted-use (L2) Data Application and Information Packet

**OPRE Report #2022-53**
**January 2022**

**Submitted to:**
Ivelisse Martinez-Beck, Ph.D., Co-Project Officer
Ann Rivera, Ph.D., Co-Project Officer

Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

**Contract #**
HHSP233201500048I

**OMB Control Number:** 0970-0391
**Expiration Date:** 10/31/2024

**Project Director**
A. Rupa Datta
NORC at the University of Chicago
55 E Monroe Street
Chicago, Illinois, 60603

**Suggested Citation:** National Survey of Early Care and Education Project Team (2022). *National Survey of Early Care and Education (NSECE) Level-2 Restricted-use (L2) Data Application and Information Packet,* OPRE Report No. 2022-53, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

**Disclaimer**
The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This documentation and other reports sponsored by the Office of Planning, Research and Evaluation are available at www.acf.hhs.gov/opre.

Sign-up for the **ACF OPRE News E-Newsletter**

Like OPRE on Facebook **facebook.com/OPRE.ACF**

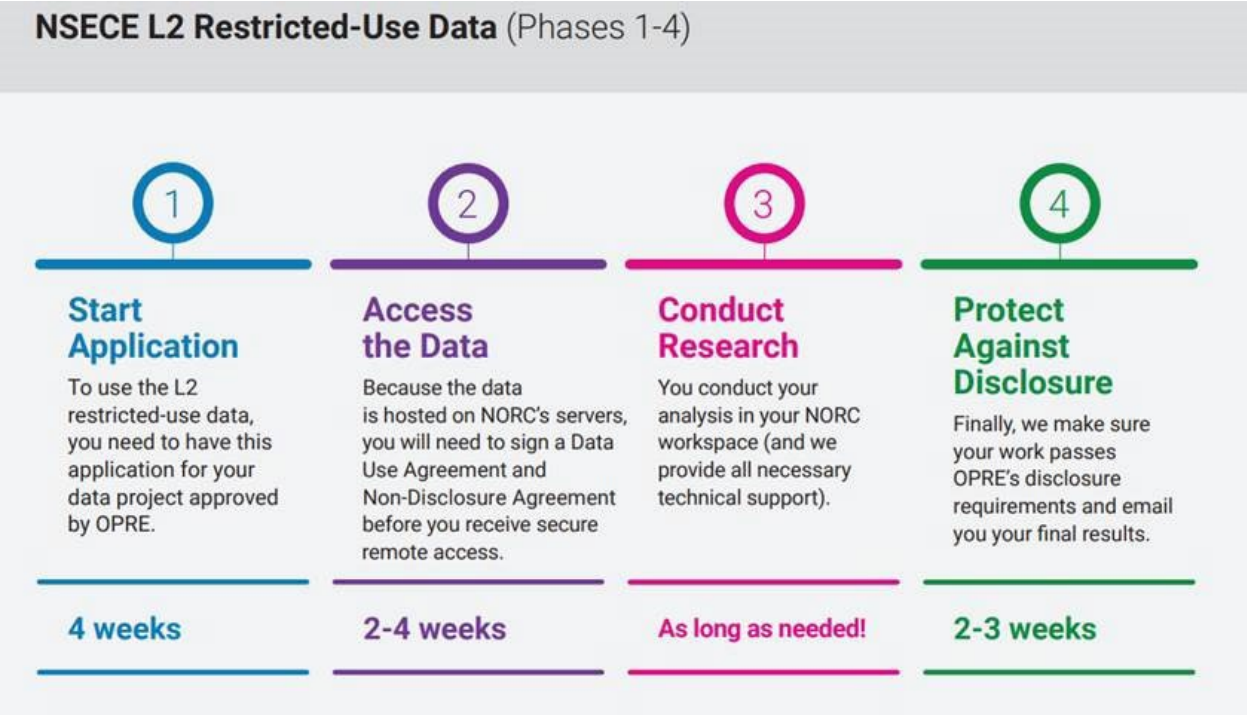
Follow OPRE on Twitter **@OPRE_ACF**

# NSECE Restricted-Use L2 Data Access Application

Thank you for your interest in using the NSECE's Level-2 restricted-use (L2) data! We here at the **NSECE Data Desk** are always excited to help data users get started with their research projects.

The NSECE Data Desk is funded by the Office of Planning, Research, and Evaluation (OPRE) in the Administration for Children and Families, U.S. Department of Health and Human Services.

This document applies to analyses using L2 data from the 2012 NSECE, the 2019 NSECE, and both the 2012 and 2019 NSECE. It describes the four phases involved in using the L2 data, from applying for access to the data to extracting your research results.




**Note:** For a full review of the NSECE data types, and how to access them, please see **Appendix 1** at the end of this document.

# Phase 1: Start Application

**Timeline: Approximately 4 weeks**

| WHAT WE NEED FROM YOU | WHAT WE WILL PROVIDE | KEY DOCUMENTS |
|---|---|---|
| 1. Initial call with NSECE Data Desk<br>2. Completed application | 3. **Initial review** of your application prior to submission to OPRE<br>4. **Answers** to any questions you may have about the NSECE design, data, or L2 process<br>5. **Submission** of application to OPRE and communication with you about approval or any needed revisions | • This document<br>• NSECE Data Types (Appendix 1) |

**Phase 1 begins when you first contact the NSECE Data Desk about accessing the NSECE L2 data.**

We will first provide this document, which we prepared to help users familiarize themselves with the L2 data and the L2 processes. Once you have had a chance to review it, we're happy to schedule an introductory call with you and your research team. We see this call as an opportunity to introduce ourselves and to understand your research goals. You do not need to have every aspect of your L2 research perfectly planned out at this phase, but for us to provide the best possible feedback you should have an understanding as to why your research project requires L2 data (as opposed to L1 or public-use data). We are happy to answer any questions you may have about the NSECE's design, its data, or the NSECE L2 research process during this call.

After the call, you should be prepared to fill out the L2 application. Once you have done so, please send the application back to us so that we can review it. At this point, we may have questions about the specific variables or data files you are requesting or suggestions for how to minimize the risk of disclosing the identity of NSECE sample members that would be part of your research. Please keep in mind that **most L2 applications require a few rounds of back and forth with the NSECE Data Desk** before they are fully completed. If this is the case with your application, please do not think that you have done something "wrong"; we merely want to make sure that you have fully minimized any disclosure risk before we ask OPRE to approve your application for L2 data.

# Phase 2: Access the Data

**Timeline: Approximately 2-4 weeks**

| WHAT WE NEED FROM YOU | WHAT WE WILL PROVIDE | KEY DOCUMENTS |
|---|---|---|
| 1. IRB approval<br>2. Data Use Agreement*<br>3. Non-disclosure Agreement*<br><br>*We provide templates for these in Appendices 2 and 3, links at right | **Access** to VPN, statistical software, and exclusive project folder for you to work in | • **Example of Data Use Agreement (DUA) Template** (APPENDIX 2)<br>• **Non-Discosure Agreement (NDA) Template** (APPENDIX 3) |

We will email the entire research team that you have listed in the L2 application once OPRE has approved it to let you know that your project has moved onto phase 2. We will then send a follow-up email at the start of each phase to remind you of the next steps for your project.

**Once phase 2 starts, you will need to do the following:**

Confirm the names and emails of all individuals working with you on your research team who will need to access the data.

Provide a copy of the IRB approval certificate for your research that you have received from your institution.

Return the data use agreement (DUA) signed by your institutional signatory for your entire project team.

Submit a signed non-disclosure agreement (NDA) **for each member** of your team who will be accessing the NSECE data. The NDA is attached to the DUA.

