STATISTICAL SCIENCE

Volume 32, Number 2 May 2017

Special Section on Complex Surveys

Introduction to the Design and Analysis of Complex Survey Data ..... *Chris Skinner and Jon Wakefield* 165
Probability Sampling Designs: Principles for Choice of Design and Balancing ..... *Yves Tillé and Matthieu Wilhelm* 176
Model-Assisted Survey Estimation with Modern Prediction Techniques ..... *F. Jay Breidt and Jean D. Opsomer* 190
Construction of Weights in Surveys: A Review ..... *David Haziza and Jean-François Beaumont* 206
Approaches to Improving Survey-Weighted Estimates ..... *Qixuan Chen, Michael R. Elliott, David Haziza, Ye Yang, Malay Ghosh, Roderick J. A. Little, Joseph Sedransk and Mary Thompson* 227
Inference for Nonprobability Samples ..... *Michael R. Elliott and Richard Valliant* 249
Fitting Regression Models to Survey Data ..... *Thomas Lumley and Alastair Scott* 265
Options for Conducting Web Surveys ..... *Matthias Schonlau and Mick P. Couper* 279
Combining Survey Data with Other Data Sources ..... *Sharon L. Lohr and Trivellore E. Raghunathan* 293

General Section

J. B. S. Haldane's Contribution to the Bayes Factor Hypothesis Test ..... *Alexander Etz and Eric-Jan Wagenmakers* 313

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 32, Number 2, May 2017. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, OH 44122, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike—Suite L2310, Bethesda, MD 20814-3998, USA.

Statistical Science
2017, Vol. 32, No. 2, 165–175
DOI: 10.1214/17-STS614

Introduction to the Design and Analysis of Complex Survey Data

Chris Skinner and Jon Wakefield

Abstract. We give a brief overview of common sampling designs used in a survey setting, and introduce the principal inferential paradigms under which data from complex surveys may be analyzed. In particular, we distinguish between design-based, model-based and model-assisted approaches. Simple examples highlight the key differences between the approaches. We discuss the interplay between inferential approaches and targets of inference and the important issue of variance estimation.

Key words and phrases: Design-based inference, model-assisted inference, model-based inference, weights, variance estimation.

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling, part I. In *Foundations of Statistical Inference* (*Proc. Sympos., Univ. Waterloo, Waterloo, Ont.*, 1970) 203–242. Holt, Rinehart and Winston, Toronto. MR0423625

BERGER, Y. G. and TILLÉ, Y. (2009). Sampling with unequal probabilities. In *Handbook of Statistics*, *Vol. 29A*, *Sample Surveys*: *Design, Methods and Applications* (D. Pfeffermann and C. R. Rao, eds.) 39–54. North-Holland, Amsterdam. MR2654632

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144

BREIDT, J. and OPSOMER, J. (2017). Model-assisted survey estimation with modern prediction techniques. *Statist. Sci.* **32** 190–205.

BREWER, K. (2002). *Combined Survey Sampling Inference*: *Weighing Basu's Elephants*. Arnold, London.

CHAMBERS, R. L. and CLARK, R. G. (2012). *An Introduction to Model-Based Survey Sampling with Applications*. Oxford Univ. Press, Oxford. MR3186498

CHEN, Q., ELLIOTT, M. R., HAZIZA, D., YANG, Y., GHOSH, M., LITTLE, R., SEDRANSK, J. and THOMPSON, M. (2017). Approaches to improving survey-weighted estimates. *Statist. Sci.* **32** 227–248.

COX, D. R. (2006). *Principles of Statistical Inference*. Cambridge Univ. Press, Cambridge. MR2278763

ELLIOTT, M. and VALLIANT, R. (2017). Inference for nonprobability samples. *Statist. Sci.* **32** 249–264.

GELMAN, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.* **22** 153–164. MR2408951

GRAUBARD, B. I. and KORN, E. L. (2002). Inference for superpopulation parameters using sample surveys. *Statist. Sci.* **17** 73–96. MR1910075

HÁJEK, J. (1971). Discussion of 'An essay on the logical foundations of survey sampling, part I', by D. Basu. In *Foundations of Statistical Inference* (*Proc. Sympos., Univ. Waterloo, Waterloo, Ont.*, 1970). Holt, Rinehart and Winston, Toronto.

HAZIZA, D. and BEAUMONT, J.-F. (2017). Construction of weights in surveys: A review. *Statist. Sci.* **32** 206–226.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

JAPEC, L., KREUTER, F., BERG, M., BIEMER, P., DECKER, P., LAMPE, C., LANE, J., O'NEIL, C. and ASHER, A. (2015). Big data in survey research: Aapor task force report. *Public Opin. Q.* **79** 839–880.

LITTLE, R. J. (2013). Calibrated Bayes, an alternative inferential paradigm for official statistics (with discussion). *J. Off. Stat.* **28** 309–372.

LOHR, S. L. (2010). *Sampling*: *Design and Analysis*, 2nd ed. Brooks/Cole Cengage Learning, Boston, MA.

LOHR, S. and RAGHUNATHAN, T. (2017). Combining survey data with other data sources. *Statist. Sci.* **32** 293–312.

LUMLEY, T. (2010). *Complex Surveys*: *A Guide to Analysis Using R*. Wiley, Hoboken, NJ.

LUMLEY, T. and SCOTT, A. (2017). Fitting regression models to survey data. *Statist. Sci.* **32** 265–278.

PFEFFERMANN, D., SKINNER, C. J., HOLMES, D. J., GOLDSTEIN, H. and RASBASH, J. (1998). Weighting for unequal selection probabilities in multilevel models. *J. R. Stat. Soc. Ser. B* **60** 23–40. MR1625668

Chris Skinner is Professor, Department of Statistics, London School of Economics and Political Science, London, United Kingdom (e-mail: c.j.skinner@lse.ac.uk). Jon Wakefield is Professor, Departments of Statistics and Biostatistics, University of Washington, Seattle, Washington, USA (e-mail: jonno@uw.edu).

RABE-HESKETH, S. and SKRONDAL, A. (2006). Multilevel modelling of complex survey data. *J. Roy. Statist. Soc. Ser. A* **169** 805–827. MR2291345

RUST, K. F. and RAO, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Stat. Methods Med. Res.* **5** 283–310.

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SCHONLAU, M. and COUPER, M. (2017). Options for conducting web surveys. *Statist. Sci.* **32** 279–292.

SCOTT, A. and SMITH, T. M. F. (1969). Estimation in multi-stage surveys. *J. Amer. Statist. Assoc.* **64** 830–840.

SEAMAN, S. R. and WHITE, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **22** 278–295.

SHAO, J. and TU, D. (2012). *The Jackknife and Bootstrap*. Springer, New York. MR1351010

SKINNER, C. J. (2003). Introduction to part b. In *Analysis of Survey Data* (R. L. Chamber and C. J. Skinner, eds.) 75–84. Wiley, Chichester. MR1978845

SKINNER, C. J., HOLT, D. and SMITH, T. M. F., eds. (1989). *Analysis of Complex Surveys*. Wiley, Chichester. MR1049386

TILLÉ, Y. and WILHELM, M. (2017). Probability sampling designs; principles for the choice of design and balancing. *Statist. Sci.* **32** 176–189.

VALLIANT, R., DEVER, J. A. and KREUTER, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, Berlin. MR3088726

VALLIANT, R., DORFMAN, A. H. and ROYALL, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley-Interscience, New York. MR1784794

Statistical Science
2017, Vol. 32, No. 2, 176–189
DOI: 10.1214/16-STS606

Probability Sampling Designs: Principles for Choice of Design and Balancing

Yves Tillé and Matthieu Wilhelm

Abstract. The aim of this paper is twofold. First, three theoretical principles are formalized: randomization, overrepresentation and restriction. We develop these principles and give a rationale for their use in choosing the sampling design in a systematic way. In the model-assisted framework, knowledge of the population is formalized by modelling the population and the sampling design is chosen accordingly. We show how the principles of overrepresentation and of restriction naturally arise from the modelling of the population. The balanced sampling then appears as a consequence of the modelling. Second, a review of probability balanced sampling is presented through the model-assisted framework. For some basic models, balanced sampling can be shown to be an optimal sampling design. Emphasis is placed on new spatial sampling methods and their related models. An illustrative example shows the advantages of the different methods. Throughout the paper, various examples illustrate how the three principles can be applied in order to improve inference.

Key words and phrases: Balanced sampling, design-based, model-based, inference, entropy, pivotal method, cube method, spatial sampling.

REFERENCES

ANTAL, E. and TILLÉ, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *J. Amer. Statist. Assoc.* **106** 534–543. MR2847968

BERGER, Y. G. (1998a). Rate of convergence to normal distribution for the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **67** 209–226. MR1624693

BERGER, Y. G. (1998b). Rate of convergence for asymptotic variance for the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **74** 149–168. MR1665125

BERGER, Y. G. and DE LA RIVA TORRES, O. (2016). Empirical likelihood confidence intervals for complex sampling designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 319–341. MR3454199

BREIDT, F. J. and CHAUVET, G. (2012). Penalized balanced sampling. *Biometrika* **99** 945–958. MR2999170

BREIDT, F. J. and OPSOMER, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statist. Sci.* **32** 190–205.

BREWER, K. R. W. (1963). Ratio estimation in finite populations: Some results deductible from the assumption of an underlying stochastic process. *Aust. J. Stat.* **5** 93–105.

BREWER, K. R. W. and DONADIO, M. E. (2003). The high entropy variance of the Horvitz–Thompson estimator. *Surv. Methodol.* **29** 189–196.

BREWER, K. R. W. and HANIF, M. (1983). *Sampling with Unequal Probabilities. Lecture Notes in Statistics* **15**. Springer, New York. MR0681289

CARDOT, H. and JOSSERAND, E. (2011). Horvitz–Thompson estimators for functional data: Asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika* **98** 107–118. MR2804213

CHAUVET, G. (2012). On a characterization of ordered pivotal sampling. *Bernoulli* **18** 1320–1340. MR2995798

CHAUVET, G., BONNÉRY, D. and DEVILLE, J. C. (2011). Optimal inclusion probabilities for balanced sampling. *J. Statist. Plann. Inference* **141** 984–994. MR2732966

CHAUVET, G. and TILLÉ, Y. (2005). Fast SAS macros for balancing samples: User's guide. Software manual, Univ. Neuchâtel. Available at http://www2.unine.ch/statistics/page10890.html.

