

Secondary Data Analysis: Ethics & Resource Toolkit

TEACHERS COLLEGE INSTITUTIONAL REVIEW BOARD

By Bemsimbom (Bemsi) Nkuo
TC IRB Research Writing & Ethics Internship
Spring 2025

Toolkit and TC NEXT: Career Enhancement

This toolkit aims to arm students in the clinical psychology (Psychology in Education) MA program with resources they could use to find publicly available data, learn data analysis skills for free, and apply to present at annual conferences in the field. Despite being tailored to the clinical psychology population, it could prove useful to anyone interested in research.

Secondary Data Analysis and the IRB

Exempt Category 4 refers to certain types of secondary research involving existing identifiable private information or biospecimens, provided specific conditions are met, such as the data being publicly available, recorded by the researcher in a way that subjects cannot be identified, or use being regulated by HIPAA or another legally authorized entity.. Please refer to the [Get to Know a Review Category: Exempt Category 4 - Secondary Research \(Identifiable Private Information\)](#) blog for more information on the classification of “secondary research,” IRB application requirements, data security, and obtaining publicly available data.

For this toolkit, here are some frequently asked questions by master-level researchers regarding secondary research:

1. **FAQ: What kinds of secondary data analysis require IRB review as human subjects research, and which do not?**
 - a. This depends on the identifiability of the data. Anonymized datasets, when properly stripped of all identifiers and when re-identification is not reasonably possible, may not qualify as human subjects research under the Common Rule. However, researchers should still consult with the IRB, as some de-identified datasets may retain indirect identifiers or be subject to data use agreements requiring oversight. Visit this resource: “[What is the difference between anonymization and de-identification](#)” for more detailed information.
2. **FAQ: If my project doesn't need IRB approval, in what cases should I still submit it to the IRB?**
 - a. When submitting research work for publishing or presentations, some conferences or journals may require a statement from your IRB that states that your project does not meet the definition of human subjects research.. The TC IRB cannot issue a retrospective determination of exemption or non-human subjects status. If you think you may need documentation for future publication or conference submission, it's best to submit an NHSR application before conducting your analysis.
 - b. Please refer to our [What is Human Subjects Research](#) Guide for detailed examples of what studies are considered human subjects research and which are not.
3. **FAQ: How do I determine whether my proposal meets the definition of research with human subjects?**
 - a. For a formal determination of whether your proposal meets the definition of research with human subjects, please complete and submit for review the Non-Human Subjects Research (NHSR) Determination application via Mentor IRB.
4. **Are there any other requirements beyond IRB review I should be aware of?**
 - a. When working with secondary datasets, always check whether the data use is governed by a Data Use Agreement (DUA), Terms of Use, or licensing restrictions that may impose additional requirements beyond IRB review.
5. **Some Useful Resources:**
 - a. [Guidance On Secondary Data Analysis For Existing Datasets](#)

- b. [IRB Guidance For Secondary Data Analysis](#)
- c. [Research Involving The Secondary Use Of Existing Data](#)

Free data sources and their descriptions

Publicly available secondary data sources can typically be accessed for free online. Below is a list of popular websites where you can find free completed datasets.

1. Statistical Databases with Downloadable Datasets

- a. [Statista](#): Find statistics and consumer survey results from over 18,000 sources on over 60,000 topics. Download data sets in a choice of formats.
- b. [Kaggle](#): Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.

2. Major Academic Repositories

- a. [Dryad Digital Repository](#): Dryad is a nonprofit membership organization that is committed to making data available for research and educational reuse now and into the future.
- b. [Harvard Dataverse](#): A free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data. (Includes the data from other Dataverse repositories, e.g., UCSD)
- c. [Zenodo](#): To ensure no one is left behind through lack of access to the necessary tools and resources, Zenodo makes the sharing, curation, and publication of data and software a reality for all researchers.

3. Additional Open Data Repositories

- a. [Figshare](#): Figshare is a repository where users can make all of their research outputs available in a citable, shareable, and discoverable manner.
- b. [Google Dataset Search](#): Similar to how Google Scholar works, Dataset Search lets you find datasets wherever they're hosted, whether it's a publisher's site, a digital library, or an author's webpage.
- c. [ICPSR \(endorsed by APA\)](#): ICPSR maintains a data archive of more than 250,000 files of research in the social and behavioral sciences.
- d. [openAFRICA](#): Searchable datasets maintained by Code for Africa, with support from the World Bank and AWS. "All the data on openAFRICA is intended to be actionable. The data is meant to help you change the world."
- e. [Registry of Research Data Repositories](#): This offers a way to search over 2,000 worldwide data repositories.

4. Local datasets

- a. [City of Seattle](#): The Open Data Program makes the data generated by the City of Seattle openly available to the public for the purpose of increasing the quality of life for our residents; increasing transparency, accountability, and comparability; promoting economic development and research; and improving internal performance management.
- b. [King County](#): Welcome to King County's Open Data portal, where you can discover, analyze, and connect with publicly available datasets published by King County agencies.

