

Microdata Confidentiality References william.e.winkler@census.gov 2010Jan15

- Abowd, J. M., and Vilhuber, L. (2008). "How Protective are Synthetic Data?" in (J. Domingo-Ferrer and V. Yucel, eds.) *Privacy in Statistical Databases 2008*, New York, N.Y.: Springer, 239-246.
- 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*, Amsterdam, The Netherlands: North Holland.
- Abowd, J. M., and Woodcock, S. D. (2004), "Multiply-Imputing Confidential Characteristics and File Links in Longitudinal Linked Data, in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases 2004*, New York: Springer .
- Adams, N. R., and Wortmann, J. C., (1989), "Security-control Methods for Statistical Databases, A Comparative Study," *ACM Computing Surveys*, 21, 515-556.
- Aggarwal, C. C., (2005), "On k-Anonymity and the Curse of Dimensionality," *Proceedings of the 31st VLDB Conference*, Trondheim, Norway, <http://www.vldb2005.org/program/paper/fri/p901-aggarwal.pdf> .
- Aggarwal, C. C., and Parthasarathy, S. (2001), "Mining Massively Incomplete Data Sets through Conceptual Reconstruction," *Proceedings of the ACM KDD Conference*, 227-232.
- Aggarwal, C. C., and Yu, P. (2004), "A Condensation Approach to Privacy Preserving Data Mining," *Proceedings of the EBDT Conference*, 183-199.
- Aggarwal, G., Feder, T., Kenthapadi, K., Motwani, R., Panigrahy, R., Thomas, D., and Zhu, A. (2005), "Anonymizing Tables," *International Conference on Database Theory*.
- Aggarwal, G., Feder, T., Kenthapadi, K., Khuller, S. Panigrahy, R., Thomas, D., and Zhu, A. (2006), "Achieving Anonymity via Clustering," *ACM PODS '06*.
- Agrawal, D., and Aggarwal, C. C. (2001), "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," *Association of Computing Machinery, Proceedings of PODS 2001*, 247-255.
- Agrawal, R., and Srikant, R. (2000), Privacy Preserving Data Mining, *Proceedings of the ACM SIGMOD 2000*, 439-450.
- Agrawal, R., Srikant, R., and Thomas, D. (2005), "Privacy Preserving OLAP," *ACM SIGMOD Conference*.
- Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2002), "Hippocratic Databases," *Very Large Databases 2002*.
- 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.
- Bacher, J., Brand, R., and Bender, S. (2002) Re-identifying Register Data by Survey Data using Cluster Analysis: An Empirical Study, *International Journal of Uncertainty, Fuzziness, Knowledge-Based Systems*, 10 (5) 589-608.
- Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., and Talwar, K. (2007), "Privacy, Accuracy, and Consistency Too: A Holistic Solution to Contingency Table Release," *PODS '07*, Beijing, China.
- Bayardo, R. J., and Agrawal, R. (2005), "Data Privacy Through Optimal K-Anonymization," *IEEE 2005 International Conference on Data Engineering*.
- 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.
- Blum, A., Dwork, C., McSherry, F., and Nissim, K. (2005), "Practical Privacy: The SuLQ Framework," *ACM SIGMOD Conference* (also <http://research.microsoft.com/research/sv/DatabasePrivacy/bdmn.pdf>).
- Brand, R. (2002), "Microdata Protection Through Noise Addition," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, New York: Springer, 97-116.
- Castro, J. (2004), "Computational Experience with Minimum-Distance Controlled Perturbation Methods," in (J. Domingo-Ferrer, ed.), *Privacy in Statistical Databases 2004*, Springer: New York.
- Chawla, S., Dwork, C., McSherry, F., Smith, A., and Wee, H. (2004), "Toward Privacy in Public Databases," Microsoft Research Technical Report, Theory of Cryptography Conference.
- Chawla, S., Dwork, C., McSherry, F., and Talwar, K. (2005), "On the Utility of Privacy-Preserving Histograms," <http://research.microsoft.com/research/sv/DatabasePrivacy/cdmt.pdf> .
- 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. A. (2004), Maximum Utility Minimum Information Loss Table Server Design of Statistical Disclosure Control of Tabular Data, in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases*, New York: Springer, 121-135.

