

Homeyra Banaeefar^a, Sarah Burchart^b, Elizabeth Moss^c, and Eszter Palvolgyi-Polyak^d
University of Michigan, Inter-university Consortium for Political and Social Research (ICPSR)
^ahbanaeef@umich.edu, ^bsdonetti@umich.edu, ^ceammoss@umich.edu, ^dppeszter@umich.edu

Project Background

ICPSR is a repository of curated, digitized research data, which follows best practice¹ by providing data users and creators with ways to cite and access datasets over time. Specifically, ICPSR provides a data citation and a URL containing the study's registered DOI (in red, below) for each study in its collection. For example:

Holahan, John, and Karpman, Michael. Health Reform Monitoring Survey, United States, First Quarter 2020. Inter-university Consortium for Political and Social Research [distributor], 2022-02-10. <https://doi.org/10.3886/ICPSR38110.v1>

As staff at ICPSR, we populate the ICPSR Bibliography of Data-related Literature, which contains publications that analyze the research data distributed by ICPSR. Authors sometimes use the study DOI to acknowledge something other than data used in their analysis. We wanted to learn more about this type of use of the study DOI.

We Evaluated 2,546* Cited ICPSR Study DOIs

1 We used an API query of Dimensions Plus²,

a large multidisciplinary database with over 69 million publications available for full text search, to search for over 11,000 ICPSR study DOIs cited in the literature.

2 We examined each publication in the search results

to determine if the DOI was used to cite data analyzed, a requirement for collection, according to the ICPSR Bibliography's strict inclusion criteria³.

Of the cited DOIs we evaluated, did they refer to data analysis?

YES
626

NO
1,031

*889 hits were bad string matches from API

3 We categorized uses of the DOI

that did not refer to the data analyzed.

Three Main Types of ICPSR Study DOI Uses, When Not Citing Data Analyzed

Source Mention 36%

Definition: The study DOI is cited in the publication when the authors refer to findings from another study, or suggest a data source for further investigation.

Example (from Conclusions section⁴):

"the distribution of the psychological distress variable from this study is similar to the distribution of negative emotional experiences in a larger, more representative daily diary study (Ryff and Almeida, 2018)."

C. Ryff, D. Almeida. Midlife in the United States (MIDUS Refresher): Daily Diary Project, 2012-2014 ICPSR [distributor, Ann Arbor, MI (2018), [10.3886/ICPSR37083.v1](https://doi.org/10.3886/ICPSR37083.v1)

Data Mention 35%

Definition: The study DOI is used to give credit for a brief statistic or data point, provided for context or background, but which has nothing to do with the data used in the main analysis.

Example (from Abstract section⁵):

There is no solid evidence regarding the percentage of officers who commit abuse against their intimate partners, and we do not know if officers commit IPV at rates higher than the general population. Estimates of OIVD perpetration range from 10% to 40% (Friedersdorf, 2014; Garvey, 2015; Gershon, 1999).

Gershon, R (1999) Police Stress and Domestic Violence in Police Families in Baltimore, Maryland, 1997-1999. ICPSR02976.v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Available at: <http://doi.org/10.3886/ICPSR02976.v1> (accessed 18 February 2020).

Instrument, Measure, or Methodology Use 29%

Definition: The study DOI is used to acknowledge the use of one or more questions from a survey instrument. Or a measure from a study is used, but not the data from that study. Or authors compare their methodological approach to that used in the cited study.

Example (from Introduction section⁶):

This study focuses on one such subjective measure, used previously in the National Survey of American Life Adolescent supplement (NSAL-A; Jackson et al., 2016). This measure assesses adolescent perceptions of family financial security (PFS) by asking whether their family has enough money to meet their needs. Although this measure has been used to examine SES with geriatric (Wolinsky et al., 2005), adult (Caldwell et al., 2011), and college student samples (Eisenberg et al., 2007), it has yet to be rigorously tested with adolescents.

