



CRITICAL EVALUATION OF LARGESCALE SECONDARY ANALYSES TO AID RESPONSIBLE RESEARCH USE AND EQUITABLE DECISION MAKING

Created by:

Tara Kulkarni, Mollie R. Weeks, and
Amanda L. Sullivan

Midwest and Plains Equity Assistance Center

August 2021





About the Centers

Great Lakes Equity Center (Center) is an educational research and service center located in Indiana University's School of Education at IUPUI. The Center engages in equity-focused technical assistance and related research with educational and community agencies focused on systemic improvements to serve all learners with particular focus on educational access, participation and outcomes for those who have been historically marginalized. Midwest and Plains Equity Assistance Center is a project of the Center and provides technical assistance related to educational equity based on student race, national origin, sex, and religion at no cost to public educational agencies throughout its 13-state region in the Midwest and Plains.



Overview

Largescale secondary data analysis—that is, analysis of a very large number of cases—is a popular and often influential research approach in education, public health, and numerous other fields. Educational professionals frequently encounter headlines in popular media outlets that promote contentious or sensational findings from studies using large datasets. From the outside, these studies are appealing given their large numbers of participants and sophisticated statistical methods. It is no wonder that there have been several secondary studies that have captured widespread attention among educators and the general public. For example, those in special education might remember controversial disproportionality research that garnered editorials in several national newspapers, interviews on cable news, and a flurry of critiques of the study's theoretical and statistical appropriateness (for discussion, see Skiba et al., 2016; Sullivan & Proctor, 2016; 2016 special issue of the journal *Multiple Voices for Ethnically Diverse Exceptional Learners*). The controversy was spurred by claims that students from racially minoritized backgrounds were *under*-identified for special education while other scholars emphasize the complex history and sociopolitical issues surrounding the education of minoritized students, particularly at the intersections of race, culture, and dis/ability. However, since critiques were primarily in academic spheres, most of the public had no access to the scholarly discourse and may have been familiar primarily or solely with the headlines. This has real world implications, as policy remediating racial disproportionality in special education has faced recent setbacks with controversial findings often cited (*Council of Parent Attorneys and Advocates v. DeVos*, 2019; Sullivan & Osher, 2019).

Largescale data analysis is a promising tool for exploring educational questions that requires responsible use on the part of both researchers and readers to ensure that study's features, including limitations, are appropriately accounted for to aid in decision-



Overview (cont.)

making that is justifiable and equitable. Largescale data analysis is often alluring to researchers and research consumers because huge numbers of participants, often in the thousands and even hundreds of thousands, and variables allow for investigations often not otherwise possible—from examination of outcomes for hard to study populations (e.g., groups comprising a small proportion of a larger population) to sophisticated statistical analyses of complex phenomena or contexts not amenable to experimental investigation (e.g., quasi-experimental analyses of contexts not amenable to experimental research; Sullivan & Field, 2013). In addition, this type of research can also afford estimates of broader contexts or populations (e.g., nationally representative samples, statewide data systems), making them attractive options for studying educational issues of broad social significance.

Yet such studies are not without limitations; namely, secondary researchers' lack of control over the study design or measures, which necessitates compromises in using variables less precise or appropriate for research questions than might have been chosen if a researcher had the opportunity to do so. In addition, largescale secondary data analysis often involves survey (e.g., phone interviews or paper questionnaires) or observational data that do not allow for testing of cause-and-effect relationships (e.g., Does A cause B? Does X improve Y?). Instead, secondary researchers are often limited to describing trends (e.g., group A is increasing more than group B, group C has better outcome than group D) or relationships (e.g., increased X is associated with increased Y) without being able to test or interpret why or how such findings occur. Rarely, if ever, do studies allow for definitive claims (e.g., 'proving' or 'disproving' something) or causal inferences, so skepticism is warranted when definitive language is used to describe study findings (e.g., use of 'is' versus 'may'). To aid in critical consumption of secondary research, the checklist in Table 1 can be used to evaluate a largescale secondary data analysis, gauging the quality of reporting, regardless of advanced statistical knowledge.