- Dandekar, R. A., 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*, New York: Springer, 175-186.
- Dandekar, R., Cohen, M., and Kirkendal, N. (2002), "Sensitive Microdata Protection Using Latin Hypercube Sampling Technique," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*, New York: Springer, 117-125.
- Defays, D., and Anwar, M. N. (1998), "Masking Microdata Using Micro-aggregation," *Journal of Official Statistics*, 14, 449-461.
- 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.
- De Wolf, P.-P. (2007), "Risk, Utility and PRAM: A Comparison of Proximity Swap and Data Shuffle for Numeric Data," in (J. Domingo-Ferrer, ed.) *Statistical Data Protection 2006*, Springer: New York, N.Y.
- Dinur, I., and Nissim, K. (2003), "Revealing Information while Preserving Privacy," *ACM PODS Conference*, 202-210.
- 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. (ed.) (2002) *Inference Control in Statistical Databases*, New York: Springer
- 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*, 14 (1), 189-201.
- Domingo-Ferrer, J., Mateo-Sanz, J., Oganian, A., and Torres, A. (2002), "On the Security of Microaggregation with Individual Ranking: Analytic Attacks," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 477-492.
- Domingo-Ferrer, J., Sebé, F., and Castellà-Roca, J. (2004), "On the Security of Noise Addition for Privacy In Statistical Databases, in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases*, New York: Springer, 149-161.
- Domingo-Ferrer, J., and Torra, V. (2001) A Quantitative Comparison of Disclosure Control Methods for Microdata in (P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz, eds.) *Confidentiality, Disclosure Control and Data Access: Theory and Practical Applications*, Amsterdam, The Netherlands: North Holland, 111-134.
- Domingo-Ferrer, J., and Torra, V. (2003), "Statistical Data Protection in Statistical Microdata Protection Via Advanced Record Linkage," *Statistics and Computing*, 13 (4), 343-354.
- Du, W., Han, Y. S., and Chen, S. (2004), "Privacy Preserving Multivariate Statistical Analysis: Linear Regression and Classification," *SIAM International Conference on Data Mining 2004*.
- Du, W., and Zhan, Z. (2003), "Using Randomized Response Techniques for Privacy Preserving Data Mining," *ACM Knowledge Discovery and Data Mining Conference 2003*, 505-510.
- DuMouchel, W., Volinsky, C., Johnson, T., Cortes, C., and Pregibon, D. (1999), "Squashing Flat Files Flatter," *Proceedings of the ACM Knowledge Discovery and Data Mining Conference*, 6-15.
- Duncan, G. T., Keller-McNulty, S. A., and Stokes, S. L. (2001), "Disclosure Risk vs. Data Utility: The R-U Confidentiality Map," Los Alamos National Laboratory Technical Report LA-UR-01-6428.
- Dwork, C. (2006), "Differential Privacy," 33rd International Colloquium on Automata, Languages and Programming – ICALP 2006, Part II, 1-12.
- Dwork, C. (2008), "Differential Privacy: A Survey of Results," in (M. Agrawal et al., eds.) TAMC 2008, LNCS 4978, 1-19.
- Dwork, C., and Lei, J. (2009), "Differential Privacy and Robust Statistics," STOC '09.
- Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006), "Calibrating Noise to Sensitivity in Private Data Analysis," 3rd Conference on Cryptography – TCC 2006, 365-384.
- Dwork, C., McSherry, F., and Talwar, K. (2007a), "The Price of Privacy and the Limits of LP Decoding," STOC '07, San Diego, CA.

- Dwork, C., McSherry, F., and Talwar, K. (2007b), "Differentially Private Marginals Release with Mutual Consistency and Error Independent Sample Size," UNECE Worksession on Statistical Data Confidentiality, Manchester, UK, at <http://www.unece.org/stats/documents/2007/12/confidentiality/wp.19.e.pdf> .
- Dwork, C. and Yekhanin, S. (2008), "New Efficient Attacks on Statistical Disclosure Control Mechanisms," Advances in Cryptology—CRYPTO 2008, to appear, also at <http://research.microsoft.com/research/sv/DatabasePrivacy/dy08.pdf> .
- Dwork, C. and Nissim, K. (2004), "Privacy-Preserving Datamining on Vertically Partitioned Databases," Microsoft Research Technical Report.