### *Additional Tips and Suggestions*

It is possible that your institutional signatory may want to revise our DUA and NDA templates. We certainly understand that the language in our templates may not fit all institutions as written; we generalized these documents so that they would be applicable to most circumstances. We welcome your institution's input, but should you need to make any changes, **please make sure to do so in Microsoft Word's track changes feature**. That way, NORC's Data Use Governance Board can easily identify your institution's suggestions. Please keep in mind that your institution and NORC's Data

Use Governance Board may require multiple rounds of back-and-forth before they can agree on the final language in these documents and that this back-and-forth may temporarily delay the start of your research.

**Please note**: only the research team's leader and the institutional signatory need to sign the DUA, but every single researcher on the research team who will have access to the L2 data will need to sign an NDA. **We will only provide NORC logins to individuals who have signed the NDA.** Once you return those signed documents, NORC will counter-sign them, and then we can get you started with the L2 data.

# Phase 3: Conduct Research

**Timeline: As long as you need!**

| WHAT WE NEED FROM YOU | WHAT WE WILL PROVIDE | KEY DOCUMENTS |
|---|---|---|
| 1. You work on your own in your designated folder.<br>2. If you need the NSECE Data Desk to generate any variables on your behalf, we will need your syntax and description of variables. | 1. **L2 data file(s)**, including variables and relevant geographic identifiers for the analytic sample you requested.<br>2. **Additional NSECE public-use or L1 data file(s) that are specific to your project needs**, including variables and relevant geographic identifiers for the analytic sample you requested.<br>3. **NSECE documentation,** including user's guides, disclosure guidelines, and documentation about design-corrected variance estimation.<br>4. *Optional*: Non-NSECE, public-use datasets if you provide detailed documentation for each variable.<br>5. *Optional*: Non-NSECE, restricted-use datasets pending discussion with the NSECE Data Desk | • **Guidelines and Information for L2 Access** (APPENDIX 4)<br>• **VPN Setup** (APPENDIX 5) |

**Phase 3 begins once we have received a signed DUA for your project and NDA's for all individuals working on the research.**

Because of the uniquely powerful design of the NSECE, the NSECE L2 data is stored on NORC's server. OPRE has deemed it necessary to take this step to protect against the disclosure of the NSECE sample members. You will access NORC's server using your web browser.

**You will be assigned a project folder for your materials on NORC's server.** The name of this project folder will begin with your research team's project number and will be followed by the Principal Researcher's last name (e.g., 2000_Jones). Only NORC staff and individuals on your research team (who have signed an NDA) will be able to access this project folder. Please note that after posting your files, the NSECE Data

Desk will not access your project folder unless you ask us to (for instance, if you ask us to execute syntax on your behalf, or if we need to carry out a disclosure review prior to extracting your completed documents).

In this folder, you will be able to access the L2 data you have requested, as well as any supporting materials such as corresponding quick tabulation data, public-use data, and L1 NSECE data.

In your project folder you will have access to Microsoft Office programs and all major statistical programs (Stata, SAS, SPSS, R, and StatTransfer).

**To keep the NSECE data secure, some activities are restricted within this work area. You will not be able to:**

- Print from the project folder
- Copy and paste to documents outside of the project folder
- Upload or download files to your computer
- Email items as attachments from the folder.

**To protect against disclosure, you are also forbidden from manually transcribing any figures or estimates from your project folder to an external document or file.** L2 data is not allowed to leave your project folder under any circumstance until it has passed a disclosure review. We discuss this process in the following section.

We will also upload any additional supporting materials to your project folder at this time. These materials may include other, non-NSECE, public-use datasets (so long as you provide documentation), or any code for statistical programs you need to conduct your analysis. We are happy to upload any supporting material so long as it does not contain information that could jeopardize the privacy of the NSECE's sample members.

**If you want us to upload any additional, non-NSECE, public-use data files to your project folder, we will need…**

- A list of all variables you want uploaded
- A list of all the sources of these variables *(i.e., the names of the datasets)*
- An assurance that this data is in fact publicly available

**If you want us to upload any additional, non-NSECE, restricted-use data files to your project folder, we will need…**

- A conversation with your research team to ensure that the non-NSECE, restricted-use data you have requested does not jeopardize the identity of NSECE sample members and that you have secured whatever data use agreement necessary for that additional, non-NSECE data.

***A note about provider cluster IDs (this does not apply to every project!):***

If your research team is interested in:

- conducting any analyses at the level of the provider cluster
- developing any measures at the level of the provider cluster
- merging provider cluster information onto the household sample

then you will need access to the NSECE's simulated provider cluster IDs.

For more information on provider clusters or on NSECE's sampling design, more generally, please see the NSECE's **Summary of Data Collection and Sampling Methodology** report, available on OPRE's website.

If this is the case, we will have already walked you through this process when you prepared your application. Just as an overview, you will have access to Simulated Cluster IDs for every center-based, home-based, or workforce case in your analytic sample. You will generate variables at the level of the provider cluster using these simulated cluster IDs. Once you have settled on the provider cluster aggregates you want to use, you provide us with the syntax, and we run it with the real provider cluster IDs. We then create summary statistics for you using your syntax and merge those aggregate statistics to the household file. This file will be posted to your designated folder with all provider cluster aggregates attached. Please see our **Disclosure Guidelines for L2 NSECE Users Working with Provider Clusters** for more information.

# Phase 4: Protect Against Disclosure

**Timeline: Approximately 2-4 weeks**

| WHAT WE NEED FROM YOU | WHAT WE WILL PROVIDE | KEY DOCUMENTS |
|---|---|---|
| 1. Description of files you want to extract<br>2. Raw and final versions of each file | 1. **General assessment of disclosure risk** (some teams in the past have found it helpful to do this over the phone to minimize back-and-forth)<br>2. **Detailed review of every cell in every table** to confirm that suppression and rounding have been properly implemented<br>3. **Extraction agreement and submission agreement to OPRE**<br>4. **Final version** of tables to you | • **L2 Reporting Guidelines** (Appendix 6)<br>• **L2 Extraction Checklist** (APPENDIX 7)<br>• **Disclosure Guidelines for L2 NSECE Users Working with Provider Clusters** (Appendix 8) |

**Phase 4 begins once you have completed your research and are ready to extract your results.** Before we can extract these results, however, we need to ensure your data meets the disclosure guidelines. This is a two-step process that we detail below.

Please keep in mind that the better you understand and adhere to the disclosure rules, the less likely your work will require multiple rounds of review. Taking the time to understand these disclosure rules early on will put you in a position to receive your files sooner.

### Part 1: General assessment of disclosure risk

When you notify us that you are ready to begin the disclosure review process, we will ask you the questions outlined below. You probably addressed some of these questions in your initial application, but the answers may have changed over the course of your research:

- **What was your analytic strategy?**
- **What was the specific subset of cases you are using in the analysis?**
- **What was the geographic reach of your estimates (e.g., national, subgroups of states, specific states)?**

- **What are the specific estimates you are planning to extract (e.g., percentage of X regression estimate of Y variable on X variables)?**
- **How did you specifically use NSECE L2 data (e.g., are you identifying any states, are states used in a regression but state-level estimates are not reported, are you combining cluster-level aggregates with information on community characteristics)?**
- **Did you use any other non-NSECE data files? If so, what are those data files, the specific variables you used, and the geographic reach of those variables?**

We need to understand the details of your analysis to assess the disclosure risk involved. Our key concern is to prevent the identification of any geographic unit below the state level, such as a county or the specific location of any sample member. We review the overall risk of your estimates and consider the extent to which your estimates, in combination with other reported estimates, may lead to the disclosure of sub-state units.

In addition, you will need to describe all the files you want to extract, as well as provide us with the final and raw versions of these files. By raw, we mean files where all figures are unrounded and unsuppressed.

**Part 2: Detailed review of every estimate!**

Once you provide us with a description of the files you want to extract, as well as raw and final versions of all the estimates in each file, we will conduct a detailed review of every cell to make sure it meets our disclosure requirements. These requirements are detailed in **Appendix 6: L2 Reporting Guidelines**. For the record, when we say we review every estimate, we really mean it! *We manually review every cell of every table before we clear your research for extraction.*

**Please understand that this process is iterative;** it may require multiple rounds of back-and-forth between you and the NSECE Data Desk before you receive your final files. Even if you have to adjust or otherwise fix only a handful of numbers in response to our feedback, we are still required to once again review *all* of the figures in your research, not just the ones you have changed. This is what can make the process so time consuming.