When encountering largescale research—or any study—that claims to disprove existing bodies of knowledge, keep in mind that *extraordinary claims require extraordinary evidence* (Sagan, 1980). Science typically evolves in small increments, and often includes contradiction of past work, but dismissing decades of evidence often relies on several rigorous studies from heterogeneous data sources and researchers (Gough & Richardson, 2018).

Critically evaluating a secondary data analysis can also mean evaluating the utility of the statistical analysis for real world contexts. Take, for example, disproportionality researchers who attempt to isolate the effect of a single factor through statistical



Overview (cont.)

models in which “all else is equal.” Critical consumers must question what conclusions can be drawn from research that artificially creates equitable environments when students’ lived experiences are influenced by systemic inequity (Sullivan et al., 2020; Whitford & Carrero, 2019). For example, can we really parse out the effects of the interaction between race and socioeconomic status within the U.S. given the historical context of segregation and systematic discrimination (Cohen et al., 2015)? Asking these questions is pivotal in evaluating the utility of the results of a secondary research, especially when generalizing findings to aid equitable decision making in schools. Thus, it is important to approach sensational headlines with healthy skepticism, remembering that every study has limitations, and no single study defines a field.



About This Tool

To aid in this process, this tool contains two parts. The first part (Table 1) is a checklist that can be applied to a review of secondary research studies or reports. The second part (Table 2) presents a summary of different statistical analysis tools, and what research questions can be answered when using different approaches commonly encountered in secondary analysis. This tool is intended to support when critically reading a research study or report. When engaging with more general research information, such as news stories or reports shared via social media, guiding considerations include:

- Does the story mention national, nationally-representative, or large sample or big data? In these cases, the study likely entailed largescale secondary analysis.
- Are the findings used to justify changes in policy or practice? This necessitates careful consideration of the quality and applicability of the study.
- Do the recommendations stem directly from the study completed or was there inferences beyond the study made to link to potential implications?
- Are the findings described in tempered or definitive terms? The latter is a sign that the findings may not have been interpreted judiciously.

Once you have located the primary source (e.g., study article or report), you can apply the checklist in Table 1. A preponderance of affirmative responses in each domain of the checklist suggest the study can be considered trustworthy. Conversely, studies missing extensive information in multiple domains may reflect poor reporting or poor study execution and should be interpreted with caution. Studies with few affirmative responses for the method, analyses, or interpretations may be especially problematic.



About this Tool (cont.)

This tool was adapted from Kulkarni et al. (2020) and Sullivan et al., 2020, as well as drawing on Anderson et al. (2011), Hahs-Vaughn et al. (2011), and Trzesniewski et al. (2011). Table 2 summarizes common methods and statistical concepts encountered in largescale secondary research distilled from Howell (2010), Payne and Payne (2011), Salkind (2010), and Sullivan and Fein (2012).

Table 1: Checklist for Critically Evaluating Largescale Secondary Research

Study Feature	Yes	No	Can't Tell	N/A	Notes
Abstract/Front Matter					
Is the study purpose <i>and</i> rationale clearly stated (i.e., what questions they plan to answer and why those questions are important)?					
Is the data source and sample explicitly described?					
Are key variables identified?					
Are the analyses and findings described?					
Do the authors identify whether they received funding for the study and whether there were any conflicts of interest (mark "yes" if reported regardless of whether the study received funding or conflicts existed)?					
Study Description					
Do the authors state why they undertook the study?					
Is related research described from a balanced perspective, including both supporting studies and conflicting research?					
Do the authors avoid causal language?					
Is the authors' theoretical and conceptual basis for the study clearly described?					
Is the relationship to educational policy, practice, or systems appropriately described?					
Are the research questions clearly stated?					
If the authors seek to make causal inferences, do they explain how the research design or method allows such claims?					