- Elamir, E. A. H. (2004), "Analysis of Re-identification Risk based on Log-Linear Model," in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases*, Springer: New York, 273-281.
- Elamir, E. A. H. and Skinner, C. J. (2006), "Record Level Measures of Disclosure Risk for Survey Microdata," *Journal of Official Statistics*, 22, 525-539.
- Elliott, M. A., Manning, A. M., and Ford, R. W. (2002), "A Computational Algorithm for Handling the Special Uniques Problem," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10 (5), 493-510.
- Elliot, M.J., Skinner, C. A., and Dale, A. (1998), "Special Uniques, Random Uniques, and Sticky Populations: Some Counterintuitive Effects of Geographic Detail in Disclosure Risk," *Research in Official Statistics*, 1, 53-68.
- Evfimievski, A. (2004), "Privacy Preserving Information Sharing," Ph.D. Dissertation, Cornell University, <http://www.cs.cornell.edu/aevf/> .
- Evfimievski, A., Gehrke, J., and Srikant, R. (2003), "Limiting Privacy Breaches in Privacy Preserving Data Mining," *ACM PODS Conference*, 211-222.
- Evfimievski, A., Srikant, R., Agrawal, R., and Gehrke, J. (2002), "Privacy Preserving Mining of Association Rules," *Association of Computing Machinery, Special Interest Group on Knowledge Discovery and Datamining 2002*.
- Fellegi, I. P., and Sunter, A. B. (1969), "A Theory for Record Linkage," *Journal of the American Statistical Association*, 64, 1183-1210.
- Fellegi, I. P. (1972), "On the Question of Statistical Confidentiality," *Journal of the American Statistical Association*, 67, 7-18.
- Fellegi, I. P. (1999), "Record Linkage and Public Policy - A Dynamic Evolution," *Proceedings of the Record Linkage Workshop 1997*, Washington, DC: National Academy Press, 3-12.
- 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. and MacIntyre, J. (2005), "Data Swapping: Variations on a Theme of Dalenius and Reiss," *Journal of Official Statistics*, 21 (2), 309-323.
- 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*, Upper Saddle River, NJ: Prentice-Hall.
- Franconi, L., Capobianchi, A., Polletini, S., and Seri, G. (2001), "Experiences in Model-Based Disclosure Protection," presented at the UNECE Workshop on Statistical Data Confidentiality, Skopje, Macedonia, May 2001.
- Fuller, W. A. (1993), "Masking Procedures for Microdata Disclosure Limitation," *Journal of Official Statistics*, 9, 383-406 (<http://www.jos.nu/Articles/abstract.asp?article=92383>).
- Fung, B. C. M., Wang, K., and Yu, P. S. (2005), Top-Down Specialization for Information and Privacy Preservation," *IEEE International Conference on Data Engineering*, 205-216.
- Ganta, S., Prasad, S., and Smith, A. (2008), Compositional Attacks and Auxiliary Information in Data Privacy," *ACM KDD '08*, 265-273.
- Gopal, R., Goes, P. and Garfinkel, R. (1998) "Confidentiality Via Camouflage: The CVC Approach to Database Query Management," in *Statistical Data Protection '98*, Eurostat, Brussels, Belgium, 1-8. (also (2002) *Operations Research*, 50 (3)).
- Gilburd, B., Schuster, A., and Wolff, R. (2004b), "k-TTP: A New Privacy Model for Large-Scale Distributed Environments," *ACM Knowledge Discovery and Data Mining Conference 2004*, 599-604.
- Gill, L. (1999), "OX-LINK: The Oxford Medical Record Linkage System," in *Record Linkage Techniques 1997*, Washington, DC: National Academy Press, 15-33.
- Gomatam, S. V., and Karr, A. (2003), "On Data Swapping of Categorical Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.

- Gouweleeuw, J.M., Kooiman, P., Willenborg, L. C. R. J., and De Wolf, P.-P. (1998), "Post Randomisation For Statistical Disclosure Control: Theory and Implementation," *Journal of Official Statistics*, 14, 463-478.
- Graham, P., Young, J., and Penny, R. (2009), "Multiply Imputed Synthetic Data: Evaluation of Hierarchical Bayesian Imputation Models," *Journal of Official Statistics*, 25 (2), 245-268.