CHEN, X. (1993). Poisson-Binomial distribution, conditional Bernoulli distribution and maximum entropy. Technical report, Dept. Statistics, Harvard Univ.

CHEN, S. X., DEMPSTER, A. P. and LIU, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika* **81**

Yves Tillé is Professor, Institut de Statistique, Université de Neuchâtel, Av. de Bellevaux 51, 2000 Neuchâtel, Switzerland (e-mail: yves.tille@unine.ch). Matthieu Wilhelm is Ph.D. Student, Institut de Statistique, Université de Neuchâtel, Av. de Bellevaux 51, 2000 Neuchâtel, Switzerland (e-mail: matthieu.wilhelm@unine.ch).

457–469. MR1311090
COCHRAN, W. G. (1977). *Sampling Techniques*. Wiley, New York.
DEVILLE, J. C. (2000). Note sur l'algorithme de Chen, Dempster et Liu. Technical report, CREST-ENSAI, Rennes.
DEVILLE, J. C. (2014). Échantillonnage équilibré exact Poissonnien. In 8*ème Colloque Francophone sur les Sondages* 1–6. Société française de statistique, Dijon.
DEVILLE, J. C. and SÄRNDAL, C. E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804
DEVILLE, J. C. and TILLÉ, Y. (2000). Selection of several unequal probability samples from the same population. *J. Statist. Plann. Inference* **86** 215–227. MR1763190
DEVILLE, J.-C. and TILLÉ, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika* **91** 893–912. MR2126040
DEVILLE, J.-C. and TILLÉ, Y. (2005). Variance approximation under balanced sampling. *J. Statist. Plann. Inference* **128** 569–591. MR2102777
FALORSI, P. D. and RIGHI, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Surv. Methodol.* **34** 223–234.
FALORSI, P. D. and RIGHI, P. (2016). A unified approach for defining optimal multivariate and multi-domains sampling designs. In *Topics in Theoretical and Applied Statistics* 145–152. Springer, New York.
FULLER, W. A. (1970). Sampling with random stratum boundaries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **32** 209–226. MR0272140
GODAMBE, V. P. and JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations I. *Ann. Math. Stat.* **36** 1707–1722. MR0240915
GRAFSTRÖM, A. (2010). Entropy of unequal probability sampling designs. *Stat. Methodol.* **7** 84–97. MR2591712
GRAFSTRÖM, A. and LISIC, J. (2016). BalancedSampling: Balanced and spatially balanced sampling. R package version 1.5.2.
GRAFSTRÖM, A. and LUNDSTRÖM, N. L. P. (2013). Why well spread probability samples are balanced? *Open J. Stat.* **3** 36–41.
GRAFSTRÖM, A., LUNDSTRÖM, N. L. P. and SCHELIN, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics* **68** 514–520. MR2959618
GRAFSTRÖM, A. and TILLÉ, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics* **24** 120–131. MR3067336
HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Čas. Pěst. Mat.* **84** 387–423.
HÁJEK, J. (1981). *Sampling from a Finite Population*. Dekker, New York.
HODGES, J. L. JR. and LE CAM, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Stat.* **31** 737–740. MR0117812
HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460
JAY BREIDT, F. and CHAUVET, G. (2011). Improved variance estimation for balanced samples drawn via the cube method. *J. Statist. Plann. Inference* **141** 479–487. MR2719512
JESSEN, R. J. (1978). *Statistical Survey Techniques*. Wiley, New York.
JIANG, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, New York.
KINCAID, T. M. and OLSEN, A. R. (2015). spsurvey: Spatial survey design and analysis. R package version 3.1.
KRUSKAL, W. and MOSTELLER, F. (1979a). Representative sampling, I: Nonscientific literature. *Int. Stat. Rev.* **47** 13–24.
KRUSKAL, W. and MOSTELLER, F. (1979b). Representative sampling, II: Scientific literature, excluding statistics. *Int. Stat. Rev.* **47** 111–127.
KRUSKAL, W. and MOSTELLER, F. (1979c). Representative sampling, III: The current statistical literature. *Int. Stat. Rev.* **47** 245–265.
KRUSKAL, W. and MOSTELLER, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895–1939. *Int. Stat. Rev.* **48** 169–195.
LAPLACE, P. S. (1847). *Théorie Analytique des Probabilités*. Imprimerie royale, Paris.
LEGG, J. C. and YU, C. L. (2010). Comparison of sample set restriction procedures. *Surv. Methodol.* **36** 69–79.
LOHR, S. (2009). *Sampling*: *Design and Analysis*. Brooks/Cole, Boston.
MARKER, D. A. and STEVENS, D. L. JR. (2009). Sampling and inference in environmental surveys. In *Sample Surveys*: *Design, Methods and Applications. Handbook of Statistics* **29** 487–512. Elsevier/North-Holland, Amsterdam.
MØLLER, J. and WAAGEPETERSEN, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, London.
NARAIN, R. D. (1951). On sampling without replacement with varying probabilities. *J. Indian Soc. Agricultural Statist.* **3** 169–174. MR0045354
NEDYALKOVA, D. and TILLÉ, Y. (2008). Optimal sampling and estimation strategies under linear model. *Biometrika* **95** 521–537. MR2443173
NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. R. Stat. Soc.* **97** 558–606. MR0121942
NEYMAN, J. (1938). Contribution to the theory of sampling human population. *J. Amer. Statist. Assoc.* **33** 101–116.
R DEVELOPMENT CORE TEAM (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
ROUSSEAU, S. and TARDIEU, F. (2004). La macro SAS CUBE d'échantillonnage équilibré, Documentation de l'utilisateur. Technical report, Insee, Paris.
ROYALL, R. M. (1970a). Finite population sampling—On labels in estimation. *Ann. Math. Stat.* **41** 1774–1779. MR0282493
ROYALL, R. M. (1970b). On finite population sampling theory under certain linear regression models. *Biometrika* **57** 377–387.
ROYALL, R. M. (1976a). Likelihood functions in finite population sampling theory. *Biometrika* **63** 605–614. MR0652530
ROYALL, R. M. (1976b). The linear least-squares prediction approach to two-stage sampling. *J. Amer. Statist. Assoc.* **71** 657–664. MR0451484
ROYALL, R. M. (1992). The model based (prediction) approach to finite population sampling theory. In *Current Issues in Statistical Inference*: *Essays in Honor of D. Basu* (M. Ghosh and P. K. Pathak, eds.). *Lecture Notes—Monograph Series* **17** 225–240. IMS, Hayward, CA. MR1194420
ROYALL, R. M. and HERSON, J. (1973a). Robust estimation in finite populations I. *J. Amer. Statist. Assoc.* **68** 880–889. MR0386087

ROYALL, R. M. and HERSON, J. (1973b). Robust estimation in finite populations. II. Stratification on a size variable. *J. Amer. Statist. Assoc.* **68** 890–893. MR0386088

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SEN, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agricultural Statist.* **5** 119–127.

STEIN, C. (1990). Application of Newton's identities to a generalized birthday problem and to the Poisson-Binomial distribution. Technical Report TC 354, Dept. Statistics, Stanford Univ., Stanford, CA.

STEVENS, D. L. JR. and OLSEN, A. R. (1999). Spatially restricted surveys over time for aquatic resources. *J. Agric. Biol. Environ. Stat.* **4** 415–428. MR1816377

STEVENS, D. L. JR. and OLSEN, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics* **14** 593–610.

STEVENS, D. L. JR. and OLSEN, A. R. (2004). Spatially balanced sampling of natural resources. *J. Amer. Statist. Assoc.* **99** 262–278. MR2061889

SUKHATME, P. V. (1954). *Sampling Theory of Surveys with Applications*. The Indian Society of Agricultural Statistics, New Delhi, India; The Iowa State College Press, Ames, Iowa. MR0062991

THOMPSON, S. K. (2012). *Sampling*, 3rd ed. Wiley, Hoboken, NJ. MR2894042

TILLÉ, Y. (2006). *Sampling Algorithms*. Springer, New York. MR2225036

TILLÉ, Y. and FAVRE, A. C. (2004). Co-ordination, combination and extension of optimal balanced samples. *Biometrika* **91** 913–927.

TILLÉ, Y. and FAVRE, A. C. (2005). Optimal allocation in balanced sampling. *Statist. Probab. Lett.* **74** 31–37. MR2189073

TILLÉ, Y. and MATEI, A. (2015). sampling: Survey sampling. R package version 2.7.

VALLIANT, R., DEVER, J. A. and KREUTER, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York. MR3088726

VALLIANT, R., DORFMAN, A. H. and ROYALL, R. M. (2000). *Finite Population Sampling and Inference*: *A Prediction Approach*. Wiley-Interscience, New York. MR1784794

YATES, F. and GRUNDY, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **15** 235–261.

Statistical Science
2017, Vol. 32, No. 2, 190–205
DOI: 10.1214/16-STS589

Model-Assisted Survey Estimation with Modern Prediction Techniques

F. Jay Breidt and Jean D. Opsomer

Abstract. This paper reviews the design-based, model-assisted approach to using data from a complex survey together with auxiliary information to estimate finite population parameters. A general recipe for deriving model-assisted estimators is presented and design-based asymptotic analysis for such estimators is reviewed. The recipe allows for a very broad class of prediction methods, with examples from the literature including linear models, linear mixed models, nonparametric regression and machine learning techniques.

Key words and phrases: Machine learning, nonparametric regression, nearest neighbors, neural network, regression trees, survey asymptotics.

REFERENCES

ARAGON, Y., GOGA, C. and RUIZ-GAZEN, A. (2006). Estimation non-paramétrique de quantiles en présence d'information auxiliaire. In *Méthodes D'Enquêtes et Sondages*. *Pratiques Européenne et Nord-américaine* (P. Lavellée and L.-P. Rivest, eds.) 377–382. Dunod, Paris.

BAFFETTA, F., CORONA, P. and FATTORINI, L. (2010). Design-based diagnostics for k-NN estimators of forest resources. *Can. J. For. Res*. **41** 59–72.

BAFFETTA, F., FATTORINI, L., FRANCESCHI, S. and CORONA, P. (2009). Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. *Remote Sens. Environ*. **113** 463–475.