Table 1: Checklist for Critically Evaluating Largescale Secondary Research (cont.)

Study Feature	Yes	No	Can't Tell	N/A	Notes
Method					
Do the authors clearly identify the data source?					
Do the authors clearly describe how they accessed the data?					
Do the authors note whether the study was approved by an Institutional Review Board and other relevant organizational bodies (e.g., agency granting data access)?					
Do the authors describe when and how the data were originally collected?					
Do the authors explain why they used the data source for their purposes?					
Do the authors identify how they selected participants from the main dataset?					
Do tables clearly provide descriptive information on participant characteristics and subgroups?					
Does the analytic sample match the population of interest as described in the introduction or purpose?					
Do the authors describe how and why they selected their study variables from the main dataset?					
Do the authors reference original study materials (e.g., technical reports, code) in explaining their choice of variables?					
Do the authors provide information to support the reliability and validity of their selected variables?					

Table 1: Checklist for Critically Evaluating Largescale Secondary Research (cont.)

Study Feature	Yes	No	Can't Tell	N/A	Notes
Method (cont.)					
Do the authors describe each variable used in their analysis (e.g., how collected in original study, any manipulation of the variable)?					
Do tables clearly provide descriptive information on all variables used in the analysis?					
Have the authors included confounding variables that could be related to the outcome in the analysis?					
Do the key measures match the purpose?					
Analyses and Findings					
Do the authors explain how they dealt with missing data, including how they determined the nature of missing data in their study?					
Do the authors describe how small frequencies were dropped or aggregated?					
Do the authors explain which analyses were used to answer their aims/research questions <i>and</i> why the analyses were used?					
Do the authors explain the statistical assumptions that had to be met to conduct their analyses?					
Do the authors report how they completed their analyses (i.e., included steps to aid replication, named statistical software, programs, etc.)?					
If the data source was described as a complex sample, do the authors describe how weighting or adjustment for sampling was applied?					
Are the results well described, including effect sizes and confidence intervals?					

Table 1: Checklist for Critically Evaluating Largescale Secondary Research (cont.)

Study Feature	Yes	No	Can't Tell	N/A	Notes
Interpretations					
Do the authors interpret their findings with an appropriate level of caution?					
Do the authors describe how their findings compare to other studies or theory?					
Do the authors note the limitations of their data source (e.g., age of data, imperfect measures, missing data, small subsamples)?					
Do the authors describe the limitations of their analyses and any caveats for drawing conclusions?					
Do the authors describe how the findings can generalize to specific populations or contexts?					
Is this context or population applicable to your setting?					
Do the authors tell how their findings might apply to policy or practice in ways that are directly linked to the information in their study?					
Additional Notes					

Table 2: Common Statistical Concepts in Largescale Secondary Research

Does the research design support correlational or causal inference?

Correlation is a statistical measure that describes the size and direction of a relationship or association between two or more variables.

- Example research question: Is higher socio-economic status associated with improved mental health outcomes?
- Example research designs: correlational, observational, survey.
- Appropriate descriptors for correlational inference: associated, related, relationship, increased probability, decreased probability, higher likelihood, lower likelihood

Causation indicates that one event is the result of the occurrence of the other event.

- Example research question: Does ADHD medication result in reduced inattentive symptoms in adolescents?
- Example research designs: Experimental studies (e.g., randomized clinical trials), quasi experimental studies.
- Appropriate words to describe causal inference: caused, due to, resulted in, increased, decreased.

How much does a finding matter? It depends on the effects found.

- An **effect size** quantifies the difference between two groups. It is the standardized mean difference between the two groups.
 - ◊ Examples of effect sizes: Cohen's d , R^2 , odd ratio, relative risk ratio
- Effect is different from statistical significance, which only tells how likely the result is due to chance alone. Statistical significance is often reported as p values, which are sensitive to sample size, making it easier for largescale research to report statistically significant results even when effects are very small and likely of little practical significance. An effect size allows us to describe the size or magnitude of the difference, which is essential to determining its practical importance.