- Grim, J., Bocek, P., and Pudil, P. (2001), "Safe Dissemination of Census Results by Means of Interactive Probabilistic Models," *Proceedings of 2001 NTTS and ETK*, Luxembourg: Eurostat, 849-856.
- Huang, Z., Du, W., and Chen, B. (2005), "Deriving Private Information from Randomized Data," *ACM SIGMOD 2005 Conference*, 37-48.
- Huckett, J. C. (2008), "Synthetic Data Methods for Disclosure Limitation," Ph.D. Thesis, Department of Statistics, Iowa State University.
- Huckett, J. C., and Larsen, M. D. (2007), "Microdata Simulation for Confidentiality Protection Using Regression Quantiles and Hot Deck," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM, 3053-3060.
- 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.
- Iyengar, V. (2002), "Transforming Data to Satisfy Privacy Constraints," *Association of Computing Machinery, Knowledge Discovery and Data Mining Conference 2002*, 279-288.
- Kantarcioglu, M., and Clifton, C. (2004a), "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," *IEEE Transactions on Knowledge and Data Engineering*, 16 (9), 1026-1037.
- Kantarcioglu, M., and Clifton, C. (2004b), "When Do Data Mining Results Violate Privacy?" *Association of Computing Machinery, Knowledge Discovery and Data Mining Conference 2004*, 599-604.
- Kargupta, H., Datta, S., Wang, Q., and Ravikumar, K. (2003) "Random Data Perturbation Techniques and Privacy Preserving Data Mining," Expanded version of best paper awarded paper from the IEEE International Conference on Data Mining, November, 2003, Orlando, FL, (also version to appear in *Knowledge and Information Systems Journal*, http://www.cs.umbc.edu/~hillol/PUBS/kargupta_privacy03a.pdf).
- Kaufmann, S., Seastrom, M., and Roey, S. (2005), "Do Disclosure Controls to Protect Confidentiality Degrade the Quality of the Data? Data from the 2003 Trends in Mathematics and Science Study," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Keller-McNulty, S., and Unger, E. (2003), "Database Systems: Inferential Security," *Journal of Official Statistics*, 9 (2), 475-499.
- 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.fcsm.gov> under Methodology reports).
- Kifer, D. (2009), "Attacks on Privacy and deFinetti's Theorem," *ACM SIGMOD Conference*.
- Kifer, D. and Gehrke, J. (2006), "Injecting Utility into Anonymized Data Sets," *ACM SIGMOD*.
- 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*, 370-374 (http://www.amstat.org/sections/SRMS/Proceedings/papers/1986_069.pdf).
- Kim, J. J. (1990), "Subdomain Estimation for the Masked Data," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 456-461 (http://www.amstat.org/sections/SRMS/Proceedings/papers/1990_075.pdf).
- Kim, J. J., and Winkler, W. E. (1995), "Masking Microdata Files," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 114-119 (http://www.amstat.org/sections/SRMS/Proceedings/papers/1995_017.pdf), longer report (<http://www.census.gov/srd/papers/pdf/rr97-3.pdf>).
- 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*, CD-ROM.
- Lambert, D. (1993), "Measures of Disclosure Risk and Harm," *Journal of Official Statistics*, 9, 313-331 (<http://www.jos.nu/Articles/abstract.asp?article=92313>).
- Lane, J. (2007), "Optimizing the Use of Microdata: An Overview of the Issues," *Journal of Official Statistics*, 23 (3), 299-317 (<http://www.jos.nu/Articles/abstract.asp?article=233299>).
- 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.
- Lakshmanan, L., Ng, R., and Ramesh, G. (2005), "To Do or Not To Do – The Dilemma of Disclosing Anonymized Data," *ACM SIGMOD Conference*, .

- Lakshmanan, L. K. S., Ng, R., Ramesh, G. (2008), "On Disclosure Risk Analysis of Anonymized Itemsets in the Presences of Prior Knowledge," *ACM Transactions on Knowledge Discovery from Data*, 2 (3), 13.1-13.44.
- LeFevre, K., DeWitt, D. and Ramakrishnan, R. (2005), "Incognito: Efficient Full-Domain K-Anonymity," *ACM SIGMOD Conference*, .