BARDSLEY, P. and CHAMBERS, R. L. (1984). Multipurpose estimation from unbalanced samples. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **33** 290–299.

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc*. **83** 28–36.

BEAUMONT, J. F. and BOCCI, C. (2008). Another look at ridge calibration. *Metron* **66** 5–20.

BEAUMONT, J.-F., HAZIZA, D. and RUIZ-GAZEN, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika* **100** 555–569.

BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist*. **12** 470–482. MR0740906

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144

BREIDT, F. J., CLAESKENS, G. and OPSOMER, J. D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika* **92** 831–846. MR2234189

BREIDT, F. J. and OPSOMER, J. D. (2000). Local polynomial regression estimators in survey sampling. *Ann. Statist*. **28** 1026–1053. MR1810918

BREIDT, F. J. and OPSOMER, J. D. (2008). Endogenous post-stratification in surveys: Classifying with a sample-fitted model. *Ann. Statist*. **36** 403–427. MR2387977

BREIDT, F. J., OPSOMER, J. D. and SANCHEZ-BORREGO, I. (2016). Nonparametric variance estimation under fine stratification: An alternative to collapsed strata. *J. Amer. Statist. Assoc*. **111** 822–833. MR3538708

BREIDT, F. J., OPSOMER, J. D., JOHNSON, A. A. and RANALLI, M. G. (2007). Semiparametric model-assisted estimation for natural resource surveys. *Surv. Methodol*. **33** 35–44.

BREIMAN, L. (1996). Bagging predictors. *Mach. Learn*. **24** 123–140.

BREIMAN, L. (2001). Random forests. *Mach. Learn*. **45** 5–32.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

CARDOT, H., GOGA, C. and LARDIN, P. (2013). Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electron. J. Stat*. **7** 562–596. MR3035266

CARDOT, H. and JOSSERAND, E. (2011). Horvitz–Thompson estimators for functional data: Asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika* **98** 107–118. MR2804213

F. Jay Breidt is Professor, Department of Statistics, Colorado State University, Fort Collins, Colorado 80523-1877, USA (e-mail: jbreidt@stat.colostate.edu). Jean D. Opsomer is Professor, Department of Statistics, Colorado State University, Fort Collins, Colorado 80523-1877, USA (e-mail: jopsomer@stat.colostate.edu).

CASSEL, C. M., SÄRNDAL, C. E. and WRETMAN, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63** 615–620. MR0445666

CHAMBERS, R. L. (1996). Robust case-weighting for multipurpose establishment surveys. *J. Off. Stat.* **12** 3–32.

COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York. MR0474575

DAHLKE, M., BREIDT, F. J., OPSOMER, J. D. and VAN KEILEGOM, I. (2013). Nonparametric endogenous post-stratification estimation. *Statist. Sinica* **23** 189–211. MR3076164

DATTA, G. S. and GHOSH, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist.* **19** 1748–1770. MR1135147

DEVILLE, J.-C. and GOGA, C. (2004). Estimation par régression par polynômes locaux dans des enquêtes sur plusieurs échantillons. In *Echantillonnage et Méthodes D'Enquêtes* (P. Ardilly, ed.) 156–162. Dunod, Paris.

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382.

ELLIOTT, M. R. and LITTLE, R. J. A. (2000). Model-based alternatives to trimming survey weights. *J. Off. Stat.* **16** 191–209.

FAY, R. E. and HERRIOT, R. A. (1979). Estimation of income from small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277.

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines (with discussion). *Ann. Statist.* **19** 1–141. MR1091842

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

FULLER, W. A. (2002). Regression estimation for survey samples (with discussion). *Surv. Methodol.* **28** 5–23.

GHOSH, M. and RAO, J. N. K. (1994). Small area estimation: An appraisal. *Statist. Sci.* **9** 55–93. MR1278679

GOGA, C. (2004). Estimation de l'évolution d'un total en présence d'information auxiliaire: Une approche par splines de régression. *C. R. Math. Acad. Sci. Paris* **339** 441–444.

GOGA, C. (2005). Réduction de la variance dans les sondages en présence d'information auxiliaire: Une approche non paramétrique par splines de régression. *Canad. J. Statist.* **33** 163–180. MR2193026

GUGGEMOS, F. and TILLÉ, Y. (2010). Penalized calibration in survey sampling: Design-based estimation assisted by mixed models. *J. Statist. Plann. Inference* **140** 3199–3212. MR2659847

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 361–374. MR0125612

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning*: *Data Mining*, *Inference and Prediction*. Springer, New York.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685.

ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029

KREWSKI, D. and RAO, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9** 1010–1019. MR0628756

LAZZERONI, L. C. and LITTLE, R. J. A. (1998). Random-effects models for smoothing poststratification weights. *J. Off. Stat.* **14** 61–78.

LI, X. and OPSOMER, J. D. (2006). Model averaging in survey estimation. In *Proceedings of the Section on Survey Research Methods*. Amer. Statist. Assoc., Alexandria, VA.

MCCONVILLE, K. (2011). Improved Estimation for Complex Surveys Using Modern Regression Techniques. Ph.D. thesis, Colorado State University.

MCCONVILLE, K. S. and BREIDT, F. J. (2013). Survey design asymptotics for the model-assisted penalised spline regression estimator. *J. Nonparametr. Stat.* **25** 745–763. MR3174295

MCROBERTS, R. E., NÆSSET, E. and GOBAKKEN, T. (2013). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sens. Environ.* **128** 268–275.

MCROBERTS, R. E., TOMPPO, E. O. and NÆSSET, E. (2010). Advances and emerging issues in national forest inventories. *Scand. J. For. Res.* **25** 368–381.

MONTANARI, G. E. and RANALLI, M. G. (2005). Nonparametric model calibration estimation in survey sampling. *J. Amer. Statist. Assoc.* **100** 1429–1442.

MONTANARI, G. E. and RANALLI, M. G. (2009). Multiple and ridge model calibration. In *Proceedings of Workshop on Calibration and Estimation in Surveys*. Statistics Canada, Ottawa, ON.

NÆSSET, E., BOLLANDSÅS, O. M., GOBAKKEN, T., GREGOIRE, T. G. and STÅHL, G. (2013). Model-assisted estimation of change in forest biomass over an 11 year period in a sample survey supported by airborne lidar: A case study with post-stratification to provide "activity data". *Remote Sens. Environ.* **128** 299–314.

OPSOMER, J. D., BREIDT, F. J., MOISEN, G. G. and KAUERMANN, G. (2007). Model-assisted estimation of forest resources with generalized additive models. *J. Amer. Statist. Assoc.* **102** 400–409. MR2370838

OPSOMER, J. D., CLAESKENS, G., RANALLI, M. G., KAUERMANN, G. and BREIDT, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 265–286. MR2412642

PARK, M. and FULLER, W. A. (2005). Towards nonnegative regression weights for survey samples. *Surv. Methodol.* **31** 85–93.

PARK, M. and FULLER, W. A. (2009). The mixed model for survey regression estimation. *J. Statist. Plann. Inference* **139** 1320–1331.

RAO, J. N. K. (2003). *Small Area Estimation*. Wiley-Interscience, New York.

RAO, J. N. K. and SINGH, A. C. (1997). A ridge-shrinkage method for range-restricted weight calibration in survey sampling (Pkg: P57-85). In *ASA Proceedings of the Section on Survey Research Methods* 57–65. Amer. Statist. Assoc., Alexandria, VA.

ROBINSON, P. and SÄRNDAL, C. E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya, Ser. B* **45** 240–248. MR0748468

RUEDA, M., SÁNCHEZ-BORREGO, I. and ARCOS, A. (2009). Mean estimation in the presence of change points. *Appl. Math. Lett.* **22** 1257–1261. MR2532550

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720

SÁNCHEZ-BORREGO, I., RUEDA, M. and MUÑOZ, J. (2012). Nonparametric methods in sample surveys. Application to the estimation of cancer prevalence. *Qual. Quant.* **46** 405–414.

SÄRNDAL, C.-E. (2010). The calibration approach in survey theory and practice. *Surv. Methodol.* **33** 99–119.

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SILVA, P. N. and SKINNER, C. J. (1997). Variable selection for regression estimation in finite populations. *Surv. Methodol.* **23** 23–32.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242

TIPTON, J., OPSOMER, J. and MOISEN, G. (2013). Properties of endogenous post-stratified estimation using remote sensing data. *Remote Sens. Environ.* **139** 130–137.

TOTH, D. and ELTINGE, J. L. (2011). Building consistent regression trees from complex sample data. *J. Amer. Statist. Assoc.* **106** 1626–1636. MR2896862

WANG, L. (2009). Single-index model-assisted estimation in survey sampling. *J. Nonparametr. Stat.* **21** 487–504. MR2571724

WANG, J. C., OPSOMER, J. D. and WANG, H. (2014). Bagging non-differentiable estimators in complex surveys. *Surv. Methodol.* **40** 189–209.

WANG, L. and WANG, S. (2011). Nonparametric additive model-assisted estimation for survey data. *J. Multivariate Anal.* **102** 1126–1140. MR2805653

WU, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika* **90** 937–951. MR2024768

WU, C. F. J. and SITTER, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.* **96** 185–193. MR1952731

ZHENG, H. and LITTLE, R. J. A. (2003). Penalized spline model-based estimation of finite population total from probability-proportional-to-size samples. *J. Off. Stat.* **19** 99–117.

ZHENG, H. and LITTLE, R. J. A. (2004). Penalized spline nonparametric mixed models for inference about a finite population mean from two-stage samples. *Surv. Methodol.* **30** 209–218.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

Statistical Science
2017, Vol. 32, No. 2, 206–226
DOI: 10.1214/16-STS608

Construction of Weights in Surveys: A Review

David Haziza and Jean-François Beaumont

Abstract. Weighting is one of the central steps in surveys. The typical weighting process involves three major stages. At the first stage, each unit is assigned a base weight, which is defined as the inverse of its inclusion probability. The base weights are then modified to account for unit nonresponse. At the last stage, the nonresponse-adjusted weights are further modified to ensure consistency between survey estimates and known population totals. When needed, the weights undergo a last modification through weight trimming or weight smoothing methods in order to improve the efficiency of survey estimates. This article provides an overview of the various stages involved in the typical weighting process used by national statistical offices.

