

## MICRODATA CONFIDENTIALITY REFERENCES W. E. Winkler Feb. 16, 2002

- Abowd, J. M. and Woodcock, S. D. (2002), "Disclosure Limitation in Longitudinal Linked Data," in (P. Doyle et al, eds.) *Confidentiality, Disclosure, and Data Access*, North Holland: Amsterdam.
- Agrawal, D. and Aggarwal, C. C. (2001), "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," *Association of Computing Machinery, Proceedings of SIGMOD*.
- Adams, N. R. and Wortmann, J. C., (1989), "Security-control Methods for Statistical Databases, A Comparative Study," *ACM Computing Surveys*, **21**, 515-556.
- Bacher, J., Bender, S. and Brand, R. (2001), "Re-identifying Register Data by Survey Data: An Empirical Study," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Bethlehem, J. A., Keller, W. J., and Pannekoek, J., (1990) "Disclosure Control of Microdata," *Journal of the American Statistical Association*, **85**, 38-45.
- Blien, U., Wirth, U., and Muller, M. (1992), "Disclosure Risk for Microdata Stemming from Official Statistics," *Statistica Neerlandica*, **46**, 69-82.
- Brand, R. (2002), "Microdata Protection Through Noise Addition," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.
- Dalenius, T. and Reiss, S.P. (1982), "Data-swapping: A Technique for Disclosure Control," *Journal of Statistical Planning and Inference*, **6**, 73-85.
- Dandekar, R., Domingo-Ferrer, J. and Sebe, F. (2002), "LHS-Based Hybrid Microdata vs Rank Swapping and Microaggregation for Numeric Microdata Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.
- Dandekar, R., Cohen, M., and Kirkendal, N. (2002), "Sensitive Microdata Protection Using Latin Hypercube Sampling Technique," in (J. Domingo-Ferrer, ed.) *Statistical Data Protection: From Theory to Application*, Springer: New York.
- Davies, S. and Moore, A. (1999), "Bayesian Networks for Lossless Dataset Compression," Association of Computing Machinery, Conference of Knowledge Discovery and Datamining.
- Defays, D. and Nanopolis, P. (1993), "Panels of Enterprises and Confidentiality: the Small Aggregates Method," in Proceedings of the 1992 Symposium on Design and Analysis of Longitudinal Surveys, 195-204.
- De Waal, A. G., and Willenborg, L.C.R.J. (1995), "Global Recodings and Local Suppressions in Microdata Sets," *Proceedings of Statistics Canada Symposium 95*, 121-132
- De Waal, A. G., and Willenborg, L.C.R.J. (1996), "A View of Statistical Disclosure Control for Microdata," *Survey Methodology*, **22**, 95-103.
- Domingo-Ferrer, J. (2001), "On the Complexity of Microaggregation," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Domingo-Ferrer, J. and Mateo-Sanz, J. M. (2001), "An Empirical Comparison of SDC Methods for Continuous Microdata in Terms of Information Loss And Re-Identification Risk," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Domingo-Ferrer, J. and Mateo-Sanz, J. M. (2002), "Practical Data-Oriented Microaggregation for Statistical Disclosure Control," *IEEE Transactions on Knowledge and Data Engineering*, to appear.
- DuMouchel, W., Volinsky, C., Johnson, T., Cortes, C. and Pregibon, D. (2000), "Squashing Flat Files Flatter," Association of Computing Machinery, Proceedings of Knowledge Discovery in Data, 6-15.
- Fellegi, I. P. (1997), "Record Linkage and Public Policy - A Dynamic Evolution," *Proceedings of the Record Linkage Workshop 1997*, Washington, DC: National Academy Press, 3-12.
- Fellegi, I. P., and Sunter, A. B. (1969), "A Theory for Record Linkage," *Journal of the American Statistical Association*, **64**, 1183-1210.
- Fienberg, S. E. (1997), "Confidentiality and Disclosure Limitation Methodology: Challenges for National Statistics and Statistical Research, commissioned by Committee on National Statistics of the National Academy of Sciences.
- Fienberg, S. E., Makov, E. U., and Sanil, A. P., (1997), "A Bayesian Approach to Data Disclosure: Optimal Intruder Behavior for Continuous Data," *Journal of Official Statistics*, **14**, 75-89.
