1 Research Contributions Over the Last Six Years

Refeered contributions

Journal articles:

Foote, A., Kutzbach, M. J., & Vilhuber, L. (2021). Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings. *Applied Economics*, *53*. https://doi.org/10.1080/00036846.2020.1841083

Vilhuber, L., & Lagoze, C. (2021). Metajelo: A metadata package for journals to support external linked objects. *International Journal of Digital Curation*, *16*(1). https://doi.org/10.2218/ijdc.v16i1.600

Alam, M. J., Dostie, B., Drechsler, J., & Vilhuber, L. (2020a). Applying data synthesis for longitudinal business data across three countries. *Statistics in Transition New Series*, *21*(4), 212–236. https://doi.org/10.21307/stattrans-2020-039

McKinney, K. L., Green, A. S., Abowd, J. M., & Vilhuber, L. (2020). Total error and variability measures with integrated disclosure limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in OnTheMap. *Journal of Survey Statistics and Methodology*, *published online*. https://doi.org/10.1093/jssam/smaa029

Vilhuber, L. (2020). Reproducibility and replicability in economics. *Harvard Data Science Review*, *2*(4). https://doi.org/10.1162/99608f92.4f6b9e67

Weinberg, D. H., Abowd, J. M., Belli, R. F., Cressie, N., Folch, D. C., Holan, S. H., Levenstein, M. C., Olson, K. M., Reiter, J. P., Shapiro, M. D., Smyth, J., Soh, L.-K., Spencer, B. D., Spielman, S. E., Vilhuber, L., & Wikle, C. K. (2018). Effects of a Government-Academic Partnership: Has the NSF-Census Bureau Research Network Helped Improve the U.S. Statistical System? *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smy023

Refereed books and book chapters:

National Academies of Sciences, Engineering, and Medicine. (2020a). *Life-cycle decisions for biomedical data: The challenge of forecasting costs*. The National Academies Press. https://doi.org/10.17226/25639

National Academies of Sciences, Engineering, and Medicine. (2020b). *Planning for long-term use of biomedical data: Proceedings of a workshop* (L. Casola, Ed.). The National Academies Press. https://doi.org/10.17226/25707

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press. https://doi.org/10.17226/25303

Vilhuber, L. (2019a). Making confidential data part of reproducible research [National Academies of Sciences, Engineering, and Medicine; Michael Cohen (Rapporteur)]. In National Academies of Sciences, Engineering, and Medicine (Ed.), *Methods to foster transparency and reproducibility of federal statistics: Proceedings of a workshop* (pp. 63–66). The National Academies Press. https://doi.org/10.17226/25305

Other refereed contributions

Abowd, J. M., Schmutte, I. M., Sexton, W. N., & Vilhuber, L. (2019). Why the economics profession must actively participate in the privacy protection debate. *AEA Papers and Proceedings*, *109*, 397–402. https://doi.org/10.1257/pandp.20191106

Pistner, M., Slavković, A., & Vilhuber, L. (2018). Synthetic data via quantile regression for heavy-tailed and heteroskedastic data. In J. Domingo-Ferrer & F. Montes (Eds.), *Privacy in statistical databases*. https://doi.org/10.1007/978-3-319-99771-1_7

Haney, S., Machanavajjhala, A., Abowd, J. M., Graham, M., Kutzbach, M., & Vilhuber, L. (2017). Utility cost of formal privacy for releasing national employer-employee statistics. *Proceedings of the 2017 International Conference on Management of Data*, 1339–1354. https://doi.org/10.1145/3035918.3035940

Vilhuber, L., & Lagoze, C. (2017). Making confidential data part of reproducible research. *Chance*. http://chance.amstat.org/2017/09/reproducible-research/

Non-refereed contributions

Cole, S., Dhaliwal, I., Sautmann, A., & Vilhuber, L. (Eds.). (2021). *Handbook on Using Administrative Data for Research and Evidence-based Policy* (v1.0). Abdul Latif Jameel Poverty Action Lab. https://doi.org/10.31485/admindatahandbook.1.0

Dwork, C., Karr, A., Nissim, K., & Vilhuber, L. (2021). On privacy in the age of covid-19. *Journal of Privacy and Confidentiality*, *10*(2). https://doi.org/10.29012/jpc.749