Key words and phrases: Calibration estimator, design-based framework, expansion estimator, propensity score adjusted estimator, unequal probability sampling, unit nonresponse, weight smoothing, weight trimming, weighting system.

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling, Part I. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 203–242. Holt, Rinehart and Winston, Toronto.

BEAUMONT, J.-F. (2005a). Calibrated imputation in surveys under a quasi-model-assisted approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 445–458. MR2155348

BEAUMONT, J.-F. (2005b). On the use of data collection process information for the treatment of unit nonreponse through weight adjustment. *Surv. Methodol.* **31** 227–231.

BEAUMONT, J.-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika* **95** 539–553. MR2443174

BEAUMONT, J.-F. and BOCCI, C. (2008). Another look at ridge calibration. *Metron—International Journal of Statistics* **LXIV** 5–20.

BEAUMONT, J.-F. and HAZIZA, D. (2016). A note on the concept of invariance in two-phase sampling designs. *Surv. Methodol.* **42** 319–323.

BREIDT, F. J. and OPSOMER, J. D. (2016). Model-assisted survey estimation with modern prediction techniques. *Statist. Sci.* **32** 190–205.

BREWER, K. R. W. (2002). *Combined Survey Sampling Inference: Weighing Basu's Elephants*. Arnold, London.

CHAMBERS, R. L. (1996). Robust case-weighting for multipurpose establishment surveys. *J. Off. Stat.* **12** 3–32.

CHANG, T. and KOTT, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95** 555–571. MR2443175

CHEN, Q., ELLIOT, M. R., HAZIZA, D., YANG, Y., GOSH, M., LITTLE, R., SEDRANSK, J. and THOMPSON, M. (2016). Weights and estimation of a survey population mean: A review. *Statist. Sci.* To appear.

CLOUTIER, E. and LANGLET, É. (2014). Aboriginal Peoples Survey, 2012: Concepts and methods guide. Statistics Canada, Catalogue No. 89-653-X—No. 002, ISBN 978-1-100-22738-2.

DA SILVA, D. N. and OPSOMER, J. D. (2006). A kernel smoothing method of adjusting for unit non-response in sample surveys. *Canad. J. Statist.* **34** 563–579. MR2347046

DA SILVA, D. N. and OPSOMER, J. D. (2009). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *Surv. Methodol.* **35** 165–176.

DEMING, W. E. and STEPHAN, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11** 427–444. MR0003527

DEVILLE, J.-C. (1998). La correction de la non-réponse par calage ou par échantillonnage équilibré. In *Actes du Colloque* de la Société Statistique du, Canada, Sherbrooke, Canada.

David Haziza is Professor of Statistics, Department of Mathematics and Statistics, Université de Montréal, C.P. 6128 Succ. Centre-Ville, Montreal, Quebec, H3J 3C7 (e-mail: david.haziza@umontreal.ca). Jean-François Beaumont is Chief, International Cooperation and Corporate Statistical Methods Division, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6 (e-mail: jean-francois.beaumont@canada.com).

DEVILLE, J.-C. (2002). La correction de la non-réponse par calage généralisé. *Actes des Journées de Méthodologie Statistique*, *Insee*.

DEVILLE, J.-C. and LAVALLÉE, P. (2006). Indirect sampling: The foundation of the generalized weight share method. *Surv. Methodol*. **32** 165–176.

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc*. **87** 376–382. MR1173804

DEVILLE, J.-C., SÄRNDAL, C.-E. and SAUTORY, O. (1993). Generalized raking procedure in survey sampling. *J. Amer. Statist. Assoc*. **88** 1013–1020.

EKHOLM, A. and LAAKSONEN, S. (1991). Weighting via response modeling in the Finnish household budget survey. *J. Off. Stat*. **7** 325–337.

ELTINGE, J. L. and YANSANEH, I. S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. consumer expenditure survey. *Surv. Methodol*. **23** 33–40.

ESTEVAO, V. M. and SÄRNDAL, C.-E. (2000). A functional form approach to calibration. *J. Off. Stat*. **16** 379–399.

ESTEVAO, V. M. and SÄRNDAL, C.-E. (2006). Survey estimates by calibration on complex auxiliary information. *Int. Stat. Rev.* **74** 127–147.

FIRTH, B. and BENNETT, K.-E. (1998). Robust models in probability sampling. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **60** 3–21. MR1625672

FOLSOM, R. E. and SINGH, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the Survey Research Methods Section* 598–603. Amer. Statist. Assoc., Alexandria, VA.

FULLER, W. A. (1966). Estimation employing post strata. *J. Amer. Statist. Assoc*. **61** 1172–1183. MR0208781

GELMAN, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci*. **22** 153–164. MR2408951

HÁJEK, J. (1971). Comment on "An essay on the logical foundations of survey sampling" by Basu, D. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 236. Holt, Rinehart, and Winston, Toronto.

HAZIZA, D. and BEAUMONT, J.-F. (2007). On the construction of imputation classes in surveys. *Int. Stat. Rev.* **75** 25–43.

HAZIZA, D. and LESAGE, E. (2016). A discussion of weighting procedures for unit nonresponse. *J. Off. Stat*. **32** 129–145.

HIDIROGLOU, M. A. and PATAK, Z. (2004). Domain estimation using linear regression. *Surv. Methodol*. **30** 67–78.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc*. **47** 663–685. MR0053460

IANNACCHIONE, V. G., MILNE, J. G. and FOLSOM, R. E. (1991). Response probability weight adjustments using logistic regression. In *Proceedings of the Survey Research Methods Section*, 637–642. Amer. Statist. Assoc., Alexandria, VA.

KALBERG, F. (2000). Survey estimation for highly skewed populations in the presence of zeroes. *J. Off. Stat*. **16** 229–241.

KALTON, G. and FLORES-CERVANTES, I. (2003). Weighting methods. *J. Off. Stat*. **19** 81–97.

KASS, V. G. (1980). An exploratory technique for investigating large quantities of categorical data. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **29** 119–127.

KIM, J. K. and KIM, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canad. J. Statist*. **35** 501–514. MR2381396

KIM, J. K., KWON, Y. and PAIK, M. H. C. (2016). Calibrated propensity score method for survey nonresponse in cluster sampling. *Biometrika* **103** 461–473. MR3509899

KIM, J. K. and PARK, M. (2010). Calibration estimation in survey sampling. *Int. Stat. Rev.* **78** 21–39.

KISH, L. (1992). Weighting for unequal P_i. *J. Off. Stat*. **8** 183–200.

KMENTA, J. (1971). *Elements of Econometrics*. MacMillan, New York.

KOTT, P. S. (2006). Using calibration weighting to adjust for nonresponse and undercoverage. *Surv. Methodol*. **32** 133–142.

KOTT, P. S. (2009). Calibration weighting: Combining probability samples and linear prediction models. In *Handbook of Statistics* 29*B*, *Sample Surveys*: *Inference and Analysis*. Elsevier, Oxford.

LAVALLÉE, P. (2007). *Indirect Sampling*. Springer, New York. MR2344875

LAVALLÉE, P. and BEAUMONT, J.-F. (2016). Weighting: Principles and practicalities. In *The SAGE Handbook of Survey Methodology* (C. Wolf, D. Joye, T. W. Smith and Y.-C. Fu, eds.) 460–476, Chapter 30. Sage Publications, Thousand Oaks, CA.

LEMAÎTRE, G. and DUFOUR, J. (1987). An integrated method for weighting persons and families. *Surv. Methodol*. **13** 199–207.

LITTLE, R. J. A. (1986). Survey nonresponse adjustments for estimates of means. *Int. Stat. Rev.* **54** 139–157.

LITTLE, R. J. A. and VARTIVARIAN, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Surv. Methodol*. **31** 161–168.

LUNDSTRÖM, S. and SÄRNDAL, C.-E. (1999). Calibration as a standard method for the treatment of nonresponse. *J. Off. Stat*. **15** 305–327.

MONTANARI, G. E. (1987). Post-sampling efficient QR-prediction in large-sample surveys. *Int. Stat. Rev.* **55** 191–202.

MONTANARI, G. E. and RANALLI, M. G. (2005). Nonparametric model calibration estimation in survey sampling. *J. Amer. Statist. Assoc*. **100** 1429–1442. MR2236453

NARAIN, R. D. (1951). On sampling without replacement with varying probabilities. *J. Indian Soc. Agricultural Statist*. **3** 169–174. MR0045354

PHIPPS, P. and TOTH, D. (2012). Analyzing establishment nonresponse using an interpretable regression tree model with linked administrative data. *Ann. Appl. Stat*. **6** 772–794. MR2976491

POTTER, F. (1990). A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the Section on Survey Research Methods* 225–230. Amer. Statist. Assoc., Alexandria, VA.

RAO, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *J. Indian Statist. Assoc*. **6** 173–180. MR0201035

RAO, J. N. K. (1966). Alternative estimators in sampling for multiple characteristics. *Sankhyā Ser. A* **28** 47–60. MR0258176

RAO, J. N. K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Off. Stat*. **10** 153–165.

RAO, J. N. K. (2005). Interplay between sample survey theory and practice: An appraisal. *Surv. Methodol*. **31** 117–138.

RAO, J. N. K. and MOLINA, I. (2015). *Small Area Estimation*, 2nd ed. Wiley, Hoboken, NJ. MR3380626

RAO, J. N. K., HIDIROGLOU, M., YUNG, W. and KOVACEVIC, M. (2010). Role of weights in descriptive and analytical inferences from survey data: An overview. *J. Indian Soc. Agricultural Statist.* **64** 129–135, 325. MR2789774

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

SALGADO, D., PÉREZ-ARRIERO, C., HERRADOR, M. and ARBUÉS, I. (2012). Letter to the Editor. *J. Off. Stat.* **28** 471–476.

SAMPFORD, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54** 499–513. MR0223051

SÄRNDAL, C.-E. (2007). The calibration approach in survey theory and practice. *Surv. Methodol.* **33** 99–119.

SÄRNDAL, C.-E. and LUNDSTRÖM, S. (2005). *Estimation in Surveys with Nonresponse*. Wiley, New York.