- Li, X.-B., and Sarkar, S. (2007), "A Tree-Based Data Perturbation Technique for Privacy-Preserving Data Mining," *IEEE Transactions on Knowledge and Data Engineering*, 18 (9), 1278-1283.
- 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 (<http://www.jos.nu/Articles/abstract.asp?article=92407>).
- Little, R. J. A., and Liu, F. (2002), "Selective Multiple Imputation of Keys for Statistical Disclosure Control in Microdata," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Little, R. J. A., and Liu, F. (2003), "Comparison of SMIke with Data-Swapping and PRAM for Statistical Disclosure Control of Simulated Microdata," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Lindell, Y. and Pinkas, B. (2002), "Privacy Preserving Data Mining," *Proceedings of Crypto 2000*, Springer LNCS 1880, 20-24.
- Liu, H., Kargupta, H., and Ryan, J. (2007), "Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining," *IEEE Transactions of Knowledge and Data Engineering*, 18 (1), 92-106.
- Machanavajjhala, A., Gehrke, and M., Goetz. (2009), "Data Publishing against Realistic Adversaries," VLDB 2009.
- Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramamiam, M. (2005), "l-Diversity: Privacy Beyond k-Anonymity," Cornell CS Dept. technical report, <http://www.cs.cornell.edu/johannes/papers/2005/publishing-icde-final.pdf> .
- Machanavajjhala, A., Kifer, D., Abowd, J. Gehrke, J., and Vilhuber, L. (2008), "Privacy: Theory meets Practice on the Map," ICDE 2008.
- Malin, B. Sweeney, L., and Newton, E. (2003), "Trail Re-identification: Learning Who You are from Where You have Been, Workshop on Privacy in Data, Carnegie-Mellon University, March 2003.
- McSherry, F. (2009), "Privacy Integrated Queries," SIGMOD 2009.
- McSherry, F., and Talwar, K. (2007), "Mechanism design via differential privacy," *Proceedings of the 48th Symposium of the Foundations of Computer Science*.
- Mera, R. (1998), "Matrix Masking Methods That Preserve Moments," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 445-450.
- 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 <http://www.census.gov/srd/www/byyear.html>).
- Moore, A. W., and Lee, M. S. (1998), "Cached Sufficient Statistics for Efficient Machine Learning with Large Datasets," *Journal of Artificial Intelligence Research*, 8, 67-91.
- Motwani, R., and Xu, Y. (2007), "Efficient Algorithms for Masking and Finding Quasi-Identifiers," VLDB '07.
- Müller, W., Blien, U., and Wirth, H. (1995), "Identification Risks of Micro Data," *Evidence from Experimental Studies. Sociological Methods and Research*, 24, 131-157.
- 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., and Sarathy, R. (2003), "A Theoretical Basis for Perturbation Methods," *Statistics and Computing*, 13 (4), 329-335.
- Muralidhar, K., and Sarathy, R. (2006a), "Data Shuffling – A New Masking Approach to Numerical Data," *Management Science*, 52 (5), 658-670.
- Muralidhar, K., and Sarathy, R. (2006b), "A Theoretical Basis for Perturbation Methods," *Journal of Statistics*, 22 (3), 507-524.
- Muralidhar, K. and Sarathy, R. (2007), " 'Easy to Implement' is Putting the Cart before the Horse: Effective Techniques for Masking Numerical Data," Federal Committee on Statistical Methodology Research Conference, to appear on CD-ROMI.

- Muralidhar, K., Sarathy, R., and Dandekar, R. (2006), "Why Swap When You Can Shuffle? A Comparison of Proximity Swap and Data Shuffle for Numeric Data," in (J. Domingo-Ferrer, ed.) *Statistical Data Protection 2006*, Springer: New York, N.Y.
- Muralidhar, K., Sarathy, R., and Parsa, R. (2001) "An Improved Security Requirement for Data Perturbation with Implications for E-Commerce," *Decision Sciences*, 32 (4), 683-698.
- Nissim, K., Raskhodnikova, S., and Smith, A. (2007), "Smooth Sensitivity and Sampling in Private Data Analysis," STOC'07, June 11-13, 2007, San Diego, California, USA.