**Once phase 4 is complete, we email you your final files and you will be able to share your results publicly!**

Let us know if you have any questions. You can reach us at **NSECE@norc.org**.

***A final note*:** Protecting the confidentiality of the NSECE data and samples requires particular protection of the geographic locations where NSECE sample members are located. Therefore, **you must design any analyses of L2 data so that the description of research methods and study context, reporting of analyses, and discussion of findings can be completed without any direct or indirect disclosure of locations below the state level**. Of course, individual sample units' identities should also not be disclosed directly or indirectly. Researchers will be responsible for gathering and documenting proof of adequate disclosure protection.

**Thank you for reading!**

# Appendices: Key Documents

Links below…


**Appendix 1: NSECE Data Types**

**Appendix 2: Example of DUA Template**

**Appendix 3: NDA Template**

**Appendix 4: Guidelines and Information for L2 Access**

**Appendix 5: VPN Initial Setup**

**Appendix 6: L2 Reporting Guidelines**

**Appendix 7: L2 Extraction Checklist**

**Appendix 8: Disclosure Guidelines for L2 NSECE Users Working with Provider Clusters**

# Appendix 1: NSECE Data Types

| Data type | Type of data | Expected users | Approximate number of variables | Available through… |
|---|---|---|---|---|
| **Main public-use (PU)** | • All questionnaire response data represented subject to disclosure considerations<br>• Extensive created variables<br>• Community characteristics from ACS data<br>• No identifying information | Academic researchers with programming knowledge and statistical expertise | 200-26,000 | **The NSECE page at the Child and Family Data Archive** (CFData Archive) |
| **Quick-tabulation (QT)** | • Some questionnaire response data<br>• Some created variables<br>• Community characteristics from ACS data<br>• No identifying information | Agency staff, policy firms, and researchers exploring data | 100-200 | **The NSECE page at the CFData Archive** |
| **Level-1 restricted-use (L1)** | • Questionnaire response data with disclosure risk | Academic researchers with programming knowledge and statistical expertise | 20-200 | Restricted-use files from the 2012 NSECE are available through **the NSECE page at the CFData Archive**<br>Please email NORC at nsece@norc.org for more information about accessing restricted-use files from the 2019 NSECE. |
| **Level-2 restricted-use (L2)** | • PSU and state identification where possible<br>• SSU identifiers (linking variables, not actual location) | Academic researchers with programming knowledge and statistical expertise who have specific research projects in mind | Varies | Email **NSECE@norc.org** for application<br>*Additional approval needed, as detailed in this document* |

# Appendix 2: Example of DUA Template

***NOTE*: This is merely an example DUA, included only for demonstration purposes. This example likely differs from the DUA a research team's institution will sign. Please read the description of Phase 2 above for more information about the DUA and its timing.**

**Data Use Agreement**
**Between**
**National Opinion Research Center (NORC)**

**And**
**NAME HERE**

This Data Use Agreement ("Agreement") is made and entered into as of DATE (the "Effective Date") by and between **National Opinion Research Center ("NORC"),** and **NAME** ("DATA USER"), as it may be amended from time to time by the agreement of the parties or as required by federal and state laws, rules, and regulations that may apply to the protection of the data shared under this Agreement. This Agreement establishes a formal data access and data use relationship between NORC and DATA USER, to enable DATA USER DESCRIBE DATA USER USE CASE HERE.

**Attachments:**

A. Details of Data Sharing

B. Data Security Requirements

**WITNESSETH:**

WHEREAS, NORC may disclose or make available to DATA USER, and DATA USER may use, maintain or create from, certain information in conjunction with STATE PROJECT NAME AND PURPOSE HERE.

WHEREAS, NORC and DATA USER are committed to compliance with applicable privacy laws, including the federal Privacy Act, the Privacy Rule and regulations promulgated thereunder; and

WHEREAS, the purpose of this Agreement is to satisfy the obligations of the parties and to ensure the integrity and confidentiality of certain information disclosed or made available to DATA USER and certain information that DATA USER uses, discloses, receives, transmits, maintains or creates, from NORC.

EXAMPLE

NOW, THEREFORE, in consideration of the foregoing recitals and other good and valuable consideration, the receipt and sufficiency of which is hereby acknowledged, the parties agree as follows:

## 1. DEFINITIONS

Terms used but not otherwise defined in this Agreement shall have the same meaning as those terms in the Privacy Rule.

A. Individual shall have the same meaning as the term "individual" in the Privacy Act of 1974 and shall mean a citizen of the United States or an alien lawfully admitted for permanent residence.

B. Personally Identifiable Information (PII), as defined in OMB Memorandum M-07-16, refers to information that can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information that is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc.

C. Privacy Rule shall mean the Standards for Privacy of Individually Identifiable Information at 45 CFR Part 160 and Part 164, Subparts A and E, as amended from time to time.

D. The Privacy Act of 1974 shall mean 5 U.S.C. § 552a.

E. Data Set shall mean the data disclosed to DATA USER by NORC as described in Attachment A and any records, reports, databases, and/or other derivative documents created from the Data Set provided under this Agreement.

## 2. SCOPE AND PURPOSE

A. This Agreement sets forth the terms and conditions pursuant to which NORC will grant DATA USER access to the Data Set and obligations and responsibilities of DATA USER in relation to the Data Set.

B. DATA USER shall use Data Sets solely for the research purposes stated in the Proposal. Data Sets may not be used for any other purposes whatsoever, including administrative, regulatory, marketing, law enforcement, judicial, or other purposes.

## 3. OBLIGATIONS AND ACTIVITIES OF DATA USER

A. DATA USER agrees to protect the Data Set according to the appropriate security standards to prevent use or disclosure of the Data Set and in accordance with the Data Security Requirements attached to this Agreement as Attachment B. DATA USER agrees to notify NORC within 1 hour of any knowledge of unauthorized access or use of the Data Set that DATA USER becomes aware of while using the Data Set.

EXAMPLE

B. DATA USER acknowledges that any and all information of any form contained in the Data Set disclosed under this Agreement shall be deemed confidential. DATA USER agrees to hold the Data Set in strict confidence and shall not copy, reproduce, sell, assign, license, market, transfer or otherwise dispose of, give or disclose such information to third parties or to use such information for any purpose whatsoever, other than as described in this Agreement or as required by applicable privacy laws. DATA USER shall not disclose the Data Set to any third party without specific written authorization from NORC. DATA USER agrees that he/she shall not attempt to capture, store or share any images, files or information accessed or attempt in any way to remove, excerpt or duplicate the Data Set.

C. DATA USER specifically agrees to comply with state and federal confidentiality and information disclosure laws, rules, and regulations applicable to information disclosed under this Agreement.

D. DATA USER assumes the entire risk related to its use of the Data Set. NORC is providing the data "as is," and NORC disclaims any and all warranties, whether express or implied, including (without limitation) any implied warranties of merchantability or fitness for a particular purpose. In no event will NORC be liable to DATA USER or any third party for any direct, indirect, incidental, consequential, special or exemplary damages or lost profit resulting from any use or misuse of the Data Set.

E. DATA USER agrees to indemnify and hold harmless NORC, its officials and employees and agents from, and defend NORC, against any liabilities or expenses (including attorney's fees and costs) arising out of or in relation to (1) any unauthorized or unlawful disclosure or release of any information obtained under this Agreement; and/or (2) any breach of the Agreement by DATA USER

## 4. TERM AND TERMINATION

A. The provisions of this Agreement shall be effective as of the Effective Date and DATA USER access to the Data Set shall terminate on DATE.