SÄRNDAL, C.-E. and SWENSSON, B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *Int. Stat. Rev.* **55** 279–294. MR0963144

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SAUTORY, O. (2003). CALMAR2: A new version of the CALMAR calibration adjustment program. In Proceedings of Statistics Canada's Symposium 2003. Available at http://www.statcan.ca/english/freepub/11-522-XIE/2003001/session13/sautory.pdf.

SCOTT, A. and SMITH, T. M. F. (1969). A note on estimating secondary caracteristics in multivariate surveys. *Sankya A* **31** 497–498.

SILVA, P. L. N. and SKINNER, C. (1997). Variable selection for regression estimation in finite populations. *Surv. Methodol.* **23** 23–32.

SKINNER, C. J. and D'ARRIGO, J. (2011). Inverse probability weighting for clustered nonresponse. *Biometrika* **98** 953–966. MR2860335

SKINNER, C. J. and WAKEFIELD, J. (2017). Introduction to the design and analysis of comlex survey data. *Statist. Sci.* **32** 165–175.

STEEL, D. G. and CLARK, R. G. (2007). Person-level and household-level regression estimation in household surveys. *Surv. Methodol.* **33** 51–60.

TILLÉ, Y. (2017). Probability sampling designs: New methods and guidelines. *Statist. Sci.* **32** 176–189.

WU, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika* **90** 937–951. MR2024768

WU, C. and LU, W. W. (2016). Calibration weighting methods for complex surveys. *Int. Stat. Rev.* **84** 79–98. MR3491280

WU, C. and RAO, J. N. K. (2006). Pseudo empirical likelihood ratio confidence intervals for complex surveys. *Canad. J. Statist.* **34** 359–375. MR2328549

WU, C. and SITTER, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.* **96** 185–193. MR1952731

Statistical Science
2017, Vol. 32, No. 2, 227–248
DOI: 10.1214/17-STS609

Approaches to Improving Survey-Weighted Estimates

Qixuan Chen, Michael R. Elliott, David Haziza, Ye Yang, Malay Ghosh, Roderick J. A. Little, Joseph Sedransk and Mary Thompson

Abstract. In sample surveys, the sample units are typically chosen using a complex design. This may lead to a selection effect and, if uncorrected in the analysis, may lead to biased inferences. To mitigate the effect on inferences of deviations from a simple random sample a common technique is to use survey weights in the analysis. This article reviews approaches to address possible inefficiency in estimation resulting from such weighting.

To improve inferences we emphasize modifications of the basic design-based weight, that is, the inverse of a unit's inclusion probability. These techniques include weight trimming, weight modelling and incorporating weights via models for survey variables. We start with an introduction to survey weighting, including methods derived from both the design and model-based perspectives. Then we present the rationale and a taxonomy of methods for modifying the weights. We next describe an extensive numerical study to compare these methods. Using as the criteria relative bias, relative mean square error, confidence or credible interval width and coverage probability, we compare the alternative methods and summarize our findings. To supplement this numerical study we use Texas school data to compare the distributions of the weights for several methods. We also make general recommendations, describe limitations of our numerical study and make suggestions for further investigation.

Key words and phrases: Design-based survey weights, finite population survey sampling, inclusion probability, weight modeling, weight trimming.

REFERENCES

BASU, D. (1971). An essay on the logical foundations of survey sampling. Part I. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 203–242. Holt, Rinehart and Winston, Toronto. MR0423625

BEAUMONT, J.-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika* **95** 539–553. MR2443174

BEAUMONT, J.-F., HAZIZA, D. and RUIZ-GAZEN, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika* **100** 555–569. MR3094437

BENRUD, C. H. et al. (1978). Final report on national assessment of educational progress: sampling and weighting activities for assessment year 08. Research Triangle Park, North Carolina: National Assessment of Education Progress.

BINDER, D. A. (1982). Nonparametric Bayesian models for sam-

Qixuan Chen is Assistant Professor, Department of Biostatistics, Columbia University, New York, New York 10032, USA (e-mail: qc2138@cumc.columbia.edu). Michael R. Elliott is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: mrelliot@umich.edu). David Haziza is Professor, Department of Mathematics and Statistics, Université de Montréal, Montreal, Quebec, Canada H3T 1J4, (e-mail: haziza@dms.umontreal.ca). Ye Yang is Graduate Student, University of Michigan, Ann Arbor, Michigan 48105, USA (e-mail: yeya@umich.edu). Malay Ghosh is Professor, Department of Statistics, University of Florida, Gainesville, Florida 32611, USA (e-mail: ghoshm@stat.ufl.edu). Roderick J. A. Little is Professor, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA (e-mail: rlittle@umich.edu). Joseph Sedransk is Professor, Joint Program in Survey Methodology, University of Maryland, College Park, Maryland 20742, USA (e-mail: jxs123@cwru.edu). Mary Thompson is Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1, (e-mail: methomps@uwaterloo.ca).

ples from finite populations. *J. R. Stat. Soc., Ser. B*. **44** 388–393. MR0693238

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J. R. Stat. Soc., Ser. A* **143** 383–430. MR0603745

BREWER, K. R. W. (1979). A class of robust sampling designs for large-scale surveys. *J. Amer. Statist. Assoc.* **74** 911–915. MR0556487

CHEN, Q., ELLIOTT, M. R. and LITTLE, R. J. A. (2010). Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling. *Surv. Methodol.* **36** 23–34.

CHEN, Q., ELLIOTT, M. R. and LITTLE, R. J. A. (2012). Bayesian inference for finite population quantiles from unequal probability samples. *Surv. Methodol.* **38** 203–214.

CHEN, Q., ELLIOTT, M. R., HAZIZA, D., YANG, Y., GHOSH, M., LITTLE, R. J., SEDRANSK, J. and THOMPSON, M. (2017). Supplement to "Approaches to Improving Survey-Weighted Estimates." DOI:10.1214/17-STS609SUPP.

COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York. MR0474575

DALÉN, J. (1986). Sampling from finite populations: Actual coverage probabilities for confidence intervals on the population mean. *J. Off. Stat.* **2** 13–24.

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804

ELLIOTT, M. R. (2007). Bayesian weight trimming for generalized linear regression models. *Surv. Methodol.* **33** 23–34.

ELLIOTT, M. R. (2008). Model averaging methods for weight trimming. *J. Off. Stat.* **24** 517–540.

ELLIOTT, M. R. (2009). Model averaging methods for weight trimming in generalized linear regression models. *J. Off. Stat.* **25** 1–20.

ELLIOTT, M. R. and LITTLE, R. J. A. (2000). Model-based alternatives to trimming survey weights. *J. Off. Stat.* **16** 191–209.

ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations. *J. R. Stat. Soc., Ser. B*. **31** 195–233. MR0270494

ERICSON, W. A. (1988). Bayesian inference in finite populations. In *Handbook of Statistics* (P. R. C. Krishnaiah and C. R. Rao, eds.) 213–246. North-Holland, Amsterdam. MR1020084

FAVRE-MARTINOZ, C., HAZIZA, D. and BEAUMONT, J.-F. (2015). A method of determining the winsorization threshold, with an application to domain estimation. *Surv. Methodol.* **41** 57–77.

FAVRE-MARTINOZ, C., HAZIZA, D. and BEAUMONT, J.-F. (2016). Robust inference in two-phase sampling designs with application to unit nonresponse. *Scand. J. Stat.* **43** 1019–1034. MR3573673

FULLER, W. A. (2009). *Sampling Statistics*. Wiley, New York.

GELMAN, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.* **22** 153–164. MR2408951

GHOSH, M. and MEEDEN, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.* **81** 1058–1062. MR0867632

GHOSH, M. and MEEDEN, G. (1997). *Bayesian Methods for Finite Population Sampling*. Chapman & Hall, London.

GODAMBE, V. and THOMPSON, M. (1986). Parameters of superpopulation and survey populations: Their relationships and estimation. *Int. Stat. Rev.* **54** 127–138. MR1469494

HÁJEK, J. (1971). Comment on a paper by D. Basu. *Foundations of Statistical Inference* 236.

HANSEN, M. H., HURWITZ, W. N. and MADOW, W. G. (1953). *Sample Survey Methods and Theory. Methods and Applications* **1**. Wiley, New York. MR0058171

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. Chapman & Hall, London. MR1082147

HAZIZA, D. and BEAUMONT, J. (2017). Construction of weights in surveys: A review. *Statist. Sci.* **32** 206–226.

HAZIZA, D., MECATTI, F. and RAO, J. N. K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *Metron* **66** 91–108.

HENRY, K. and VALLIANT, R. (2012). Comparing alternative weight adjustment methods. In *Proceedings of the Section on Survey Research Methods* 4696–4710. Amer. Statist. Assoc., Alexandria, VA.

HOLT, D. and SMITH, T. M. F. (1979). Post stratification. *J. R. Stat. Soc., Ser. A* **142** 33–46.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

HULLIGER, B. (1995). Outlier robust Horvitz–Thompson estimator. *Surv. Methodol.* **21** 79–87.

ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029

KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458

KIM, J. K. and SKINNER, C. J. (2013). Weighting in survey analysis under informative sampling. *Biometrika* **100** 385–398. MR3068441

KISH, L. (1995). The hundred years' wars of survey sampling. *Stat. Transit.* **2** 813–830.

KOKIC, P. (1998). On winsorization in business surveys. In *Proceedings of the Survey Methods Section* 237–239. Statistical Society of Canada, Ottawa.

KOKIC, P. N. and BELL, P. A. (1994). Optimal winsorising cutoffs for a stratified finite population estimator. *J. Off. Stat.* **10** 419–435.

LAZZERONI, L. C. and LITTLE, R. J. A. (1998). Random effects models for smoothing post-stratification weights. *J. Off. Stat.* **14** 61–78.

LITTLE, R. J. A. (1983). Comment on "An evaluation of model-dependent and probability sampling inferences in sample surveys" by M. H. Hansen, W. G. Madow and B. J. Tepping. *J. Amer. Statist. Assoc.* **78** 797–799.

LITTLE, R. J. A. (1991). Inference with survey weights. *J. Off. Stat.* **7** 405–424.

LITTLE, R. J. A. (2003). Bayesian methods for unit and item nonresponse. In *Analysis of Survey Data* (R. L. Chambers and C. J. Skinner, eds.) 289–306. Wiley, Chichester. MR1978857

LITTLE, R. J. A. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *Amer. Statist.* **60** 213–223.