Common Statistical Methods

<i>Types of Analyses</i>	<i>Use</i>	<i>Example</i>
1-Way ANOVA	Tests the differences in a scale-level dependent variable by a nominal-level variable having 2 or more categories.	Are there differences in IQ scores by grade (freshman, sophomore, junior, and senior)?

Table 2: Common Statistical Concepts in Largescale Secondary Research (cont.)

<i>Common Statistical Methods (cont.)</i>		
<i>Types of Analyses</i>	<i>Use</i>	<i>Example</i>
2-Way (Factorial) ANOVA	Used to examine the interaction between the two independent variables. Interactions indicate that differences are not uniform across all categories of the independent variables.	Are there differences in IQ scores by grade (freshman, sophomore, junior, and senior) and sex (male, female)? We may find that females (or males) have a higher IQ overall, but the effect is not equal across grades.
Bivariate Regression	Tests how a change in the independent variable predicts the level of change in the dependent variable.	To what extent does ACT score predict college GPA?
Multiple Regression	Tests how a change in two or more independent variables (or the combined effect of independent variables) predict the level of change in the dependent variable.	To what extent does GRE score and undergraduate GPA predict the number of years to complete a doctoral program?
Logistic Regression	Tests how a change in one or more independent variables (or the combined effect of independent variables) predict the level of change in a categorical/dichotomized dependent variable.	Is there an association between externalizing behavior and a diagnosis of anxiety among males?



Key Terms

Secondary data analysis: Research using data collected by primary researchers and made available to others to use for research

Large scale data: A substantially large data source. In education, this often refers to studies or datasets, with thousands (and sometime hundreds of thousands) of participants or case, from school districts, state departments, federal agencies, or federally funded data collection efforts. Common sources include statewide testing data or enrollment data from school, district, or state information systems; federal datasets from the U. S. Department of Education's (USDOE) National Center for Education Statistics, Office of Special Education, and Office for Civil Rights; and USDOE funded studies such as the Early Childhood Longitudinal Studies and National Longitudinal Transition Studies.

Effect size: A measure of the magnitude of a difference between groups which is often used to communicate the practical significance of a study.

Disproportionality: The extent to which group membership (e.g., race, sex, language status, disability status, etc.) affects the probability of receiving a certain outcome (e.g., special education eligibility, suspension).



References

- Anderson, J. P., Prause, J., & Silver, R. C. (2011). A step-by-step guide to using secondary data for psychological research. *Social and Personality Psychology Compass*, 5(1), 56-75. doi:10.1111/j.1751-9004.2010.00329.x
- Cohen, D. R., Burns, M. K., Riley-Tillman, C., & Hosp, J. L. (2015). Are minority students under- or overrepresented in special education? *Communiqué*, 44(2). *Council of Parent Attorneys and Advocates v. Betsy DeVos*. 18-cv-1636 (D.C. Cir., 2019). Retrieved from https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2018cv1636-31
- Gough, D., & Richardson, M. (2018). Systematic reviews. In P. Brough (Ed.), *Advanced research methods for applied psychology: Design, analysis, and reporting* (pp. 75-87). New York, NY: Routledge.
- Hahs-Vaughn, D. L., McWayne, C. M., Bulotsky-Shearer, R. J., Wen, X., & Faria, A. (2011). Methodological considerations in using complex survey data: An applied example with the Head Start Family and Child Experiences survey. *Evaluation Review*, 35, 269-303. doi:10.1177/0193841X1141207
- Howell, D. C. (2010). *Statistical methods for psychology* (7th edition). Boston, MA: Wadsworth Cengage Learning.
- Kulkarni, T., Weeks, M., & Sullivan, A. L. (2020). Large-scale secondary data analysis—Part 2: For practitioners. *NASP Communiqué*, 48(6), 4-6.
- Payne, G., & Payne, J. (2011). *Key concepts in social research*. Newbury Park, CA: SAGE Publications, Ltd. <https://dx.doi.org/10.4135/9781849209397>
- Sagan, C. (1980). *Cosmos* (1st ed.). New York, NY: Random House.
- Salkind, N. J. (Ed.). (2010). *Encyclopedia of research design*. Newbury Park, CA: SAGE Publications, Inc. <https://dx.doi.org/10.4135/9781412961288.n31>
- Skiba, R. J., Artiles, A. J., Kozleski, E. B., Losen, D. J., & Harry, E. G. (2016). Risk and consequences of oversimplifying educational inequities: A response to Morgan et al. (2015). *Educational Researchers*, 45, 221-225. <https://doi.org/10.3102/0013189X16644606>
- Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and mathematics skills?: A propensity score weighting analysis. *Journal of School Psychology*, 51(2), 243-260. <https://doi.org/10.1016/j.jsp.2012.12.004>
- Sullivan, A. L., & Osher, D. (2019). IDEA's double bind: A synthesis of disproportionality policy interpretations. *Exceptional Children*, 85(4), 395-412. <https://doi.org/10.1177/0014402918818047>
- Sullivan, A. L., & Proctor, S. (2016). The shield or the sword?: Revisiting the debate on racial disproportionality in special education and implications for school psychologists. *School Psychology Forum*, 10, 278-288.