- Onn, S. (2007), "Entry Uniqueness in Margined Tables," in (J. Domingo-Ferrer, ed.) *Statistical Data Protection 2006*, Springer: New York, N.Y.
- Owen, A. (2003), "Data Squashing by Empirical Likelihood," *Data Mining and Knowledge Discovery*, 7 (1), 101-113.
- Paas, G. (1988), "Disclosure Risk and Disclosure Avoidance for Microdata," *Journal of Business and Economic Statistics*, 6, 487-500.
- Palley, M. A., and Simonoff, J. S. (1987), "The Use of Regression Methodology for the Compromise of Confidential Information in Statistical Databases," *ACM Transactions on Database Systems*, 12 (4), 593-608.
- Polletini, S. (2003), "Maximum Entropy Simulation for Microdata Protection," *Statistics and Computing*, 13 (4), 307-320.
- Polletini, S., Franconi, L., and Stander, J. (2002), "Model Based Disclosure Protection," in (J. Domingo-Ferrer, ed.) *Inference Control in Statistical Databases*: New York: Springer.
- Polletini, S., and Stander, J. (2004), "A Bayesian Hierarchical Model Approach to Risk Estimation in Statistical Disclosure Limitation," in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases*, New York: Springer, 247-261
- Raghunathan, T. E. (2003), Evaluation of Inferences from Multiple Synthetic Data Sets Created Using Semiparametric Approach, Panel on Confidential Data Access for Research Purposes, Committee On National Statistics, October 2003.
- Raghunathan, T. E., Lepkowski, J. M., Hoewyk, J. V., and Sollenberger, P. (1998), "A Multivariate Technique for Multiply Imputing Missing Values Using a Series of Regression Models," Survey Research Center, University of Michigan.
- Raghunathan, T. E., and Reiter, J. P. (2007), "The Multiple Adaptations of Multiple Imputation," *Journal of the American Statistical Association*, 102, 1462-1471.
- Raghunathan, T.E., Reiter, J. P., and Rubin, D.R. (2003), "Multiple Imputation for Statistical Disclosure Limitation," *Journal of Official Statistics*, 19, 1-16.
- Raghunathan, T.E., and Rubin, D.R. (2000), "Multiple Imputation for Disclosure Limitation" University of Michigan, Department of Biostatistics technical report
- Reiss, J.P. (1984), "Practical Data Swapping: The First Steps," *ACM Transactions on Database Systems*, 9, 20-37.
- Reiter, J.P. (2002), "Satisfying Disclosure Restrictions with Synthetic Data Sets," *Journal of Official Statistics*, 18, 531-543.
- Reiter, J.P. (2003a), "Inference for Partially Synthetic, Public Use Data Sets," *Survey Methodology*, 181-189.
- Reiter, J.P. (2003b), Estimating Probabilities of Identification for Microdata, Panel on Confidential Data Access for Research Purposes, Committee On National Statistics, October 2003.
- Reiter, J.P. (2005a), "Releasing Multiply Imputed, Synthetic Public-Use Microdata: An Illustration and Empirical Study," *Journal of the Royal Statistical Society, A*, 168 (1), 185-205.
- Reiter, J. P. (2005b), "Estimating Risk of Identify Disclosure in Microdata," *Journal of the American Statistical Association*, 100, 1103-1112.
- Reiter, J. P. (2008), "Selecting the number of imputed datasets when using multiple imputation for missing data and disclosure protection," *Statistics and Probability Letters*, 78, 15-20.
- Reiter, J. P. and Mitra, R. (2009), "Estimating risks of identification disclosure in partially synthetic data," *Journal of Privacy and Confidentiality*, 01 (01), 99-110.
- 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.
- Samarati, P. (2001), "Protecting Respondents' Identity in Microdata Release," *IEEE Transactions on Knowledge and Data Engineering*, 13 (6), 1010-1027.
- Samarati, P., and Sweeney, L. (1998), "Protecting Privacy when Disclosing Information: k-anonymity and Its Enforcement through Generalization and Cell Suppression," Technical Report, SRI International.

- Sarathy, R., and Muralidhar, K. (2002), "The Security of Confidential Numerical Data in Databases," *Information Systems Research*, 48 (12), 1613-1627.