LITTLE, R. J. A. (2012). Calibrated Bayes: An alternative inferential paradigm for official statistics. *J. Off. Stat.* **28** 309–372.

MATEI, A. and TILLÉ, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *J. Off. Stat.* **21** 543–570.

MOLINA, C. and SKINNER, C. (1992). Pseudo-likelihood and quasi-likelihood estimation for complex sampling schemes. *Comput. Statist. Data Anal.* **13** 395–405.

MORENO-REBOLLO, J. L., MUÑOZ-REYES, A. and MUÑOZ-PICHARDO, J. (1999). Influence diagnostic in survey sampling: Conditional bias. *Biometrika* **86** 923–928. MR1741987

MORENO-REBOLLO, J. L., MUÑOZ-REYES, A., JIMÉNEZ-GAMERO, M. D. and MUÑOZ-PICHARDO, J. (2002). Influence diagnostic in survey sampling: Estimating the conditional bias. *Metrika* **55** 209–214. MR1912392

NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.* **97** 558–625. MR0121942

O'HAGAN, A. (1995). Fractional Bayes factors for model comparison. *J. R. Stat. Soc., Ser. B.* **57** 99–138. MR1325379

PENG, Y., LITTLE, R. J. A. and RAGHUNATHAN, T. E. (2004). An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics* **60** 598–607. MR2089434

PFEFFERMANN, D. and SVERCHKOV, M. (2009). Inference under informative sampling. In *Sample Surveys: Inference and Analysis* (D. Pfeffermann and C. R. Rao, eds.) **29B** 455–487. Elsevier, Amsterdam.

POTTER, F. (1988). Survey of procedures to control extreme sampling weights. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* 453–458. Amer. Statist. Assoc., Alexandria, VA.

POTTER, F. (1990). A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* 225–230.

RAO, J. N. K. (1966). Alternative estimators in sampling for multiple characteristics. *Sankhyā Ser. A* **28** 47–60. MR0258176

RAO, J. N. K. and WU, C.-F. J. (1988). Resampling inference with complex survey data. *J. Amer. Statist. Assoc.* **83** 231–241. MR0941020

RAO, J. N. K., WU, C. F. J. and YUE, K. (1992). Some recent work on resampling methods for complex surveys. *Surv. Methodol.* **18** 209–217.

RIVEST, L.-P. and HIDIROGLOU, M. (2004). Outlier treatment for disaggregated estimates. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* 4248–4256. Amer. Statist. Assoc., Alexandria, VA.

RIVEST, L.-P. and HURTUBISE, D. (1995). On Searls' winsorized means for skewed populations. *Surv. Methodol.* **21** 107–116.

ROYALL, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57** 377–387.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

RUBIN, D. B. (1983). Comment on "An evaluation of model-dependent and probability-sampling inferences in sample surveys" by M. H. Hansen, W. G. Madow, and B. J. Tepping. *J. Amer. Statist. Assoc.* **78** 803–805.

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172. MR0760681

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York. MR0899519

SÄRNDAL, C.-E. (2007). The calibration approach in survey theory and practice. *Surv. Methodol.* **33** 99–119.

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SCOTT, A. J. (1977). On the problem of randomization in survey sampling. *Sankhyā* **39** 1–9.

SCOTT, A. J. and SMITH, T. M. F. (1969). Estimation in multi-stage surveys. *J. Amer. Statist. Assoc.* **64** 830–840.

SUGDEN, R. A. and SMITH, T. M. F. (1984). Ignorable and informative designs in survey sampling inference. *Biometrika* **71** 495–506. MR0775395

TAMBAY, J. L. (1988). Integrated approach for the treatment of outliers in sub-annual economic surveys. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* 229–234. Amer. Statist. Assoc., Alexandra, VA.

THOMPSON, M. E. (1997). *Theory of Sample Surveys. Monographs on Statistics and Applied Probability* **74**. Chapman & Hall, London. MR1462619

TILLÉ, Y. (2017). Sampling designs: New methods and guidelines. *Statist. Sci.* To appear.

VALLIANT, R., DORFMAN, A. H. and ROYALL, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley, New York. MR1784794

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Stat. Soc., Ser. B.* **40** 364–372. MR0522220

WU, C.-F. and DENG, L.-Y. (1983). Estimation of variance of the ratio estimator: An empirical study. In *Scientific Inference, Data Analysis and Robustness* (G. E. P. Box, T. Leonard and C. F. J. Wu, eds.) 245–277. Academic Press, Orlando, FL. MR0772772

WU, C. and LU, W. W. (2016). Calibration weighting methods for complex surveys. *Int. Stat. Rev.* **84** 79–98. MR3491280

ZANGENEH, S. Z. and LITTLE, R. J. A. (2015). Bayesian inference for the finite population total from a heteroscedastic probability proportional to size sample. *Journal of Survey Statistics and Methodology* **3** 162–192.

ZHENG, H. and LITTLE, R. J. A. (2003). Penalized spline model-based estimation of the finite population total from probability-proportional-to-size samples. *J. Off. Stat.* **19** 99–117.

ZHENG, H. and LITTLE, R. J. A. (2005). Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *J. Off. Stat.* **21** 1–20.

Statistical Science
2017, Vol. 32, No. 2, 249–264
DOI: 10.1214/16-STS598

Inference for Nonprobability Samples

Michael R. Elliott and Richard Valliant

Abstract. Although selecting a probability sample has been the standard for decades when making inferences from a sample to a finite population, incentives are increasing to use nonprobability samples. In a world of "big data", large amounts of data are available that are faster and easier to collect than are probability samples. Design-based inference, in which the distribution for inference is generated by the random mechanism used by the sampler, cannot be used for nonprobability samples. One alternative is quasi-randomization in which pseudo-inclusion probabilities are estimated based on covariates available for samples and nonsample units. Another is superpopulation modeling for the analytic variables collected on the sample units in which the model is used to predict values for the nonsample units. We discuss the pros and cons of each approach.

Key words and phrases: Coverage error, hierarchical regression, quasi-randomization, reference sample, selection bias, superpopulation model.

REFERENCES

ALVAREZ, R., SHERMAN, R. and VAN BESELAERE, C. (2003). Subject acquisition for web-based surveys. *Polit. Anal.* **11** 23–43.

BAKER, R., BRICK, J., BATES, N., COUPER, M., COURTRIGHT, M., DENNIS, J., DILLMAN, D., FRANKEL, M., GARLAND, P., GROVES, R., KENNEDY, C., KROSNICK, J., LAVRAKAS, P., LEE, S., LINK, M., PIEKARSKI, L., RAO, K., THOMAS, R. and ZAHS, D. (2010). AAPOR report on online panels. *Public Opin. Q.* **74** 711–781.

BAKER, R., BRICK, J. M., BATES, N. A., BATTAGLIA, M., COUPER, M. P., DEVER, J. A., GILE, K. and TOURANGEAU, R. (2013a). Report of the AAPOR Task Force on Non-probability Sampling. Technical report, American Association for Public Opinion Research, Deerfield, IL.

BAKER, R., BRICK, J. M., BATES, N. A., BATTAGLIA, M., COUPER, M. P., DEVER, J. A., GILE, K. and TOURANGEAU, R. (2013b). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology* **1** 90–143.

BERZOFSKY, M., WILLIAMS, R. and BIEMER, P. (2009). Combining probability and non-probability sampling methods: Model-aided sampling and the O*NET data collection program. *Survey Practice*.

BETHLEHEM, J. (2010). Selection bias in web surveys. *Int. Stat. Rev.* **78** 161–188.

BINDER, D. and ROBERTS, G. (2009). Imputation of business survey data. In *Handbook of Statistics*, *Sample Surveys*: *Inference and Analysis*, *Volume* 29*B* (D. Pfeffermann and C. Rao, eds.). Elsevier, Amsterdam.

BRICK, J. (2015). Compositional model inference. In *Proceedings of the Section on Survey Research Methods* 299–307. Amer. Statist. Assoc., Alexandria, VA.

CAVALLO, A. and RIGOBON, R. (2016). The billion prices project: Using online prices for measurement and research. *The Journal of Economic Perspectives* 151–178.

CHEN, J. K.-T. (2015). Using LASSO to Calibrate Nonprobability Samples using Probability Samples. Ph.D. thesis, Univ. Michigan, Ann Arbor, MI.

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172

CLEMENT, S. (2016). How the Washington Post-SurveyMonkey 50-state poll was conducted. Available at https://www.washingtonpost.com/news/post-politics/wp/2016/09/06/how-the-washington-post-surveymonkey-50-state-poll-was-conducted/.

COUPER, M. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods* **7** 145–156.

COWLING, D. (2015). Election 2015: How the opinion polls got it wrong. Available at http://www.bbc.com/news/uk-politics-32751993. BBC News online; accessed 06-November-2016.

DEVER, J., RAFFERTY, A. and VALLIANT, R. (2008). Internet surveys: Can statistical adjustments eliminate coverage bias? *Sur-*

Michael R. Elliott is Professor, Biostatistics Department & Research Professor, Institute for Social Research, University of Michigan, ISR Rm 4068, 426 Thompson St., Ann Arbor, Michigan 48109, USA (e-mail: mrelliott@umich.edu). Richard Valliant is Research Professor, Institute for Social Research, University of Michigan & Joint Program in Survey Methodology, University of Maryland, 1218 Lefrak Hall, College Park, Maryland 20742, USA (e-mail: rvallian@umd.edu).

vey Research Methods **2** 47–62.

DEVER, J. and VALLIANT, R. (2010). A comparison of variance estimators for poststratification to estimated control totals. *Surv. Methodol.* **36** 45–56.

DEVER, J. and VALLIANT, R. (2014). Estimation with non-probability surveys and the question of external validity. In *Proceedings of Statistics Canada Symposium* 2014. Statistics Canada, Ottawa, ON.

DEVER, J. and VALLIANT, R. (2016). GREG estimation with undercoverage and estimated controls. *Journal of Survey Statistics and Methodology* **4** 289–318.

DEVILLE, J. (1991). A theory of quota surveys. *Surv. Methodol.* **17** 163–181.

DONG, Q., ELLIOTT, M. and RAGHUNATHAN, T. (2014). A nonparametric method to generate synthetic populations to adjust for complex sample designs. *Surv. Methodol.* **40** 29–46.