B. The confidentiality provisions of this Agreement shall survive the termination of this Agreement.

## 5. MISCELLANEOUS

A. The parties agree to take such action as is necessary to amend this Agreement from time to time as is necessary for NORC to comply with the requirements of applicable privacy laws.

B. The respective rights and obligations of DATA USER under Articles 3 and 5 of this Agreement shall survive termination of this Agreement.

EXAMPLE

C. Any ambiguity in this Agreement shall be resolved to permit NORC to comply with applicable privacy laws.

D. Order of Precedence. This Agreement shall be interpreted as a unified contractual document with the clauses and the Attachments having equal effect, except that in the event of any inconsistency between them, first the Articles in this Agreement will apply, then the Data Security Requirements (Attachment B), then the Statement of Work (Attachment A), and then the remainder of the Attachments incorporated into this Agreement.

E. There are no intended third party beneficiaries to this Agreement. Without in any way limiting the foregoing, it is the parties' specific intent that nothing contained in this Agreement gives rise to any right or cause of action, contractual or otherwise, in or on behalf of the individuals whose PII is used or disclosed pursuant to this Agreement.

F. No provision of this Agreement may be waived or modified except by an agreement in writing signed by the waiving party. A waiver of any term or provision shall not be construed as a waiver of any other term or provision.

G. The persons signing below have the right and authority to execute this Agreement and no further approvals are necessary to create a binding agreement.

H. In the event of any conflict between the terms and conditions stated within this Agreement and those contained within any other agreement or understanding between the parties, written, oral or implied, the terms of this Agreement shall govern. Without limiting the foregoing, no provision of any other agreement or understanding between the parties limiting the liability of DATA USER to NORC shall apply to the breach of any covenant in this Agreement by DATA USER.

IN WITNESS WHEREOF, the parties have executed this Agreement effective upon the Effective Date set forth above.

**APPROVED BY:**

| NORC | DATA USER |
|---|---|
| Signature: | Signature: |
| Print Name: | Print Name: |
| Title: | Title: |
| Date: | Date: |

EXAMPLE

## ATTACHMENT A

## DETAILS OF DATA SHARING

1. DATA USER shall not attempt to re-identify respondents or sample locations, including without limitation persons, families, households, schools, establishments, organizations, economic units or any other entities ("Respondents" or "respondent locations"); nor will any list of identities or raw data elements be published or otherwise distributed

2. DATA USER shall not publish any work containing information or results derived from Data Sets that identifies a Respondent, Household, or Organization.

3. DATA USER shall respond to NORC and requests for information on research progress and preliminary findings.

4. DATA USER shall ensure that all persons who have access to the Data Set shall be fully apprised as to the confidential nature of the Data Set, the safeguards required to protect the information, and the applicable civil and criminal sanctions and penalties associated with any intentional or non-intentional disclosure must sign and submit a non-disclosure agreement specific to the Data Sets for each Approved Researcher that will be accessing the Data Sets. Approved Researchers are the DATA USER and any other researchers that the DATA USER needs to grant access to the Data Sets. In order for an additional researcher(s) to become an Approved Researcher with a right to access the Data Sets, they must sign and submit to NORC a completed copy of the Non-Disclosure Agreement attached hereto as Exhibit A.

EXAMPLE

## ATTACHMENT B

## DATA SECURITY REQUIREMENTS

1. **Security of NORC Data:** To prevent loss and unauthorized access or disclosure, DATA USER shall use current State or Federal regulation required controls while accessing the Data Set.

2. **Facility & Workstation Security:** For any DATA USER facility containing personnel or systems used in the viewing, processing, or otherwise accessing Data Sets, the DATA USER shall implement the following security requirements:

    a) DATA USER Workstations performing NORC related functions shall be positioned so that the Data Set is not visible to unauthorized personnel.

    b) In accordance with federal regulations, within 1 hour of the time when a DATA USER becomes aware that the identity or identifying information of any Individual, Household, or Organization may have been disclosed, DATA USER shall advise NORC of the disclosure, inadvertent or otherwise. DATA USER shall use his/her reasonable efforts to destroy, retract, or otherwise safeguard from further dissemination of the materials containing the Respondent, Household, or Organization's identity or containing information from which the Respondent, Household, or Organization's identity may be determined.

    c) DATA USER personnel must lock their workstations when they are away from their desks. Workstations shall be configured to automatically lock after no more than 10 minutes of inactivity.

3. **Anti-Malware:** DATA USER's computer involved with accessing or processing the Data Set are protected with up-to-date anti-malware software. DATA USER shall have a process in place for issuing regular updates to anti-malware software and conducting regular scans of the environments containing the Data Set.

4. **Access to Data:** DATA USER agrees to maintain the following controls: DATA USER shall notify NORC within 24 hours when any personnel with access to the NORC network has left the project or is no longer employed by DATA USER.

EXAMPLE

# Appendix 3: NDA Template

Between NORC, RESEARCHER ORGANIZATION & Approved Researchers

This non-disclosure agreement applies to all **Approved Researchers** who at any point may view material contained in the Data Sets made available by NORC under the DUA with RESEARCHER ORGANIZATION dated DATE. Materials in the Data Sets may contain personally identifiable, sensitive, and/or proprietary information. Users with access to Data Sets agree to protect the information contained therein. Prior to accessing the Data Sets, approved users must agree to conduct themselves in accordance with NORC's policies regarding ethical conduct as well as that which is established by NORC's physical, data, and computer security protocols. My signature on this document attests that I have read and understood the following non-disclosure policies and agree to honor them.

- I certify that I am authorized to access the Data Sets listed above.
- I agree that I will not disclose to anyone who is not also authorized under this agreement information contained in the Data Sets except with NORC's prior written approval.
- I agree to conduct my viewing of Data Sets with due caution, logging out when I am not at my computer and preventing unauthorized viewing of or access to my machine while I am accessing secure data. I agree not to produce any unauthorized transcriptions or copies of any of the raw data.
- Those accessing the Data Sets are reminded that the Economic Espionage Act of 1996 prohibits the theft or intentional conversion of trade secrets to the economic benefit of anyone other than the owner of the trade secrets. Penalties for violation of this law include prison terms of up to 10 years and substantial fines. Thus, I must never knowingly disclose or use any trade secrets or confidential information that I access while viewing the Data Sets.
- The information and data contained in the Data Sets, if disclosed to unauthorized sources, may violate the Privacy Act of 1974 (5 U.S.C. § 552a) and may result in financial loss or adverse legal actions.

Printed Name _____

Signature _____

Organization _____

Date _____

# Appendix 4: Guidelines and Information for L2 Access

### Getting Started with Access

Research teams will work with the L2 NSECE data inside the NORC secure environment where each team will be assigned a project folder for their materials. The name of this project folder will begin with the research team's project number and be followed by the Principal Researcher's last name (e. g., 2000_Jones). Only NORC staff and individuals on the research team will be able to access this folder. Research teams will access their project folder through a secure portal that can be reached through an internet browser. To gain access, you will need 1) a NORC account for each individual on your team and 2) a computer with a security certificate installed. The NSECE team will create NORC accounts for you and your team members as the administrative details are being sorted out. Once all paperwork has been received, we will provide you with your usernames and log-ins. Then you can install the security certificate on your designated computer and confirm access to the work area. (Instructions for installing the certificate are included in this packet.) All authorized researchers on your team with NORC accounts will be mapped to your project folder and be able to access the files located there. Data are always available without restriction.

### System Recommendations

The NORC system has been designed to work with computers running a Windows operating system (version 8 or higher). If you have an operating system other than Windows and/or use a different web browser, you may still be able to access the NORC environment, but it might take a little longer to setup your access. All machines (whether running Windows or not) must be updated with the current service pack for their operating system and have antivirus software installed.