- Sarathy, R., Muralidhar, K., and Parsa, R. (2002), Perturbing Non-Normal Attributes: The Copula Approach, *Management Science*, 48 (12), 1613-1627
- Scheuren, F., and Winkler, W. E. (1993), "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 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.
- Skinner, C. J., and Elliot, M. A. (2001), "A Measure of Disclosure Risk for Microdata," *Journal of the Royal Statistical Society*, 64 (4), 855-867.
- Skinner, C. J., and Holmes, D. J. Estimating the Re-identification Risk per Record in Microdata, *Journal of Official Statistics*, 14 (1998) 361-372.
- Skinner, C. J. and Shlomo, N. (2007), "Assessing Identification Risk in Survey Data Using Loglinear Models," *Journal of the American Statistical Association*, 103 (483), 989-1001, also at <http://eprints.soton.ac.uk/48103/>.
- 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.
- Sweeney, L. (2002), "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 10 (5), 571-588.
- Sweeney, L. (2004), "Optimal Anonymity using K-similar, a New Clustering Algorithm," manuscript.
- Takemura, A. (2002), "Local Recoding and Record Swapping by Maximum Weight Matching for Disclosure Control of Microdata Sets," *Journal of Official Statistics*, 18 (2), 275-289.
- Tendick, P., and Matloff, N. (1994), "A Modified Random Perturbation Method for Database Security," *ACM Transactions on Database Systems*, 19, 47-63.
- Thibaudeau, Y. (2004), "An Algorithm for Computing Full Rank Minimal Sufficient Statistics with Application to Confidentiality Protection," *UNECE Statistical Journal*, to appear.
- Thibaudeau, Y., and Winkler, W.E. (2002), "Bayesian Networks Representations, Generalized Imputation, and Synthetic Microdata Satisfying Analytic Restraints," Statistical Research Division Report RR 2002/09 at <http://www.census.gov/srd/www/byyear.html>.
- Thibaudeau, Y., and Winkler, W.E. (2004), "Full Rank Minimal Statistics for Disclosure Limitation and Variance Estimation: A Practical Way to Release Count Information," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Torra, V. (2004), "OWA Operators in Data Modeling and Re-identification," *IEEE Transactions on Fuzzy Systems*, 12 (5), 652-660.
- Torra, V., Domingo-Ferrer, J., and Abowd, J. (2007), "Using Mahalanobis-Distance Record Linkage for Disclosure Risk Assessment," in (J. Domingo-Ferrer, ed.) *Statistical Data Protection 2006*, Springer: New York, N.Y.
- Torra, V., and Miyamoto, S. (2004), "Evaluating Fuzzy Clustering Algorithms for Microdata Protection," in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases*, New York: Springer, 175-186.
- Trottini, M., and Fienberg, S. E. (2002), "Modelling User Uncertainty for Disclosure Risk and Data Utility," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 511-528.
- Van Den Hout, A., and Van Der Heijden, P. G. M. (2002), "Randomized Response, Statistical Disclosure Control, and Misclassification: A Review," *International Statistical Review*, 70 (2), 269-288.
- 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, New York: Springer.
- Willenborg, L., and De Waal, T. (2000), *Elements of Statistical Disclosure Control*, Vol. 155, Lecture Notes in Statistics, New York: Springer.

- Willenborg, L. and Van Den Hout, A. (2006), "Peruco: A Method for Producing Safe and Consistent Microdata," *International Statistical Review*, 74 (2), 271-284.
- Winglee, M., Valliant, R., Clark, J., Lim, Y., Weber, M., and Strudler, M. (2002), "Assessing Disclosure Protection for the SOI Public Use File," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, CD-ROM.
- Winkler, W. E. (1994), "Advanced Methods for Record Linkage," *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 467-472 (<http://www.census.gov/srd/papers/pdf/rr94-5.pdf>).
- Winkler, W. E. (1995), "Matching and Record Linkage," in (B. G. Cox et al, ed.) *Business Survey Methods*, New York: J. Wiley, 355-384 (also <http://www.fcsm.gov/working-papers/wwinkler.pdf>).
- Winkler, W. E. (1997), "Views on the Production and Use of Confidential Microdata," Statistical Research Division report RR 97/01 at <http://www.census.gov/srd/www/byyear.html>.