ELLIOTT, M. (2009). Combining data from probability and non-probability samples using pseudo-weights. *Survey Practice*.

ELLIOTT, M. R. and DAVIS, W. W. (2005). Obtaining cancer risk factor prevalence estimates in small areas: Combining data from two surveys. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **54** 595–609. MR2137256

ELLIOTT, M. and LITTLE, R. J. A. (2000). Model averaging methods for weight trimming. *J. Off. Stat.* **16** 191–209.

ELLIOTT, M., RESLER, A., FLANNAGAN, C. and RUPP, J. (2010). Combining data from probability and non-probability samples using pseudo-weights. *Accident Analysis and Prevention* **42** 530–539.

ENTEN, H. (2014). Flying Blind Toward Hogan's Upset Win In Maryland. Available at http://fivethirtyeight.com/datalab/governor-maryland-surprise-brown-hogan/. FiveThirtyEight online; accessed 06-November-2016.

FERRARI, S. L. P. and CRIBARI-NETO, F. (2004). Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31** 799–815. MR2095753

FILE, T. and RYAN, C. (2014). Computer and internet use in the United States: 2013. Available at http://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-28.pdf. US Census Bureau; accessed 06-November-2016.

FROST, S., BROUWER, K., FIRESTONE-CRUZ, M., RAMOS, R., RAMOS, M., LOZADA, R., MAGIS-RODRIGUEZ, C. and STRATHDEE, S. (2006). Respondent-driven sampling of injection drug users in two U.S.-Mexico border cities: Recruitment dynamics and impact on estimates of hiv and syphilis prevalence. *Journal of Urban Health* **83** 83–97.

GILE, K. J. and HANDCOCK, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociol. Method.* **40** 285–327.

GOSNELL, H. F. (1937). How accurate were the polls? *Public Opin. Q.* **1** 97–105.

HAZIZA, D. and BEAUMONT, J.-F. (2017). Construction of weights in surveys: A review. *Statist. Sci.* **32** 206–226.

HECKATHORN, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Soc. Probl.* **44** 174–199.

HOLT, D. and SMITH, T. M. F. (1979). Poststratification. *J. R. Stat. Soc., A* **142** 33–46.

KAIZAR, E. (2015). Incorporating both randomized and observational data into a single analysis. *Annual Review of Statistics and Its Application* **2** 49–72.

KEIDING, N. and LOUIS, T. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *J. R. Stat. Soc., A* **179** 319–376. MR3461587

KOHUT, A., KEETER, S., DOHERTY, C., DIMOCK, M. and CHRISTIAN, L. (2012). Assessing the representativeness of public opinion surveys. Available at http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/. Pew Research Center; accessed 06-November-2016.

KORN, E. and GRAUBARD, B. (1999). *Analysis of Health Surveys*. Wiley, New York.

LEBLANC, M. and TIBSHIRANI, R. (1998). Monotone shrinkage of trees. *J. Comput. Graph. Statist.* **7** 417–433.

LEE, S. and VALLIANT, R. (2009). Estimation for volunteer panel web surveys uing propensity score adjustment and calibration adjustment. *Sociol. Methods Res.* **37** 319–343.

LIEBERMANN, O. (2015). Why were the Israeli election polls so wrong? Available at http://www.cnn.com/2015/03/18/middleeast/israel-election-polls/. CNN online; accessed 06-November-2016.

LITTLE, R. J. A. (1982). Models for nonresponse in sample surveys. *J. Amer. Statist. Assoc.* **77** 237–250.

LITTLE, R. J. A. (2003). Bayesian methods for unit and item nonresponse. In *Analysis of Survey Data* (R. Chambers and C. Skinner, eds.). Wiley, Chichester.

LUMLEY, T. and SCOTT, A. (2017). Fitting regression models to survey data. *Statist. Sci.* **32** 265–278.

MADIGAN, D., STANG, P., BERLIN, J., SCHUEMIE, M., OVERHAGE, J., SUCHARD, M., DUMOUCHEL, W., HARTZEMA, W. and RYAN, P. (2014). A systematic statistical approach to evaluating evidence from observational studies. *Annual Review of Statistics and Its Application* **1** 11–39.

MURPHY, J., LINK, M., CHILDS, J., TESFAYE, C., DEAN, E., STERN, M., PASEK, J., COHEN, J., CALLEGARO, M. and HARWOOD, P. (2015). Social media in public opinion research. *Public Opin. Q.* **78** 788–794.

NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* **97** 558–625. MR0121942

O'MUIRCHEARTAIGH, C. and HEDGES, L. V. (2014). Generalizing from unrepresentative experiments: A stratified propensity score approach. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63** 195–210. MR3234340

RAO, J. N. K. and WU, C. F. J. (1988). Resampling inference with complex survey data. *J. Amer. Statist. Assoc.* **83** 231–241. MR0941020

RAO, J. N. K., WU, C. F. J. and YUE, K. (1992). Some recent work on resampling methods for complex surveys. *Surv. Methodol.* **18** 209–217.

RIVERS, D. (2007). Sampling for web surveys. Amazon Web Services. Available at https://s3.amazonaws.com/yg-public/Scientific/Sample+Matching_JSM.pdf.

ROSENBAUM, P. and RUBIN, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

ROYALL, R. (1970). On finite population sampling theory under certain linear regression models. *Biometrika* **57** 377–387.

ROYALL, R. (1971). Linear regression models in finite population sampling theory. In *Foundations of Statistical Inference*

(V. Godambe and D. Sprott, eds.). Holt, Rinehart, and Winston, Toronto.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

RUBIN, D. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *J. Amer. Statist. Assoc.* **74** 318–328.

RUBIN, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134. MR0600538

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SCHONLAU, M. and COUPER, M. (2017). Options for conducting web surveys. *Statist. Sci.* **32** 279–292.

SCHONLAU, M., VAN SOEST, A. and KAPTEYN, A. (2007). Are "Webographic" or attitudinal questions useful for adjusting estimates from web surveys using propensity scoring? *Survey Research Methods* **1** 155–163.

SCHONLAU, M., WEIDMER, B. and KAPTEYN, A. (2014). Recruiting an Internet panel using respondent-driven sampling. *J. Off. Stat.* **30** 291–310.

SIMON, H. (1956). Rational choice and the structure of the environment. *Psychological Review* **63** 129–138.

SIRKEN, M. (1970). Household surveys with multiplicity. *J. Amer. Statist. Assoc.* **65** 257–266.

SMITH, T. M. F. (1976). The foundations of survey sampling: A review. *J. Roy. Statist. Soc. Ser. A* **139** 183–204. MR0445669

SMITH, T. M. F. (1983). On the validity of inferences from non-random samples. *J. R. Stat. Soc., A* **146** 394–403. MR0769995

SQUIRE, P. (1988). Why the 1936 literary digest poll failed. *Public Opin. Q.* **52** 125–133.

STUART, E. A., COLE, S. R., BRADSHAW, C. P. and LEAF, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *J. R. Stat. Soc., A* **174** 369–386. MR2898850

STURGIS, P., BAKER, N., CALLEGARO, M., FISHER, S., GREEN, J., JENNINGS, W., KUHA, J., LAUDERDALE, B. and SMITH, P. (2016). Report of the Inquiry into the 2015 British general election opinion polls. Available at http://eprints.ncrm.ac.uk/3789/1/Report_final_revised.pdf. accessed 06-November-2016.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242

US ENERGY INFORMATION ADMINISTRATION (2016). Weekly petroleum status report. Available at https://www.eia.gov/petroleum/supply/weekly/pdf/appendixb.pdf. US Department of Energy online: accessed 06-November-2016.

VALLIANT, R. and DEVER, J. A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociol. Methods Res.* **40** 105–137. MR2758301

VALLIANT, R., DEVER, J. A. and KREUTER, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York. MR3088726

VALLIANT, R., DORFMAN, A. H. and ROYALL, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley, New York.

VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6**.

VONK, T. W. E., VAN OSSENBRUGGEN, R. and WILLEMS, P. (2006). The effects of panel recruitment and management on research results. Available at https://www.esomar.org/web/research_papers/Web-Panel_1476_The-effects-of-panel-recruitment-and-management-on-research-results.php. ESOMAR; accessed 06-November-2016.

WANG, W., ROTHSCHILD, D., GOEL, S. and GELMAN, A. (2015). Forecasting elections with non-representative polls. *Int. J. Forecast.* **31** 980–991.

ZHOU, H., ELLIOTT, M. and RAGHUNATHAN, T. (2016a). Multiple imputation in two-stage cluster samples using the weighted finite population Bayesian bootstrap. *Journal of Survey Statistics and Methodology* **4** 139–170.

ZHOU, H., ELLIOTT, M. and RAGHUNATHAN, T. (2016b). Synthetic multiple imputation procedure for multi-stage complex samples. *J. Off. Stat.* **32** 251–256.

ZHOU, H., ELLIOTT, M. and RAGHUNATHAN, T. (2016c). A two-step semiparametric method to accommodate sampling weights in multiple imputation. *Biometrics* **72** 242–252. MR3500593

Statistical Science
2017, Vol. 32, No. 2, 265–278
DOI: 10.1214/16-STS605

Fitting Regression Models to Survey Data

Thomas Lumley and Alastair Scott

Abstract. Data from complex surveys are being used increasingly to build the same sort of explanatory and predictive models used in the rest of statistics. Although the assumptions underlying standard statistical methods are not even approximately valid for most survey data, analogues of most of the features of standard regression packages are now available for use with survey data. We review recent developments in the field and illustrate their use on data from NHANES.

Key words and phrases: Complex sampling, statistical graphics.

REFERENCES

BATES, D., MÄCHLER, M., BOLKER, B. and WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48.

BEAUMONT, J.-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika* **95** 539–553. MR2443174

BEAUMONT, J.-F., BÉLIVEAU, A. and HAZIZA, D. (2015). Clarifying some aspects of variance estimation in two-phase sampling. *Journal of Survey Statistics and Methodology* **3** 524–542.

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144

BRESLOW, N. E. and DAY, N. E. (1980). *Statistical Methods in Cancer Research*, *Volume I—The Analysis of Case-Control Studies*. IARC Publications, Paris.