### Issues Accessing the NORC Environment

There are times when external users have difficulty getting access set up, unexpectedly lose access to the NORC environment, or experience other IT-related problems. If you are having an issue, we will be able to help during regular business hours (9am-5pm Central Time, Monday-Friday, except holidays). We will be able to let you know within one full business day when your access should be restored. If you experience a technical support problem, it may be necessary for NORC technical support staff to be on the phone or virtually connect to your machine with you. We will work to

accommodate your schedule, but your availability will affect how quickly we are able to resolve your problem.

Someone from the NSECE Data Desk will work with you and help you to facilitate calls with our IT team and get these issues resolved.

*If you experience issues and need some IT help, please contact:*

**NSECE@norc.org**
1-888-TRY-NORC

**Working in the NORC Secure Environment**

**To keep the NSECE data secure, some activities are restricted within your work area. Individuals with external access to the NSECE data will not be able to:**

- **print from the project folder**
- **copy and paste to documents outside of the project folder**
- **upload or download files to their computer**
- **email items as attachments from the folder**

If you have items you need available in the work area, these should be specified in the research plan and they will be made available for you. If you would like something that has not been specified in the research plan, please see the section below for common requests and instructions.

**Contacting Us with Questions and Requests**

If you have questions about the work area or have additional requests, please contact us.

Although research teams can access their project folders any time of day, NORC operates during normal business hours (9am -5pm CT; Monday – Friday, except holidays). If an issue arises outside of these hours, it will not be addressed until staff are in the office and can respond to the request. Below we list some common requests along with instructions on how to proceed. A summary table is included at the end of the list with the estimated time we expect a response to take.

- **Upload Code:** You may request that we upload code (Stata, SAS, SPSS, R) that you have written outside of the NORC secure environment. That code needs to run on NSECE data files as they are posted on the Child and Family Data Archive. **We will upload code within two business days of receiving your request, which you may email to NSECE@norc.org with the subject line: L2 upload request (PROJECT NUMBER).**
- **Add a New Researcher:** You may ask that we add a new researcher to your team. **We will add such a person within 5 business days of receiving that person's signed NDA.** (It can be scanned and e-mailed to **NSECE@norc.org with the subject line: L2 additional user (PROJECT NUMBER).)**
- **Upload Additional, Non-NSECE, Public-Use Data:** Your project may require incorporation of public-use data outside of the NSECE (for example, the CCDF policy database). Before we can upload any additional data, you must first provide us with a list of all the variables you want to upload, a list of all the sources of these variables (i.e. the names of the datasets they belong to), and an assurance that this data is in fact public-use. Once you email us that information, then you can email us the data you want us to upload, either as a link or as an attachment. **Files will be uploaded within five days of receipt of request to NSECE@norc.org with the subject line: L2 file upload request (PROJECT NUMBER). Please note that there are specific reporting restrictions governing use of publicly available data in conjunction with NSECE data. What this means is that if you combine any non-NSECE, public-use data with the NSECE's L2 data, the L2 Reporting Guidelines will still apply to your results.**
- **Upload Restricted-Use Data Outside of NSECE:** If your project requires incorporation of restricted-use data outside of the NSECE, separate arrangements will be made for handling of those data.
- **Upload a Document:** You may request uploading of a document or other file to the secure environment. **We will post within three business days or let you know of further questions/issues.**
- **Consult on Implementation of Disclosure Guidelines:** You may wish to consult with us on implementation of disclosure guidelines or small modifications to your approved research support plan. **We will review and respond to such requests within five business days.**
- **Upload Other NSECE Data Sets (not restricted-use):** If you wish to request upload of other NSECE data sets (QT, Main Public-Use or approved Level 1) for any of the four surveys), **we will provide you access within 2 business days.**

- **Download of Tables or Other Material:** If you wish to request download of tables or other material from the secure environment, please submit a request and document compliance with the disclosure guidelines you have received. This may require, for example, that you provide rounded/unrounded, suppressed/unsuppressed versions of your tables for inspection. **We will provide feedback on any request (including providing the files if approved) within 10 business days.**
- **Revise Research Plan:** If you wish to revise your research plan and receive new disclosure approval and guidelines, submit a revised application (Phases 1 and 2, as needed) or just amended tables and documentation. **We will provide revised guidelines within 10 business days.** An introduction of new data sets or external variables or a revised L2 data file request are considered revisions of your research plan.

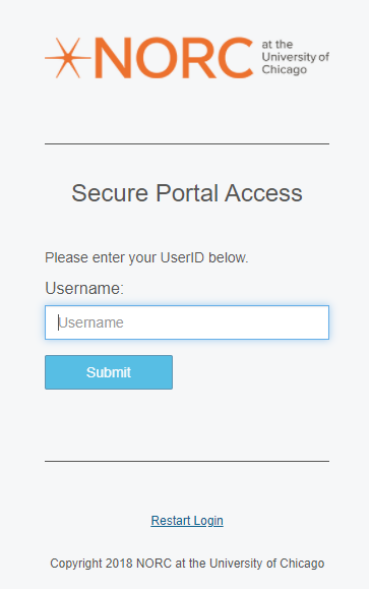
| Request/Activity | Estimated # business days |
|---|---|
| Upload code to project folder | 2 |
| Add a new researcher to project team | 5 |
| Upload a public-use data file outside of NSECE | 5 |
| Upload a restricted-use data file outside of NSECE | Varies |
| Upload a document | 3 |
| Consult on implementation of disclosure guidelines | 5 |
| Upload a standard NSECE data set | 2 |
| Download tables or other material | 10 |
| Revise research plan | 10 |
| Estimate for custom data request | 3 |

# Appendix 5: VPN Setup

NORC has its own VPN (Virtual Private Network), which you will need to access to work with your NSECE data. To do so, please follow these steps.

1. Please go to the VPN site at **http://remote.norc.org/**

   *(The website will redirect to **https://eidp.norc.org/secureauth30/**, which is perfectly fine)*




2. Enter your username, which is your lastname-firstname (ex. Duff-Bryan).

3. Select your phone number or email, choose your preferred delivery method for NORC's 2-factor verification, and click submit.

4. Enter the 6-digit code you received from the method you chose in step 3.

5. Enter your VPN/NORC password and click submit.

**Notes:**

- If this is your first time signing in to the VPN, you may be asked to download two files: "AGEE_setup.exe" and "nsepa_setup64.exe". You can trust these!

- If you receive an error message or experience any difficulty at any time, please call us at 1-888-TRY-NORC.

# Appendix 6: L2 Reporting Guidelines

**Disclosure Requirements for NSECE Research Using L2 Data**

This document specifies what's required of NSECE data users to prevent disclosing

- The identities of NSECE study subjects
- The sample size of any specific state included in the NSECE
- The presence or absence of any sub-state geographic unit in the NSECE sample, such as county, city, or school district
- The location of a center-based or home-based provider, workforce member, or household in the sample

The requirements in this document apply to

- Analyses using L2 data from the 2012 NSECE, the 2019 NSECE, and analyses using L2 data from both the 2012 and 2019 NSECE.
- Reporting of unweighted and weighted estimates, including
  - frequencies, means, coefficients from multivariate analysis, standard errors associated with these estimates, and any graphical representation of these estimates, among others

**Any document, material, output, or other information that an NSECE L2 data user requests to remove from the NORC secure environment must meet these requirements.** Only by meeting these requirements will a research team have permission to disseminate or communicate any findings from analyses of the NSECE L2 data to the general public (i.e., to any individual who has not signed a non-disclosure agreement).

Though the requirements are nonnegotiable, research teams may find there are different ways to go about complying with them. We provide some ideas for compliance in throughout this appendix.

As you can see in **General Guidelines for L2 Files** section below, each NSECE survey has a different suppression threshold. If a research team is working with data from different NSECE surveys, the team will need to make sure that estimates meet the disclosure requirements for each survey. For example, if you are using L2 data from the center-based survey, you will need to suppress any weighted or unweighted estimates from that data for any cells with 50 or fewer cases. Likewise, if you are using data from the workforce survey, you will need to suppress any weighted or unweighted estimates from that data for any cells with 10 or fewer cases. In this workforce example, you cannot just rely on the suppressing 50 or fewer cases for all the data; each data must follow its own suppression threshold.

