

**Reiter, J. P.**

**Significance tests for multi-component estimands from multiply imputed, synthetic microdata.** (English) [Zbl 1061.62201](#)

*J. Stat. Plann. Inference* 131, No. 2, 365-377 (2005).

Summary: To limit the risks of disclosures when releasing data to the public, it has been suggested that statistical agencies release multiply imputed, synthetic microdata. For example, the released microdata can be fully synthetic, comprising random samples of units from the sampling frame with simulated values of variables. Or, the released microdata can be partially synthetic, comprising the units originally surveyed with some collected values, e.g., sensitive values at high risk of disclosure or values of key identifiers, replaced with multiple imputations. This article presents inferential methods for synthetic data for multi-component estimands, in particular procedures for Wald and likelihood ratio tests. The performance of the procedures is illustrated with simulation studies.

**MSC:**

[62P99](#) Applications of statistics  
[62H15](#) Hypothesis testing in multivariate analysis  
[62G10](#) Nonparametric hypothesis testing

Cited in **5** Documents

**Keywords:**

[Confidentiality](#); [Disclosure](#); [Multiple imputation](#); [Significance tests](#); [Synthetic data](#)

**Software:**

[BayesDA](#)

**Full Text:** [DOI](#)

**References:**

- [1] Abowd, J.M.; Woodcock, S.D., Disclosure limitation in longitudinal linked data, (), 215-277
- [2] Dandekar, R.A.; Cohen, M.; Kirkendall, N., Sensitive micro data protection using Latin hypercube sampling technique, (), 117-125 · [Zbl 1051.68629](#)
- [3] Dandekar, R.A.; Domingo-Ferrer, J.; Sebe, F., LHS-based hybrid microdata versus rank swapping and microaggregation for numeric microdata protection, (), 153-162 · [Zbl 1051.68630](#)
- [4] Fienberg, S.E., Steele, R.J., Markov, U.E., 1996. Statistical notions of data disclosure avoidance and their relationship to traditional statistical methodology: data swapping and log-linear models. In: Proceedings of Bureau of Census 1996 Annual Research Conference, pp. 87-105.
- [5] Fienberg, S.E.; Markov, U.E.; Steele, R.J., Disclosure limitation using perturbation and related methods for categorical data, *J. official statist.*, 14, 485-502, (1998)
- [6] Franconi, L.; Stander, J., A model based method for disclosure limitation of business microdata, *Statistician*, 51, 1-11, (2002)
- [7] Franconi, L., Stander, J., 2003. Spatial and non-spatial model-based protection procedures for the release of business microdata. *Statist. Comput.* 13, 295-306.
- [8] Fuller, W.A., Masking procedures for microdata disclosure limitation, *J. official statist.*, 9, 393-406, (1993)
- [9] Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B., *Bayesian data analysis*, (1995), Chapman & Hall London
- [10] Kennickell, A.B., Multiple imputation and disclosure protection: the case of the 1995 survey of consumer finances, (), 248-267
- [11] Li, K.H., 1985. Hypothesis testing with multiple imputation—with emphasis on mixed-up frequencies in contingency tables. Ph.D. Thesis, Department of Statistics, University of Chicago.
- [12] Li, K.H.; Raghunathan, T.E.; Rubin, D.B., Large sample significance levels from multiply imputed data using moment-based statistics and an  $\text{F}$  reference distribution, *J. amer. statist. assoc.*, 86, 1065-1073, (1991)
- [13] Little, R.J.A., Statistical analysis of masked data, *J. official statist.*, 9, 407-426, (1993)
- [14] Liu, F., Little, R.J.A., 2002. Selective multiple imputation of keys for statistical disclosure control in microdata. In: *ASA Proceedings of the Joint Statistical Meetings*, pp. 2133-2138.

- [15] Meng, X.I.; Rubin, D.B., Performing likelihood ratio tests with multiply-imputed data sets, *Biometrika*, 79, 103-111, (1992) · [Zbl 0754.62041](#)
- [16] Polettini, S., 2003. Maximum entropy simulation for microdata protection. *Statist. Comput.* 13, 307-320
- [17] Polettini, S.; Franconi, L.; Stander, J., Model-based disclosure protection, (), 83-96 · [Zbl 1051.68774](#)
- [18] Raghunathan, T.E., 1987. Large sample significance levels from multiply-imputed data. Ph.D. Thesis, Department of Statistics, Harvard University.
- [19] Raghunathan, T.E.; Reiter, J.P.; Rubin, D.B., Multiple imputation for statistical disclosure limitation, *J. official statist.*, 19, 1-16, (2003)
- [20] Reiter, J.P., Satisfying disclosure restrictions with synthetic data sets, *J. official statist.*, 18, 531-544, (2002)
- [21] Reiter, J.P., 2003. Inference for partially synthetic, public use microdata sets. *Surv. Methodology* 29, 181-188
- [22] Reiter, J.P., 2004. Releasing multiply-imputed, synthetic public use microdata: an illustration and empirical study. *J. Roy. Statist. Soc. Ser. A*, forthcoming. · [Zbl 1099.62138](#)
- [23] Rubin, D.B., Multiple imputation for nonresponse in surveys, (1987), Wiley New York · [Zbl 1070.62007](#)
- [24] Rubin, D.B., Discussionstatistical disclosure limitation, *J. official statist.*, 9, 462-468, (1993)
- [25] Schafer, J.L., Analysis of incomplete multivariate data, (2000), Chapman & Hall London
- [26] Willenborg, L.; de Waal, T., Elements of statistical disclosure control, (2001), Springer New York · [Zbl 0973.62009](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.