- Winkler, W. E. (1998), "Re-identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata," *Research in Official Statistics*, 1, 87-104, <http://www.census.gov/srd/papers/pdf/rrs2005-09.pdf>.
- Winkler, W. E. (2002a), "Using Simulated Annealing for k-anonymity," Statistical Research Division report RR 2002/07 at <http://www.census.gov/srd/www/byyear.html>.
- Winkler, W. E. (2002b), "Single Ranking Micro-aggregation and Re-identification," Statistical Research Division report RR 2002/08 at <http://www.census.gov/srd/www/byyear.html>.
- Winkler, W. E. (2004a), Masking and Re-identification Methods for Public-Use Microdata: Overview and Research Problems, in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Database*, Springer: New York, 231-247, also <http://www.census.gov/srd/papers/pdf/rrs2004-06.pdf>.
- Winkler, W. E. (2004b), Re-identification Methods for Masked Microdata, in (J. Domingo-Ferrer and V. Torra, eds.), *Privacy in Statistical Databases*, Springer: New York, 216-230, also <http://www.census.gov/srd/papers/pdf/rrs2004-03.pdf>.
- Winkler, W. E. (2005), "Modeling and Quality of Masked Microdata," *American Statistical Association, Proceedings of the Section on Survey Research Method*, CD-ROM, also <http://www.census.gov/srd/papers/pdf/rrs2006-01.pdf>.
- Winkler, W. E. (2007a), "Analytically Valid Discrete Microdata and Re-identification," <http://www.census.gov/srd/papers/pdf/rrs2007-19.pdf>.
- Winkler, W. E. (2007b), "Examples of Easy-to-implement, Widely Used Masking Methods for which Analytic Properties are not Justified," <http://www.census.gov/srd/papers/pdf/rrs2007-21.pdf>.
- Winkler, W. E. (2008), "General Discrete-data Modeling Methods for Producing Synthetic Data with Reduced Re-identification Risk that Preserve Analytic Properties," IAB Workshop on Confidentiality and Disclosure, http://fdz.iab.de/en/FDZ_Events/SDC-Workshop.aspx, Nuremberg, Germany, November 20-21, 2008 (also <http://www.census.gov/srd/papers/pdf/rrs2010-02>).
- Winkler, W. E. (2009), "Should Social Security numbers be replaced by modern, more secure identifiers?" *Proceedings of the National Academy of Science*.
- Winkler, W. E. (2010), "General Discrete-data Modeling Methods for Creating Synthetic Data with Reduce Re-identification Risk that Preserve Analytic Properties," <http://www.census.gov/srd/papers/pdf/rrs2010-02.pdf>.
- Woo, M., Reiter, J. P., Oganian, A., Karr, A. F. (2009) "Global measures of data utility for microdata masked for disclosure limitation," *Journal of Privacy and Confidentiality*, 01 (01), 111-124.
- Xiao, X., and Tao, Y. (2007a), "m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets," *ACM SIGMOD*, 689-700.
- Xiao, X., and Tao, Y. (2007b), "Anatomy: Simple and Effective Privacy Preservation," *VLDB*, 139-150.
- Xiao, X., and Tao, Y. (2008a), "Dynamic Anonymization: Accurate Statistical Analysis with Privacy Preservation," *ACM SIGMOD*, 107-120.
- Xiao, X., and Tao, Y. (2008b), "Output Perturbation with Query Relaxation," *VLDB*, 857-869.
- 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*, New York: Springer, 135-151, (also <http://www.census.gov/srd/papers/pdf/rrs2002-01.pdf>).
- Yang, Z., Zhong, S., and Wright, R. (2005), "Anonymity Preserving Data Collection," *ACM KDD Conference*, 334-343.
- Zhang, N., Wang, S., and Zhao, W. (2005), "A New Scheme on Privacy-Preserving Data Classification," *ACM KDD Conference*, 374-383.
- Zhong, S., Yang, Z., and Wright, R. (2005), "Privacy-Enhancing k-Anonymization of Customer Data," *ACM Principals of Database Systems Conference 2005*, 139-147.
- Zhu, Y., and Liu, L. (2004), "Optimal Randomization for Privacy Preserving Data Mining," *ACM Knowledge Discovery and Data Mining Conference 2004*, 761-765.