### General Guidelines for L2 Files

The reporting requirements regarding suppression and display of estimates defined for NSECE L2 data establish the following:

| | Suppression of weighted and unweighted estimates based on specific unweighted cell size | Significant figures of unweighted and weighted estimates (including counts, proportions, standard errors, regression coefficients, and other statistical measures) |
|---|---|---|
| **Center-based** | 50 or less | 3 significant figures |
| **Workforce** | 10 or less | 3 significant figures |
| **Household** | 20 or less | 3 significant figures |
| **Home-based Listed** | 20 or less | 3 significant figures |
| **Home-based Unlisted** | 10 or less | 3 significant figures |

### Specific Requirements for Reporting of L2 Data

#### Unweighted cell sizes

In any tabulation that includes L2 data directly or indirectly, unweighted cell sizes must be rounded to the nearest 20 for all data files.

#### At least 3 states

Any reported ‘cell’ must represent a *population* of interest that contains at least 3 states (where DC can be considered a state). A ‘cell’ corresponds to a single cell in univariate tables, any cell in cross-tabulations, totals for a row or column in cross-tabulations, unweighted counts associated with measures of dispersion or central tendency, and underlying unweighted counts of any single multivariate estimate. ‘Cell’ can apply also to non-statistical estimates, for example, text, footnotes, visual

representations, or other reporting about NSECE research that uses or contextualizes the NSECE data analyzed in the research.

**Example:**

If a cell reports data on a particular policy that allows low-income households with a child 0 to 3 years old to register in a specific early care and education program identified in the data, the researcher may use datasets other than the NSECE to demonstrate that there are at least 3 states in the country that have such policy for households with those same demographic characteristics (i.e., low-income households with a child 0 to 3 years old). **If there are not 3 such states, then researcher must demonstrate that at least 3 counties exist in any state that does have such policy.**

- If a cell in a particular analysis includes fewer than 3 states (for example, because only a single state is being studied), then that cell is treated as if it were a state-specific estimate that is reported at the state level and must represent a population that covers at least 3 counties within the state.

### State-specific estimates require the same disclosure steps

Researchers may identify specific states when reporting their findings. However, the same disclosure principles identified earlier apply here, including requirements for suppression of small cells, rounding of unweighted sample sizes, and number of significant figures that may be reported.

- To report a state-specific estimate, whether it reports state name, researchers must demonstrate that the estimate represents a population that covers at least 3 counties within the state.

**Example:**

**An estimate for urban population within a state must be supported with evidence that the state has 3 or more urban counties.** If that estimate is for Head Start centers in those urban counties, then documentation is necessary that at least 3 urban counties in the state have Head Start programs.

### County-level disclosure review

Any analysis that uses county-level characteristics will undergo additional disclosure review. Examples of county-level characteristics are urbanicity level, county-level policies within a state, or variation in the timing of policies' implementation within a

state. Any state-specific estimate for DC, whether it is labeled as DC, will undergo additional review as if it were a county-level characteristic.

### Geographically diverse requirements

It is the researcher's responsibility to demonstrate that a represented population is adequately geographically diverse. In the absence of such demonstration, the estimate is assumed to pose excessive disclosure risk and will not be approved for release from the secure environment.

## Working with Provider Clusters

To minimize disclosure risk, the NORC team produced **Simulated Provider Cluster IDs**. L2 research teams only have access to these Simulated IDs. Simulated Provider Cluster IDs are designed so that, across all clusters, summary statistics produced with Simulated IDs are very similar to those produced with real Provider Cluster IDs.

The process of working with provider clusters is designed to unfold in the following way:

- **Generation of cluster aggregates using Simulated Provider Cluster IDs**

  **Provide syntax.** Research teams provide NORC syntax that creates aggregates at the cluster level, using Simulated Cluster IDs.

**Here is some sample syntax in Stata that would allow research teams to create cluster aggregates:**

```
// Please update the following variables
// NUM = Number of clusters
// VAR = Variable used to create the cluster aggregate
// VAR_cluster_wt = Cluster aggregate variable name
// STAT = Statistic that will be calculated (p50, sum, mean, etc)

egen group = group(L2_PCLUSTER_FALSE_ID)

gen VAR_cluster_wt=.
forvalues i = 1/NUM {
      sum VAR [w=L2_CB_WEIGHT_PCLUSTER] if group ==`i',detail
      replace VAR_cluster_wt = r(STAT) if group ==`i'
}

// Keep only one observation per cluster
keep l2_pcluster_id group VAR_cluster_wt
duplicates drop
count
```

**The following syntax allows for calculation of standard errors associated with any statistics users are generating for a particular cluster X:**

```
// SVYSTAT = Statistic available for the SVY command in Stata (e.g. svy: mean)

svyset [pweight = L2_CB_WEIGHT_PCLUSTER], strata(L2_STRATA)
svy linearized: SVYSTAT VAR if group == X
```

**Generate summary statistics.** Research teams generate summary statistics of cluster aggregates. A table with summary statistics may include a list of variables, unweighted number of providers included in each estimate, and relevant unweighted and weighted estimates (such as means or proportion of any center-based providers in the cluster who have a certain characteristic). Research teams include commands used to generate these summary statistics in the same syntax file, so that NORC can generate the same estimates using real Provider Cluster IDs.

**Research teams provide NORC a list of all the cluster aggregates the team intends to generate (i.e., a list of variable names and labels).** This will allow NORC to make sure all cluster aggregates produced match the research team's intent.

**Generate unweighted count of providers.** For each cluster aggregate, the research team generates an unweighted count of providers per provider cluster. For instance, if the research team produces a variable such as the proportion of center-based providers serving children of a particular age, then the research team would

produce a variable that identifies the number of center-based providers included in that proportion and that count will be calculated for each provider cluster. The main purpose of these counts is to allow the research team to understand how disclosure guidelines will affect the data they will be able to access.

**Confirm categorization of counts.** In case the research team intends to create cluster aggregates defined as weighted or unweighted counts, the research team would need to specify in their syntax how to categorize these counts. Based on the disclosure guidelines listed below (see **Disclosure Guidelines for L2 NSECE Users Working with Provider Clusters**), research teams are allowed to access counts when counts are defined as categorical variables, with a maximum of 5 different categories.

**Specify desired cluster aggregates.** Per the same disclosure guidelines, research teams specify which two cluster aggregates they want to access for all provider clusters, regardless of the number of providers in the cluster.

### Re-generation of cluster aggregates using real Provider Cluster IDs

**Updated summary statistic report.** NORC re-runs all syntaxes using real Provider Cluster IDs and generates an updated summary statistic report.

### Disclosure review

- **NORC reviews data file with cluster aggregates and implements disclosure guidelines,** specified in section **Disclosure guidelines for L2 NSECE users working with provider clusters**. Based on these guidelines, NORC may request further clarification from the team such as the meaning of certain cluster aggregates or the geographic distribution of certain provider attributes. NORC may need to ask the research team to specify variables in different ways (e.g., transform continuous variables into categorical variables).
- **NORC suppresses select provider clusters.** Except for the two cluster aggregates specified by the research team, NORC will suppress cluster aggregates for households or children located in provider clusters with fewer than 3 providers.
- **Determine who generates ratios and performs analysis.** An important distinction is whether research teams access the data file (“see the data”), generate these ratios themselves, and run the final analysis themselves, or whether NORC does all or some of these steps on behalf of the research team.
    - The data that research teams can see are more limited than the data they can utilize without seeing. The data that research teams can see are limited by the disclosure rules outlined below. So, for instance, the household data file posted in the team’s secure folder will display unweighted and weighted counts in categories and teams would only be able to see up to two variables

for households located in provider clusters that have fewer than 3 center-based providers.