- Fienberg, S. E., Makov, E. U., and Steel, R. J. (1998), "Disclosure Limitation using Perturbation and Related Methods for Categorical Data," *Journal of Official Statistics*, **14**, 485-502.
- Frakes, W. and Baeza-Yates, R. (1992), "Information Retrieval - Data Structures and Algorithms," Prentice-Hall: Upper Saddle River, N.J.
- Franconi, L., Capobianchi, A., Polletini, S., and Seri, G. (2001), "Experiences in Model-Based Disclosure Protection," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Fuller, W. A. (1993), "Masking Procedures for Microdata Disclosure Limitation," *Journal of Official Statistics*, **9**, 383-406.

- Gill, L. (1999), "OX-LINK: The Oxford Medical Record Linkage System," in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 15-33.
- Gopal, R., P. Goes, and R. Garfinkel, "Confidentiality Via Camouflage: The CVC Approach to Database Query Management," in *Statistical Data Protection '98*, Eurostat, Brussels, Belgium, 1-8.
- Grim, J., Bocek, P., and Pudil, P. (2001), "Safe Dissemination of census Results by Means of Interactive Probabilistic Models," *Proceedings of 2001 NITS and ETK*, Eurostat: Luxembourg, 849-856.
- Hwang, J. T. (1986), "Multiplicative Error-in-Variables Models with Applications to Recent Data Released by the U.S. Department of Energy," *Journal of the American Statistical Association*, **81** (395), 680 - 688.
- Kennickell, A. B. (1999), "Multiple Imputation and Disclosure Control: The Case of the 1995 Survey of Consumer Finances," in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 248-267 (available at <http://www.fcsn.gov>).
- Kim, J. J. (1986), "A Method for Limiting Disclosure in Microdata Based on Random Noise and Transformation," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 303-308.
- Kim, J. J. (1990), "Subdomain Estimation for the Masked Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 456-461.
- Kim, J. J., and Winkler, W. E. (1995), "Masking Microdata Files," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 114-119.
- Kim, J. J., and Winkler, W. E. (2001), "Multiplicative Noise for Masking Continuous Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, to appear.
- Lambert, D. (1993), "Measures of Disclosure Risk and Harm," *Journal of Official Statistics*, **9**, 313-331.
- Lawrence, C., Zhou, J.L., and Tits, A. L. (1997), "User's Guide for CFSZP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear Inequality Constraints," Unpublished, Electrical Engineering Dept. and Institute for Systems Research, University of Maryland.
- Liew, C. K., Choi, U. J. and Liew, C. J. (1991), "A Data Distortion by Probability Distribution," *ACM Transactions on Database Systems*, **10**, 395-411.
- Little, R. J. A. (1993), "Statistical Analysis of Masked Data," *Journal of Official Statistics*, **9**, 407-426.
- Mera, R. (1998), "Matrix Masking Methods That Preserve Moments," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 445-450.
- Moore, A. (1999), "Very Fast EM-based Mixture Model Clustering using Multiresolution KD-Trees," *Neural Information Processing Systems 11*.
- Moore, A. and Lee, M. S. (1998), "Cached Sufficient Statistics for Efficient Machine Learning with Large Datasets," *Journal of Artificial Intelligence Research*, **8**, 67-91.
- Moore, A., Schneider, J., and Deng, K. (1997), "Efficient Locally Weighted Polynomial Regression Predictions," *Proceedings of the 1997 International Machine Learning Conference*, Morgan Kaufmann Publishers.
- Moore, R. (1995), "Controlled Data Swapping Techniques For Masking Public Use Data Sets," U.S. Bureau of the Census, Statistical Research Division Report rr96/04, (available at <https://www.census.gov/srd/www/byyear.html>).
- Muralidhar, K., Batrah, D. and Kirs, P.J. (1995), "Accessibility, Security, and Accuracy in Statistical Databases : The Case for the Multiplicative Fixed Data Perturbation Approach," *Management Science* **41**(9). 1549-1584
- Muralidhar, K., Parsa, R. and Sarathy, R. (1999), "A General Additive Data Perturbation Method for Database Security," *Management Science*, **45**(10), 1399-1415.
- Muralidhar, K. and Sarathy, R. (1999) "Security of Random Data Perturbation Methods," *ACM Transactions on Database Systems*, **24**(4), 487-493.