Schmutte, I. M., & Vilhuber, L. (2021). Balancing privacy and data usability: An overview of disclosure avoidance methods. In S. Cole, I. Dhaliwal, A. Sautmann, & L. Vilhuber (Eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy* (v1.0). Abdul Latif Jameel Poverty Action Lab. https://doi.org/10.31485/admindatahandbook.1.0

Shen, J. H., & Vilhuber, L. (2021). Physically protecting sensitive data. In S. Cole, I. Dhaliwal, A. Sautmann, & L. Vilhuber (Eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy* (v1.0). Abdul Latif Jameel Poverty Action Lab. https://doi.org/10.31485/admindatahandbook.1.0

Vilhuber, L. (2021a). Expansion, perspectives, and challenges. *Journal of Privacy and Confidentiality*, *11*(1). https://doi.org/10.29012/jpc.777

Vilhuber, L. (2021b). Report by the AEA data editor. *AEA Papers and Proceedings*, *111*, 808–17. https://doi.org/10.1257/pandp.111.808

Alam, M. J., Dostie, B., Drechsler, J., & Vilhuber, L. (2020b). *Applying data synthesis for longitudinal business data across three countries* (arXiv:2008.02246). arXiv. https://arxiv.org/abs/2008.02246

Foote, A. D., Kutzbach, M. J., & Vilhuber, L. (2020, October). *Replication code and data for: Recalculating ... How Uncertainty in Local Labor Market Definitions Affects Empirical Findings* ([data and code] No. v20201007) (See also https://larsvilhuber.github.io/MobZ/README.html.). Labor Dynamics Institute. https://doi.org/10.5281/zenodo.4072428

Lange, F., Vilhuber, L., & Gordon, N. (2020, July). *Uncertainty in times of covid-19: Raw survey data* ([data] No. v20200622-clean). Labor Dynamics Institute. https://doi.org/10.5281/zenodo.3966534

Vilhuber, L., Connolly, M., Koren, M., Llull, J., & Morrow, P. (2020, December). *A template README for social science replication packages*. https://doi.org/10.5281/zenodo.4319999

Vilhuber, L., Turrito, J., & Welch, K. (2020). Report by the AEA data editor. *AEA Papers and Proceedings*, *110*, 764–75. https://doi.org/10.1257/pandp.110.764

Vilhuber, L. (2019b). Report by the AEA data editor. *AEA Papers and Proceedings*, *109*, 718–29. https://doi.org/10.1257/pandp.109.718

Slavković, A., & Vilhuber, L. (2018). Remembering Stephen Fienberg. *Journal of Privacy and Confidentiality*, *8*(1). https://doi.org/10.29012/jpc.685

Vilhuber, L. (2018). Relaunching the Journal of Privacy and Confidentiality. *Journal of Privacy and Confidentiality*, *8*(1). https://doi.org/10.29012/jpc.706

Forthcoming contributions

- "Reproducibility and Transparency versus Privacy and Confidentiality: Reflections from a Data Editor", *Journal of Econometrics*, submitted.
- "Teaching for large-scale Reproducibility Verification", *Journal of Statistics and Data Science Education*, revised and submitted.

2 Other Research Contributions

Editorial activities: Managing Editor of the Journal of Privacy and Confidentiality, 2017-, Data Editor for the American Economic Association, 2018-, Guest Editor for special issue on Linked Employer-Employee Data, Journal of Econometrics, 2019-2021, Associate Editor of the Harvard Data Science Review, 2019-

Committees and Panels: Committee on the Future of Scholarly Communications (Cornell University), member 2019-2022; Contribution to activities of the (U.S.) National Academies of Science, Engineering, and Medicine: • Forecasting Costs for Preserving, Archiving, and Promoting Access to Biomedical Data • Transparency and Reproducibility of Federal Statistics for the National Center for Science and Engineering Statistics • Presentation at "Workshop on Transparency and Reproducibility in Federal Statistics" • Presentation to "Committee on Reproducibility and Replicability in Science" Member of the Board of Directors of the Canadian Research Data Centre Network/ Réseau canadien des Centres de données de recherche, 2019- ; American Statistical Association, Committee on Privacy and Confidentiality, member 2016-2021, chair 2021; Comité scientifique du Centre d'accès sécurisé aux données (CASD) [Scientific advisory board of the Center for secure data access, France], member 2012-2014, chair 2014-; Conseil d'orientation du Centre interuniversitaire québécois de statistiques sociales (CIQSS)/ Advisory board of the Quebec Inter-University Centre for Social Statistics (QICSS) [Canada], member, 2014-; Social Sciences and Humanities Research Council of Canada (SSHRC), Adjudication Committee for Partnership Grants, member 2018-2019; National Science Foundation, Methodology, Measurement, and Statistics (MMS) Advisory Panel, member 2016-2018;