- However, if the research team wants to create ratios that use as the denominator a continuous version of a weighted count of providers, they may generate the appropriate syntax and NORC will run this syntax to (i) produce the variable, (ii) generate summary statistics for this variable, and (iii) include this variable in any multivariate model the team is planning to estimate. In this example, the research team would not have direct access to a data file with the value of this variable for each household record, but they could utilize the whole range of the data if NORC manipulates the data files.

### Disclosure review prior to presenting to non-restricted members.

Once research teams are getting ready to present these estimates to anyone besides the L2 research team, they would have another discussion with NORC about disclosure. A key disclosure risk is the identification of sub-state units, so if any estimates are specified in a way that may lead to this re-identification, NORC will ask the research team to suppress some estimates or generate more aggregated versions of these estimates.

### Merge of cluster aggregates to household data file

- **Confirm relevant unit of analysis.** Once all disclosure issues have been sorted out, NORC merges cluster aggregates to the household file, making the cluster provider aggregates an attribute of the household or children in the household, depending on the research team's relevant unit of analysis. NORC will only merge cluster aggregates to the analytic subpopulation specified by the research team. Cases not included in the analytic subpopulation will not have any information on cluster aggregates.
- **NORC posts data file in CSV format in the research team's secure project folder and notifies team by email.**

## Disclosure Guidelines for L2 NSECE Users Working with Provider Clusters

### Provider clusters that have 1 or 2 providers are subject to stricter disclosure rules.

- This applies to analyses using cluster data from the 2012 NSECE, the 2019 NSECE, and both the 2012 and 2019 NSECE.
  - **Stricter rules apply to providers of any type**, not necessarily providers in the analytic subpopulation of interest, such as providers serving children 0-5.
  - **Stricter rules involve releasing up to two aggregates for these provider clusters.**

- The L2 research team decides which aggregates they would like to have access to, presumably a selection of two of the aggregates the team has created. NORC will merge these two aggregates to all household cases in the analytic sample.
- For the research team, a key value of having these aggregates is to compare households located in provider clusters with 1 or 2 providers to households located in provider clusters that have 3 or more providers.
- In the final household-level file the research team has access to, households with 1 or 2 providers will have suppressed data on all L2 variables, except for these two aggregates.

**All weighted and unweighted counts are only released as categorical variables.**

For each of these variables, there is a maximum of 5 categories or bins. The research team may decide how to create these categories based on summary statistics that rely on Simulated Cluster IDs. The research team provides NORC syntax that generates counts in this categorical fashion.

**Generate aggregate measures.**

- NORC encourages teams to generate aggregate measures (such as means and medians) that include all providers, since proportions and dummy indicators that include all providers in the cluster (e.g., any providers with a specific characteristic) carry less disclosure risk and are more statistically robust than variables that only include a subset of providers in the provider cluster, such as means or medians for providers with particular characteristics.
- For aggregates (such as means or medians) that only use a subset of providers, values based on fewer than 3 providers will be suppressed.

**National level estimates.**

- These disclosure rules apply to teams aiming to produce household or child-level estimates at the national level.
- Additional disclosure rules apply if the L2 research team wants to carry out the analysis at other levels of geography, such as subgroups of states or single states.

**Disclosure Guidelines for L2 NSECE Users Working with 2012 and 2019 State-level Information**

- **L2 reporting guidelines for state-level analysis using only the 2012 NSECE are appropriate for users carrying out state-level analysis using only the 2019 NSECE and for users carrying out state-level analysis using both the 2012 and 2019 NSECE.**
  - As a reminder, estimates should be suppressed if they are based on unweighted counts of:
    - 50 or less for CB
    - 10 or less for WF
    - 20 or less for HH
    - 20 or less for listed HB, and
    - 10 or less for unlisted HB.

    (We know we mentioned this before, but it bears repeating!)
  - All estimates displayed should have a maximum of three significant digits.
  - Unweighted estimates should be rounded to the nearest 20.
  - For national level analysis, users need to provide evidence that each estimate includes a minimum of 3 states.
  - For state level analysis, users need to provide evidence that each estimate includes a minimum of 3 counties.
- **Even though the reporting guidelines are the same, the evidence users need to provide may be more extensive for analyses that use both 2012 and 2019 NSECE data.**

  Below we present three scenarios that discuss key disclosure challenges associated with different types of analysis users may carry out.

  **SCENARIO 1**: State level estimates that compare HHs or providers between 2012 and 2019, but no NSECE files are not linked, and no external data or policy variation are merged to NSECE.

| Examples of estimates | Example of research question users may be interested in | L2 data needed |
|---|---|---|
| **Example # 1** | What is the average wage of center-based teachers in 2012 and 2019 in a particular state? | State ID for 2012 and 2019 in CB data files |
| **Example # 2** | What is the percentage of children using center-based care in 2012 and 2019 in a particular state? | State ID for 2012 and 2019 for HH data files |
| **Example # 3** | What is the number of home-based providers with a bachelor's degree in 2012 and 2019 in a particular state? | State ID for 2012 and 2019 for listed and unlisted HB data files |

**L2 guidelines**: Users need to implement suppression and rounding. They would also need to demonstrate that estimates apply to three or more counties within the state *separately for 2012 and 2019*.

**Potential challenges**: If users are tabulating teachers' wages by type funding, and there is only one county in the state that has Early Head Start services, then providing an estimate for teachers in Early Head Start programs would disclose the presence of that county in the NSECE sample.

**SCENARIO 2**: State level estimates that compare HHs or providers because of change in policies between 2012 and 2019, and the source of policy data is external to the NSECE.

| 2012 | 2019 | |
|---|---|---|
| | **Counties in state X provide public pre-K** | **Counties in state X do *not* provide public pre-K** |
| **Counties in state X provide public pre-K** | Set of counties # 1 | Set of counties # 2 |
| **Counties in state X do *not* provide public pre-K** | Set of counties # 3 | Set of counties # 4 |

**L2 guidelines**: Users need to implement suppression and rounding. They would also need to demonstrate there are at least 3 counties in the population in each cell. In this example, if all four types of policy change were observed in the population, the user would show that there are a minimum of 12 counties that exist within this particular state exhibiting the specified policy variation.

**Potential challenge**: If one of these cells contains only one county in the state, then estimates about that change would reveal the presence of that county in the NSECE sample. So, each of the four sets of counties 1-4 would need to contain at least 3 counties, even if not all those counties are in the NSECE data.

**SCENARIO 3**: National level estimates that compare HHs or providers across states that change policies between 2012 and 2019.

| 2012 | 2019 | |
|---|---|---|
| | **States with limited public pre-K** | **States with extensive public pre-K** |
| **States with limited public pre-K** | Set of states # 1 | Set of states # 2 |
| **States with extensive public pre-K** | Set of states # 3 | Set of states # 4 |

**L2 guidelines**: Users need to implement suppression and rounding. Users would also need to demonstrate that there are at least 3 states in each cell. We anticipate this requirement may be difficult to fulfill for some rare policies or for users dealing

with change in more than one policy. In this case, users are encouraged to think about combining policies or aggregating cells.

**Potential challenge**: In the case of cells that have less than 3 states per cell, users would need to treat estimates as state-level estimates and demonstrate there is no county level information revealed in the state-estimate provided. For example, in a state where the only public pre-K is within localities rather than state-wide, inclusion of the state in this type of tabulation could reveal which counties within the state are in the NSECE data. The general question posed to users would be whether there is any information in the estimates provided that is unique to specific counties in that state.

# Appendix 7: L2 Extraction Checklist

**Checklist for L2-users**

- **File naming convention**
  - **All file names should be dated**, specifying the date (MMDDYEAR) at the end of the name of the file.
  - If file is updated in the process of receiving feedback from NORC, save the file with a new date.
- **Location of files**
  - **P:\Directory ID\ID_TEAM\NORC\Disclosure review\Files to extract" folder**
  - Place documents and files you want to access outside of the secure Citrix environment.
  - Depending on the L2 data you are working with, your folder may exist within the P:\Directory ID or P:\8939 directory.