Jackson et al., 2016. National Survey of American Life - Adolescent Supplement (NSAL-A), 2001-2004: Version 1. ICPSR - Interuniversity Consortium for Political and Social Research (2016), [10.3886/ICPSR36380.V1](https://doi.org/10.3886/ICPSR36380.V1)

ICPSR Studies Whose DOIs Are Often Cited for This Use:

National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2018 [Public Use] (ICPSR 21600)

Equality of Educational Opportunity (COLEMAN) Study (EEO), 1966 (ICPSR 6389)

Longitudinal Study of Generations, California, 1971, 1985, 1988, 1991, 1994, 1997, 2000, 2005 (ICPSR 22100)

Survey of Inmates in State and Federal Correctional Facilities, [United States], 2004 (ICPSR 4572)

Research on Pathways to Desistance [Maricopa County, AZ and Philadelphia County, PA]: Subject Measures, 2000-2010 (ICPSR 29961)

India Human Development Survey-II (IHDS-II), 2011-12

Collaborative Psychiatric Epidemiology Surveys (CPES), 2001-2003 [United States] (ICPSR 20240)

National Hospital Discharge Survey, 2007 (ICPSR 28162)

National Intimate Partner and Sexual Violence Survey (NISVS): General Population Survey Raw Data, 2010 (ICPSR 34305)

National Survey on Drug Use and Health (NSDUH) Series (various studies)

University of Washington - Beyond High School (UW-BHS) (ICPSR 33321)

Iowa Youth and Families Project, 1989-1992 (ICPSR 26721)

Midlife in the United States (MIDUS 1), 1995-1996 (ICPSR 2760) and Midlife in the United States (MIDUS 2), 2004-2006 (ICPSR 4652)

Project on Human Development in Chicago Neighborhoods: Community Survey, 1994-1995 (ICPSR 2766)

Population Assessment of Tobacco and Health (PATH) Study [United States] Restricted-Use Files (ICPSR 36231)

Takeaways

Existing citations are convenient but need clarity: having an existing citation and DOI provided by an archive may create an easy way to "name" or identify something about the data that the author feels should be acknowledged. Indeed, ICPSR's own instructions to "cite data right" may have had an unintended side effect: authors know they must cite data, but when to do so is still not clear.

Further exploration needed of author citation behaviors by conducting direct qualitative interviews with a sample of the authors whose publications did not cite data used in their analysis. This will help us understand what circumstances lead to authors using repository-provided DOIs and if those DOIs are used in an effective way.



Make it easier for authors! Archives and repositories that provide DOIs for datasets should:

- provide clear instructions about when it is appropriate for the DOI to be cited, and
- create and make visible other citation forms authors can use when data are not the actual object being cited, e.g., methodological concepts, constructed measures, survey instruments, mere mentions of sources, and codebooks.

Instructors' pedagogy around data transparency should include:

- requirements to cite data, and
- guidance regarding how and when to use data DOIs.

Acknowledgments

Many thanks to the NSF-funded Measuring the Impact of Curatorial Actions (MICA) staff at the University of Michigan, who created the API query and generated the search results we examined and categorized. Special thanks go to Sara Lafia and Leo Fan. This material is based upon work supported by the National Science Foundation under grant 1930645.

References

¹National Science Foundation. (n.d.). Dear colleague letter: Effective practices for making research data discoverable and citable (Data sharing) (NSF22055). Retrieved April 4, 2022, from <https://www.nsf.gov/pubs/2022/nsf22055/nsf22055.jsp>

²Digital Science. (2018-). Dimensions [Software] available from <https://app.dimensions.ai>. Accessed December 2021, under license agreement. <https://plus.dimensions.ai/support/solutions/articles/23000018794-how-can-i-cite-or-acknowledge-dge-dimensions->

³Moss, E., Cave, C., & Lyle, J. (2015). Chapter 4: Sharing and citing research data: A repository's perspective. In H.K.P. Jayasuriya (Ed.), *Big Data, Big Challenges in Evidence-based Policy Making* (pp. 47-65). Eagan, MN: West Academic Publishing. <https://hdl.handle.net/2027.42/115490>

⁴Anderson, A. R., & Fowers, B. J. (2020). Lifestyle behaviors, psychological distress, and well-being: A daily diary study. *Social Science & Medicine*, 263, 113263. <https://doi.org/10.1016/j.socscimed.2020.101546>

⁵Prost, S. G., Saunders, D. G., & Oehme, K. (2020). Childhood family violence and officer responses to officer-involved domestic violence: Effects of cumulative and resolved trauma. *International Journal of Police Science & Management*, 22(2), 194-207. <https://doi.org/10.1177/14613557209707641>

⁶Hammond, M. A., Khurana, A., & Stormshak, E. A. (2021). Adolescent measures of family socioeconomic status: Reliability, validity, and effects on substance use behaviors in adolescence and young adulthood. *Preventive Medicine Reports*, 21, 101317. <https://doi.org/10.1016/j.pmedr.2021.101317>