5. Additional datasets recommended by Seattle Pacific University (SPU) faculty

- a. [Curate Science](#): A database of studies published in the sciences (and particularly social sciences/psychology), many with open data. Search by topic (e.g., personality, romantic relationships, etc.) and find papers with accompanying datasets.
- b. [Iris Data Set](#): The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of

machine learning algorithms.

The Iris Data Set contains three classes of 50 instances each, where each class refers to a type of iris plant.

- c. [Kaggle - Titanic Survivor Data](#): This classic dataset includes training and testing data (to practice building models and seeing if they cross-validate), trying to predict whether people survived the Titanic disaster in 1912.
- d. [Kaggle - World Happiness Data](#): This is a classic dataset for analysis. Over several years, it looks at national happiness and other data by country. It contains categorical and continuous data and has data nested within countries over time.
- e. [Linear Regression Data Sets for Machine Learning](#): Here, there are open linear regression datasets for free downloading. Some of the datasets include sample regression tasks to complete with the data.
- f. [Project Implicit - Open Data for Implicit Racial Bias and More](#): Project Implicit was founded to foster the dissemination and application of implicit social cognition. The site contains many gigabytes of datasets looking at IAT data in psychology (fast, automatic learned associations with race, gender, etc.) and is an important repository of datasets for research in psychology.

Source/Reference for Data Repositories: [Seattle Pacific Library](#)

Checking the Credibility of Secondary Data Sources

1. [How to know if a source is credible](#)
2. [The advantages and limitations of secondary data](#)
3. [A guide to secondary data analysis](#)

Free Data Analysis Learning Sites

After obtaining your dataset, you may need resources to help analyze it. Here are some platforms and free courses you can take to enhance your skills.

1. Columbia University's available resources
 - o [LinkedIn Learning](#): Available only to current Columbia/TC students
2. [10 free data analysis courses](#)
3. [Data Analysis Courses and Certificates on Coursera](#)
4. [Data Analysis Classes at Harvard University](#)
5. [Columbia University Data Analysis - a collaboration with edX](#)
6. [MIT OpenCourseWare courses on Statistics](#)

Annual conferences in the field of psychology and neuroscience & Provost Grants.

Here is a list of annual conferences where you can submit your abstracts after completing your data analysis. After you select the conference(s) you would like to attend, you can visit the [Provost Grants for Professional Development and Conference Presentations website](#) to learn more about the eligibility and grant application requirements.

1. [Psychonomic Society](#): The Psychonomic Society holds its Annual Meeting in different locations throughout North America each year. These gatherings bring together researchers and professionals in cognitive psychology and related fields to share their latest findings and network with colleagues. The locations and dates for upcoming Annual Meetings are determined in advance to allow attendees to plan their travel and participation.
2. [Cognitive Neuroscience Society](#): Each year, the Cognitive Neuroscience Society holds an annual meeting in the spring. The purpose of the meeting is to bring together researchers from around the world to share the

latest studies in cognitive neuroscience. This 4-day event is filled with invited symposia, posters, awards, a keynote address, and most importantly, the opportunity to connect with colleagues.