3 Most Significant Career Research Contributions

1. Vilhuber, L., Connolly, M., Koren, M., Llull, J., & Morrow, P. (2020, December). *A template README for social science replication packages*. https://doi.org/10.5281/zenodo.4319999

The template README has already been used by several hundred manuscripts, and greatly improved reproducibility of journal articles.

2. Machanavajjhala, A., Kifer, D., Abowd, J. M., Gehrke, J., & Vilhuber, L. (2008). Privacy: Theory meets practice on the map. *International Conference on Data Engineering (ICDE)*, 277–286. https://doi.org/10.1109/ICDE.2008.4497436
   Documents the first large-scale implementation of differential privacy, an innovative and greatly superior way of publishing data while preserving respondents privacy.

3. Abowd, J. M., Stephens, B. E., Vilhuber, L., Andersson, F., McKinney, K. L., Roemer, M., & Woodcock, S. D. (2009). The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators. In T. Dunne, J. B. Jensen, & M. J. Roberts (Eds.), *Producer dynamics: New evidence from micro data*. University of Chicago Press. http://www.nber.org/chapters/c0485
   Documents the publication of a novel data resource, still being produced by the U.S. Census Bureau, with 1,000s of downloads, and use far beyond academia. Innovative in its use of disclosure avoidance to provide detailed tabular statistics at unprecedented levels.

4. Abowd, J., & Vilhuber, L. (2008, September). How protective are synthetic data. In J. Domingo-Ferrer & Y. Saygın (Eds.), *Privacy in statistical database* (pp. 239–246, Vol. 5262). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87471-3_20
   One of the earliest publications identifying the privacy protective power of synthetic (artificial data).

5. Abowd, J. M., & Vilhuber, L. (2005). The sensitivity of economic statistics to coding errors in personal identifiers. *Journal of Business and Economic Statistics*, *23*(2), 133–152. http://www.jstor.org/stable/27638803
   Often cited to illustrate that administrative data is not perfect, and can be biased for reasons (at the time) not well appreciated by the economics and statistics communities.

4 Career Interruptions and Special Circumstances

None.

5 Contributions to Training

Graduate thesis committees: Ph.D. Committee member for: Raghunathan, Kalyani (2015) Green, Andrew (2017). M.Sc. Committee member for: Daniel Li (2019)
Undergraduate involvement in research: Since 2015, I have involved about 90 students in research projects that I am involved with. The bulk of these have been trained and employed as part of the LDI Replication Lab, verifying reproducibility for the American Economic Association.

6 Relevant Experience

As Data Editor for the American Economic Association, I have organized the LDI Replication Lab. Each semester, including the summer semester, we train 12-15 students in methods to verify data provenance and reproducibility. Over 75 students have been trained so far, and at any point in time, about 15-25 students are active in the lab. Over 1,000 articles have been evaluated since the Lab opened.

I have also organized several conferences, including being lead organizer on the "Conference on Models of Linked Employer-Employee Data" (with Sloan and Cornell funding). A conference volume, to be published by the Journal of Econometrics, is currently being finalized.

As lead PI for the second half of the NSF-funded "RCN: Coordination of the NSF-Census Research Network" project, we organized two workshops per year, with presentations from researchers working on eight different NSF-funded projects. Working papers by the network were archived (preserved), and the entire research coordination website was preserved at the end of the grant. An NSF-organized mid-term evaluation deemed the work of the Coordination Office to be fully satisfactory. A summary publication of the network's work was published by the *Journal of Survey Statistics and Methodology* (Weinberg et al, 2018), and the network was awarded a prize by the American Statistical Association for excellent networking between academia and government.