- Muralidhar, K., Sarathy, R. and R. Parsa, R. (2002) "An Improved Security Requirement for Data Perturbation with Implications for E-Commerce," *Decision Sciences* (Forthcoming).
- Paas, G. (1988), "Disclosure Risk and Disclosure Avoidance for Microdata," *Journal of Business and Economic Statistics*, **6**, 487-500.
- Pollitini, S., Franconi, L., and Stander, J. (2002), "Model Based Disclosure Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.
- Raghunathan, T.E. and Rubin, D.R. (2000), "Multiple Imputation for Disclosure Limitation" technical report.
- Reiss, J.P. (1984), "Practical Data Swapping: The First Steps," *ACM Transactions on Database Systems*, **9**, 20-37.
- Roque, G. M. (2000), "Masking Microdata Files with Mixtures of Multivariate Normal Distributions," Ph.D. Dissertation, Department of Statistics, University of California at Riverside.
- Rubin, D. B. (1993), "Satisfying Confidentiality Constraints through the Use of Synthetic Multiply-imputed Microdata," *Journal of Official Statistics*, **91**, 461-468.
- Sarathy, R. and K. Muralidhar, K. (2002), "The Security of Confidential Numerical Data in Databases," *Information Systems Research* (Forthcoming).
- Scheuren, F., and Winkler, W. E. (1996), "Recursive Merging and Analysis of Administration Lists," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 123-128

- (presently available on <http://www.amstat.org> in the Section on Government Statistics).
- Scheuren, F. and W. E. Winkler, W. E. (1997), "Regression Analysis of Data Files that are Computer Matched – Part II," *Survey Methodology*, 157-165).
- Schlörer, J. (1981), "Security of Statistical Databases: Multidimensional Transformation," *ACM Transactions on Database Systems*, **6**, 91-112.
- Stander, J., and Franconi, L. (2001), "A Model-Based Disclosure Limitation Method for Business Microdata," presented at the UNECE Workshop On Statistical Data Editing, Skopje, Macedonia, May 2001.
- Sullivan, G., and Fuller, W. A. (1989), "The Use of Measurement Error to Avoid Disclosure," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 802-807.
- Sullivan, G., and Fuller, W. A. (1990), "Construction of Masking Error for Categorical Variables," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 435-439.
- Sweeney, L. (1999), "Computational Disclosure Control for Medical Microdata: The Datafly System" in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 442-453.
- Tendick, P. and Matloff, N. (1994), "A Modified Random Perturbation Method for Database Security," *ACM Transactions on Database Systems*, **19**, 47-63.
- Thibaudeau, Y. and Winkler, W.E. (2002), "Bayesian Networks Representations, Generalized Imputation, and Synthetic Microdata satisfying Analytic Restraints," Statistical Research Division report at <https://www.census.gov/srd/www/byyear.html>, to appear.
- Van Gewerden, L., Wessels, A., and Hundepol, A. (1997), "Mu-Argus Users Manual, Version 2," Statistics Netherlands, Document TM-1/D.
- Willenborg, L. and De Waal, T. (1996), *Statistical Disclosure Control in Practice*, Vol. 111, Lecture Notes in Statistics, Springer-Verlag, New York.
- Willenborg, L. and De Waal, T. (2000), *Elements of Statistical Disclosure Control*, Vol. 155, Lecture Notes in Statistics, Springer-Verlag, New York.
- Winkler, W. E. (1988), "Using the EM Algorithm for Weight Computation in the Fellegi-Sunter Model of Record Linkage," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 667-671.
- Winkler, W. E. (1989), "Near Automatic Weight Computation in the Fellegi-Sunter Model of Record Linkage," *Proceedings of the Fifth Census Bureau Annual Research Conference*, 145-155.
- Winkler, W. E. (1993), "Improved Decision Rules in the Fellegi-Sunter Model of Record Linkage," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 274-279.
- Winkler, W. E. (1994), "Advanced Methods for Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods*, pp. 467-472.
- Winkler, W. E. (1995), "Matching and Record Linkage," in B. G. Cox (ed.) *Business Survey Methods*, New York: J. Wiley, 355-384.
- Winkler, W. E. (1998), "A Re-identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata," *Research in Official Statistics*, **1**, 87-104.
- Yancey, W.E., Winkler, W.E., and Creecy, R. H. (2002) "Disclosure Risk Assessment in Perturbative Microdata Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, Springer: New York.