3. [American Psychological Association \(APA\)](#): Each year, we bring the APA Convention to a different U.S. city so psychologists across the country can access all the events that are offered (and get a little sightseeing in while they're at it!).
4. [International Positive Psychology Association \(IPPA\) World Congress](#): The IPPA World Congress is a place to share (and learn from) findings, insights, and information, through symposia, panels, poster sessions, galleries, workshops, ignite sessions, keynotes, and roundtable discussions that focus on leading-edge research, as well as applications of the fascinating science across a variety of fields, industries, and areas – from business and workplaces to education, to medicine and the greater healthcare sector, to the impact on students and wellbeing, and other areas such as belonging, spirituality, and meaning.
5. [Association for Psychological Science \(APS\) Annual Convention](#): The APS Annual Convention is an international conference dedicated to psychological science, offering a platform for invited and submitted presentations across various areas of the field. These include disciplines such as Biological/Neuroscience, Cognitive, Industrial/Organizational, Methodology, Social, Clinical Science, Developmental, General Psychology, Personality/Emotion, and Teaching. This diverse content showcases the breadth of psychological research and practice, bringing together scholars, practitioners, and educators from around the world to discuss advancements and innovations in these areas.
6. [International Conference on Applied Psychology and Behavioral Science \(ICAPBS\)](#): The International Conference on Applied Psychology and Behavioral Sciences aims to bring together leading academic scientists, researchers, and research scholars to exchange and share their experiences and research results on all aspects of Applied Psychology and Behavioral Sciences. It also provides a premier interdisciplinary platform for researchers, practitioners, and educators to present and discuss the most recent innovations, trends, and concerns, as well as practical challenges encountered and solutions adopted in the fields of Applied Psychology and Behavioral Sciences.
7. [Society of Pediatric Psychology Annual Conference \(SPPAC\)](#): For the last decade, the annual meetings for the Society of Pediatric Psychology have centered on defining the ever-changing state of the art of our field. This has required intention and willingness to adapt to rapidly shifting healthcare and political landscapes as we work together to trailblaze a path forward for our science and clinical practice. As a profession, we are learning to stretch our skill sets beyond the traditional confines of our roles as clinician-scientists to meet the call to be advocates and change agents across science, practice, training, and policy. SPPAC 2024 provided the space to come together and reflect on the work we have done as a field, pushing us to collectively define the path forward. Building off the energy and momentum from SPPAC 2024, we are excited to announce the conference theme: Moving Forward with Hope: Advancing Health Equity in Pediatric Psychology through Science, Practice, and Advocacy.
8. [The Society of Industrial and Organizational Psychology \(SIOP\) Annual Conference](#): The 2025 SIOP Annual Conference, hosted at the Colorado Convention Center in the vibrant city of Denver, Colorado, promises to be a pinnacle event in the field. Renowned for its commitment to advancing the science and practice of industrial-organizational psychology, SIOP brings together leading experts, researchers, and practitioners from around the globe.
9. [Asian Conference on Psychology & Behavioral Science](#): The ACP Conferences have proven to be a great opportunity for engaging in interdisciplinary dialogue, speaking to scholars, and learning from other experts from around the world and from a variety of academic disciplines. The interdisciplinary and international focus of the conference draws world-class speakers and keeps people coming back year after year.
10. [International Conference on Psychiatry and Psychology \(ICP\)](#): The International Conference on Psychiatry and Psychology aims to bring together leading academic scientists, researchers, and research scholars to exchange and share their experiences and research results on all aspects of Psychiatry and Psychology. It also

provides a premier interdisciplinary platform for researchers, practitioners, and educators to present and discuss the most recent innovations, trends, and concerns, as well as practical challenges encountered and solutions adopted in the fields of Psychiatry and Psychology.

11. [International Conference on Developmental and Adolescence](#): The International Conference on Developmental Psychology and Adolescence aims to bring together leading academic scientists, researchers, and scholars to exchange and share their experiences and research results on all aspects of Developmental Psychology and Adolescence. It also provides a premier interdisciplinary platform for researchers, practitioners, and educators to present and discuss the most recent innovations, trends, and concerns, as well as practical challenges encountered and solutions adopted in the fields of Developmental Psychology and Adolescence.

Tips for first-time presenters (with a focus on secondary data analysis)

Presenting research for the first time, especially work based on secondary data, can feel both exciting and overwhelming. Here are some tips to help you navigate the journey from dataset to conference presentation:

1. **Start by understanding your dataset:** Before you even think about writing an abstract, make sure you've explored your dataset thoroughly. Know what each variable means, how the data were collected, and any limitations. If you're using a public dataset, cite the source clearly and check whether any data use agreements or acknowledgments are required.
2. **Narrow your research question:** Secondary datasets can be large and overwhelming. Focus your question on something manageable, interesting, and aligned with the dataset's structure. Think about what kind of story your data can realistically tell.
3. **Run a feasibility check:** Do you have enough cases for your analysis? Are the variables of interest complete? Check for missing data or strange values early on to avoid surprises later in the process.
4. **Draft a preliminary analysis plan:** Even if your methods will evolve, having a clear analysis plan will help you stay organized. Know what statistical tests or techniques you'll use and why they make sense for your research question and data type.
5. **Document everything:** Track your data cleaning, coding decisions, and analytical steps. This will help you later when writing your abstract and answering audience questions about your process.
6. **Only then begin drafting your abstract:** Once you've completed your analysis and feel confident about your findings, start writing your abstract. Make it clear that your project is a secondary analysis (e.g., "This study conducted a secondary analysis of publicly available data from..."), and highlight the relevance of your research question, methods, and key insights.
7. **Choose the right format and conference:** Poster presentations are often ideal for first-time presenters using secondary data; they allow for more interactive feedback and are less high-pressure than oral talks. Look for conferences that welcome student work or have dedicated categories for early-career researchers.
8. **Prepare, polish, and practice:** Whether you're creating a slide deck or a poster, aim for clarity and simplicity. Practice explaining your work to someone unfamiliar with your topic (like a friend outside of psychology) to make sure your message is clear.
9. **Be ready to talk about ethics and IRB:** Even with secondary data, conference attendees may ask about data privacy, consent, and IRB status. Be prepared to explain whether your work qualified for exemption and how you ensured ethical use of the dataset.
10. **Own your expertise:** You've spent time working with the data, so you are the expert on your piece of the puzzle. Conferences are a great way to get feedback, ask questions, and grow your confidence as a researcher!