References

- Sullivan, A. L., Weeks, M., Kulkarni, T., & Nguyen, T. (2020). Large-scale secondary data analysis—Part 1: For researchers. *NASP Communiqué*, 48(5), 16-19. (* = equal contribution)
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3444174/>
- Trzesniewski, K. H., Donnellan, M. B., & Lucas, R. E. (Eds.). (2011). *Secondary data analysis: An introduction for psychologists*. Washington, DC: American Psychological Association.
- Whitford, D. K., & Carrero, K. M. (2019). Divergent discourse in disproportionality research: A response to Kauffman and Anastasiou (2019). *Journal of Disability Policy Studies*, 30, 91-104. <https://doi.org/10.1177/1044207318822264>

Resources

- **How Statistics Can Lie** - <https://www.youtube.com/watch?v=TiaVGBEFiZs>
- **Effect Size** - <https://www.youtube.com/watch?v=6uYNVCy-8NA>
- **Thresholds for Interpreting Effect Sizes** - http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html
- Weeks, M., Kulkarni, T., Kim, J., & Sullivan, A. L. (2020). Large-scale secondary data analysis—Part 3: Sense and skepticism when engaging with popular media. *NASP Communiqué*, 48(7), 28-30.
- Kulkarni, T., Weeks, M., & Sullivan, A. L. (2020). Large-scale secondary data analysis—Part 2: For practitioners. *NASP Communiqué*, 48(6), 4-6.

IMPACT:

Educate, Engage, Empower — For Equity



Great Lakes Equity Center

902 West New York St.
Indianapolis, IN 46202
317-278-3493 - glec@iupui.edu
glec.education.iupui.edu



IUPUI

SCHOOL OF EDUCATION

IUPUI School of Education 902

West New York St.
Indianapolis, IN 46202
317-274-6801 - llines@iupui.edu
education.iupui.edu

Recommended citation: Kulkarni, T., Weeks, M. R., & Sullivan, A. L. (2021). Critical evaluation of largescale secondary analyses to aid responsible research use and equitable decision making. *Equity Tool*. Midwest & Plains Equity Assistance Center (MAP EAC).

Disclaimer: Midwest and Plains Equity Assistance Center is committed to the sharing of information regarding issues of equity in education. Reference in this tool to any specific publication, person, or idea is for the information and convenience of the public and does not necessarily reflect the views and opinions of Midwest and Plains Equity Assistance Center. The contents of this tool were developed under a grant from the U.S. Department of Education (S004D110021). However, these contents do not necessarily represent the policy of the Department of Education, and endorsement by the Federal Government should not be assumed.