  - **P:\Directory ID\ID_TEAM\NORC\Disclosure review\Files to review" folder**
  - Place documents to assist NORC's review process, such as the summary of files, tables with raw estimates (unsuppressed and unweighted), and potentially relevant information about prevalence of population in counties or states.
- **Necessary information to initiate review.** The research team should generate a summary file where they specify:
  - **Name of each file**
  - **Key content of the file** (e.g. table with predicted probabilities derived from a multinomial regression with Y dependent variable and Z covariates; graph in which axis X represents X and axis Y represents Y)
  - **Preferred format for extraction** (e.g. PDF, Excel, Word, text file)
  - **Population used in analysis** or criteria used in sub-setting the data (e.g., state X for listed home-based providers serving children 0-5).
    - If a single table includes different subpopulations, specify sample selection criteria of different columns

- **If any external data sources are merged into NSECE data** and characteristics of these data sources
- **NSECE files used** (e.g. household file, including community characteristics X, Y and Z; center-based file with linkages to counties)
- **Geographic reach of estimates** (e.g. state-level, national-level).

| **Document name** (for tables, specify names for Table A and B) | **Content** | **Format** | **Analytic sample** | **External data sources** | **NSECE data file used and key L2 variables** | **Geographic reach of estimates** |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

- **Create two tables.** If the team wants to extract tables (in any kind of format –PDF, Excel, etc.), create two versions of each table:

  - **Table A – P:\Directory ID\ID_TEAM\NORC\Disclosure review\Files to extract**

  Table A will only include estimates that are meant to be distributed to users *other than* L2 NSECE researchers (e.g. presentation at a professional conference, journal article submission, and grant report).

    - Table A will include properly rounded and suppressed estimates, and it will display (rounded and suppressed) unweighted estimates whenever it's appropriate to make those estimates publicly available.
    - The identification of geographic units will need to be redacted.

  - **Table B – P:\Directory ID\ID_TEAM\NORC\Disclosure review\Files to review**

  Table B will include unweighted case counts for every estimate associated with each Table A.

    - Table B should also include unrounded, unsuppressed, and unweighted case counts for all weighted estimates and unweighted cell counts the research team aims to circulate.
    - NORC will use these unsuppressed, unrounded case counts to verify that disclosure guidelines on suppression were properly implemented.

- **Rounding of estimates.**
  - Standard rounding is 3 significant digits.
- **Suppression.**
  - Suppression of weighted estimates is based on the number of unweighted cases in each cell. The definition of a small cell varies across NSECE survey data, as specified in the table below.

| | Suppression of weighted and unweighted estimates based on specific unweighted cell size |
|---|---|
| **Center-based** | 50 or less |
| **Workforce** | 10 or less |
| **Household** | 20 or less |
| **Home-based Listed** | 20 or less |
| **Home-based Unlisted** | 10 or less |

- **Unweighted estimates.**
  - Standard rounding is nearest 20.
- **Do not mention specific sub-state units.**
  - At no point there should be any mention of specific sub-state units (e.g., City of Chicago, rural counties in Kansas, cities with Head Start programs in Oregon, etc.).
  - This includes tables, text, data visualizations, and documentation.
- **Verify at least 3 states contributed to estimate.**
  - For each estimate generated using L2 variables provide documentation that verifies that at least 3 states (DC can be considered a state) could have contributed to that estimate
    - If the estimate refers to the effect of a specific policy, team must demonstrate that at least 3 states have that same policy for the same population included in the analysis
  - If the analysis is at the state level, provide information that verifies that at least 3 counties within each state could have contributed to that estimate
    - E.G.: If the estimate refers to the timing of a specific policy across counties within a state, team must demonstrate that at least 3 counties in each state shared the same timing.

- Note that states or counties refer to the population, not necessarily to the NSECE sample.
- Most teams will only have access to some states or counties, so it is likely teams will rely on data other than the NSECE to provide this evidence.

- **Round and suppress all cell date before extraction.**
  - Whenever the team wants to extract an Excel file, make sure the actual figures in the cells are rounded and suppressed, not just what is visible when the document is printed.
- **Researchers must comply with these minimum requirements.**
  - Researchers are welcome to exceed these standards, for example, by reporting only 2 significant digits, suppressing cells that exceed the minimum cell size requirements, etc.

# Appendix 8: Disclosure Guidelines for L2 NSECE Users Working with Provider Clusters

**Provider clusters that have 1 or 2 providers are subject to stricter disclosure rules.**

This applies to analyses using cluster data from the 2012 NSECE, the 2019 NSECE, and both the 2012 and 2019 NSECE.

This applies to providers of any type, not necessarily providers in the analytic subpopulation of interest, such as providers serving children 0-5. Stricter rules involve releasing up to two aggregates for these provider clusters.

- **Decide on aggregates.** The L2 research team decides which aggregates they would like to have access to, presumably a selection of two of the aggregates the team has created.
- NORC will merge these two aggregates to all household cases in the analytic sample
  - For the research team, a key value of having these aggregates is to compare households located in provider clusters with 1 or 2 providers to households located in provider clusters that have 3 or more providers.
  - In the final household-level file the research team has access to, households with 1 or 2 providers will have suppressed data on all L2 variables, except for these two aggregates.

**All weighted and unweighted counts are only released as categorical variables.** For each of these variables, there is a maximum of 5 categories or bins. The research team may decide how to create these categories based on summary statistics that rely on Simulated Cluster IDs.

The research team provides NORC syntax that generates counts in this categorical fashion.

- **Research teams should generate aggregate measures that include all providers.** NORC encourages teams to generate aggregate measures (such as means and medians) that include all providers because they will have a lower risk of disclosure. For aggregates (such as means or medians) that only use a subset of providers, values based on fewer than 3 providers will be suppressed.

| Measures with Less Disclosure Risk | Measures with More Disclosure Risk |
|---|---|
| Proportions and dummy indicators that include all providers in the providers cluster (e.g. any providers with a specific characteristic) carry less disclosure risk and are more statistically robust | Variables that only include a subset of providers in the provider cluster, such as means or medians for providers with particular characteristics carry more disclosure risk. |

**Clarification on disclosure rules for analysis of household, child-level, and subgroups.** These disclosure rules apply to teams aiming to produce household or child-level estimates at the national level. Additional disclosure rules apply if the L2 team wants to carry out the analysis at other levels of geography, such as subgroups of states or single states.

### Helpful Hints on Protecting Against Disclosure in NSECE Analyses

Researchers may choose how they will comply with the requirements presented in the **Disclosure Requirements for NSECE Research Using L2 Data** section. If an estimate violates these reporting guidelines, researchers may choose to either suppress this particular estimate, and potentially some associated estimates, or alternatively redefine the analysis so that estimates produced do not present any disclosure risk.

These hints are not intended to guide researchers about how to redefine their analyses. Researchers may use any strategy they deem appropriate to redefine the analysis, including specifying a different analytic sample, recoding variables into more aggregated categories, or removing variables from the analysis.

**Below are some ideas for adaptations, if necessary.**
**Researchers may want to…**

- Run analysis with original data but consider collapsing categories (or creating meaningful classifications) to report
- Consider rounding techniques to report data
- Truncate some extreme values in analysis
- Check if there is enough variability in the metrics used to avoid reporting on unique combinations.
  - For instance, researchers may calculate interquartile range, coefficient of variation (ratio of the standard deviation to the mean)
- Consider adding random noise without changing the distribution
- Randomly eliminate some data points and use imputation techniques to recover data based on remainder of data
- Limit geographic detail in combination with demographic characteristics
- Use robust techniques that make less likely the identification of geographic units
  - For instance, adopting a random effects regression model instead of a fixed effects regression model (which typically uses dummy variables to account for geographic units)
- Be aware that some multivariate techniques (factor analysis, latent class analysis, cluster analysis) may yield unique solutions that can lead to inferential disclosure