BRESLOW, N. E., ROBINS, J. M. and WELLNER, J. A. (2000). On the semi-parametric efficiency of logistic regression under case-control sampling. *Bernoulli* **6** 447–455. MR1762555

BRESLOW, N. E., LUMLEY, T., BALLANTYNE, C. M., CHAMBLESS, L. E. and KULICH, M. (2009a). Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: Applications in epidemiology. *Statistics in Biosciences* **1** 32–49.

BRESLOW, N. E., LUMLEY, T., BALLANTYNE, C. M., CHAMBLESS, L. E. and KULICH, M. (2009b). Using the whole cohort in the analysis of case-cohort data. *Am. J. Epidemiol.* **169** 1398–1405.

CARR, D. B., LITTLEFIELD, R. J., NICHOLSON, W. L. and LITTLEFIELD, J. S. (1987). Scatterplot matrix techniques for large *N*. *J. Amer. Statist. Assoc.* **82** 424–436. MR0898351

CHAMBERS, R. L. and SKINNER, C. J., eds. (2003). *Analysis of Survey Data*. Wiley, Chichester. MR1978840

CHAUDHURI, S., HANDCOCK, M. S. and RENDALL, M. S. (2008). Generalized linear models incorporating population level information: An empirical-likelihood-based approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 311–328. MR2424755

CHEN, Y.-H. and CHEN, H. (2000). A unified approach to regression analysis under double-sampling designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 449–460. MR1772408

DAVIES, R. B. (1980). Algorithm AS 155: The distribution of a linear combination of χ^2 random variables. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **29** 323–333.

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804

DUMOUCHEL, W. H. and DUNCAN, G. J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *J. Amer. Statist. Assoc.* **78** 535–543.

ELLIOTT, M. R. (2007). Bayesian weight trimming for generalized linear regression models. *Surv. Methodol.* **33** 23–34.

ELLIOTT, M. R. (2009). Model averaging methods for weight trimming in generalized linear regression models. *J. Off. Stat.* **25** 1–20.

FABRIZI, E. and LAHIRI, P. (2007). A design-based approximation to the BIC in finite population sampling. Technical Report 4, Dipartimento di Matematica, Statistica, Informatica e Applicazioni, Università degli Studi di Bergamo.

FAREBROTHER, R. W. (1984). Algorithm AS 204: The distribution of a positive linear combination of χ^2 random variables. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **33** 332–339.

NATIONAL CENTER FOR HEALTH STATISTICS (1994). Plan and Operation of the Third National Health and Nutrition Examination Survey, 1976–1980. Number 32 in Series 1: Programs and Collection Procedures.

FULLER, W. A. (1975). Regression analysis for sample survey. *Sankhyā*, *Series C* **37** 117–132.

FULLER, W. A. (2009). *Sampling Statistics*. Wiley, Hoboken, NJ.

GELMAN, A. (2007). Struggles with survey weighting and regression modeling. *Statist. Sci.* **22** 153–164. MR2408951

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* **31** 1208–1211. MR0123385

Thomas Lumley is Professor of Biostatistics and Alastair Scott is Emeritus Professor, Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand (e-mail: t.lumley@auckland.ac.nz; a.scott@auckland.ac.nz).

GUGGENBERGER, P. (2010a). The impact of a Hausman pretest on the size of a hypothesis test: The panel data case. *J. Econometrics* **156** 337–343.

GUGGENBERGER, P. (2010b). The impact of a Hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory* **26** 369–382.

HARMS, T. and DUCHESNE, P. (2010). On kernel nonparametric regression designed for complex survey data. *Metrika* **72** 111–138.

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica* **46** 1251–1271.

HEERINGA, S., WEST, B. T. and BERGLUND, P. A. (2010). *Applied Survey Data Analysis*. CRC Press, Boca Raton, FL.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

KIM, J. K. and SKINNER, C. J. (2013). Weighting in survey analysis under informative sampling. *Biometrika* **100** 385–398. MR3068441

KOENKER, R. and BASSET, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KORN, E. L. and GRAUBARD, B. I. (1998). Scatterplots with survey data. *Amer. Statist.* **52** 58–69.

KORN, E. L. and GRAUBARD, B. I. (1999). *Analysis of Health Surveys*. Wiley, New York.

KOTT, P. S. and CHANG, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *J. Amer. Statist. Assoc.* **105** 1265–1275. MR2752620

KUONEN, D. (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika* **86** 929–935.

LIN, D., TAO, R., KALSBEEK, W., ZENG, D., GONZALEZ, F. II, FERNANDEZ-RHODES, L., GRAFF, M., KOCH, G. G., NORTH, K. and HEISS, G. (2014). Genetic association analysis under complex survey sampling: The Hispanic Community Health Study/Study of Latinos. *American Journal of Human Genetics* **95** 675–688.

LITTLE, R. J. A. (2012). Calibrated Bayes: An alternative inferential paradigm for official statistics. *J. Off. Stat.* **28** 309–372.

LUMLEY, T. (2010). *Complex Surveys*: *A Guide to Analysis Using R*. Wiley, Hoboken, NJ.

LUMLEY, T. (2015). survey: Analysis of complex survey samples. R package version 3.30-3. Available at https://cran.r-project.org/package=survey.

LUMLEY, T. and SCOTT, A. J. (2013). Partial likelihood ratio tests for the Cox model under complex sampling. *Stat. Med.* **32** 110–123. MR3017887

LUMLEY, T. and SCOTT, A. J. (2014). Tests for regression models fitted to survey data. *Aust. N. Z. J. Stat.* **56** 1–14. MR3200288

LUMLEY, T. and SCOTT, A. (2015). AIC and BIC for modeling with complex survey data. *Journal of Survey Statistics and Methodology* **3** 1–18.

LUMLEY, T., SHAW, P. A. and DAI, J. Y. (2011). Connections between survey calibration estimators and semiparametric models for incomplete data. *Int. Stat. Rev.* **79** 200–220.

MAGEE, L. (1998). Improving survey-weighted least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 115–126. MR1625652

MORRISON, J., LAURIE, C., MARAZITA, M., SANDERS, S., OFFENBACHER, S., SALAZAR, C., CONOMOS, M., THORNTON, T., JAIN, D., LAURIE, C., KERR, K., PAPANICOLAOU, G., TAYLOR, K., KASTE, L., BECK, J. and SHAFFER, J. (2016). Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL). *Human Molecular Genetics* **25** 807–816.

MUTHÉN, L. K. and MUTHÉN, B. O. (2012). *Mplus User's Guide*, 7th ed. Muthén & Muthén, Los Angeles, CA.

PFEFFERMANN, D., KRIEGER, A. M. and RINOTT, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statist. Sinica* **8** 1087–1114. MR1666233

PFEFFERMANN, D. and SIKOV, A. (2011). Imputation and estimation under nonignorable nonresponse in household surveys with missing covariate information. *J. Off. Stat.* **27** 181–209.

PFEFFERMANN, D. and SVERCHKOV, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhyā*, *Series B* **61** 166–186. MR1720710

PFEFFERMANN, D., SKINNER, C. J., HOLMES, D. J., GOLDSTEIN, H. and RASBASH, J. (1998). Weighting for unequal selection probabilities in multilevel models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 23–40. MR1625668

PRENTICE, R. L. and PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411. MR0556730

RABE-HESKETH, S. and SKRONDAL, A. (2006). Multilevel modelling of complex survey data. *J. Roy. Statist. Soc. Ser. A* **169** 805–827. MR2291345

RAO, J. N. K. and SCOTT, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *J. Amer. Statist. Assoc.* **76** 221–230. MR0624328

RAO, J. N. K. and SCOTT, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Ann. Statist.* **12** 46–60. MR0733498

RAO, J. N. K., SCOTT, A. J. and SKINNER, C. J. (1998). Quasi-score tests with survey data. *Statist. Sinica* **8** 1059–1070.

RAO, J. N. K., VERRET, F. and HIDIROGLOU, M. A. (2014). A weighted composite likelihood approach to inference for two-level models from survey data. *Surv. Methodol.* **39** 263–282.

RAO, J. N. K., YUNG, W. and HIDIROGLOU, M. A. (2002). Estimating equations for the analysis of survey data using poststratification information. *Sankhyā*, *Series A* **64** 364–378. MR1981764

RIVERA, C. and LUMLEY, T. (2015). Using the whole cohort in the analysis of countermatched samples. *Biometrics* **72** 382–391.

ROBINS, J. M., HERNÁN, M. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemology. *Epidemiology* **11** 550–560.

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

ROSENBAUM, P. L. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

ROTNITZKY, A. and JEWELL, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77** 485–497. MR1087838

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

RUST, K. F. and RAO, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Stat. Methods Med. Res.* **5** 283–310.

SÄRNDAL, C.-E. (2007). The calibration approach in survey theory and practice. *Surv. Methodol.* **33** 99–119.

SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (2003). *Model Assisted Survey Sampling*. Springer, Berlin. MR1140409

SCOTT, A. and WILD, C. (2002). On the robustness of weighted methods for fitting models to case-control data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 207–219. MR1904701

SKINNER, C. and MASON, B. (2012). Weighting in the regression analysis of survey data with a cross-national application. *Canad. J. Statist.* **40** 697–711. MR2998857

SOLON, G., HAIDER, S. J. and WOOLDRIDGE, J. (2013). What are we weighting for? Working Paper 18859, National Bureau of Economic Research, Cambridge, MA.

STATACORP (2015). Stata Statistical Software: Release 14. StataCorp LP, College Station, TX.

STØER, N. C. and SAMUELSEN, S. O. (2012). Comparison of estimators in nested case-control studies with multiple outcomes. *Lifetime Data Anal.* **18** 261–283. MR2945403

THOMAS, D. R. and RAO, J. N. K. (1987). Small-sample comparisons of level and power for simple goodness-of-fit statistics under cluster sampling. *J. Amer. Statist. Assoc.* **82** 630–636. MR0898369

UNWIN, A., THEUS, M. and HOFMANN, H., eds. (2007). *Graphics of Large Datasets*: *Visualizing a Million*. Springer, New York.

VALLIANT, R. (1993). Poststratification and conditional variance estimation. *J. Amer. Statist. Assoc.* **88** 89–96. MR1212481

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

VERRET, F., RAO, J. and HIDIROGLOU, M. A. (2015). Model-based small area estimation under informative sampling. *Surv. Methodol.* **41** 333–347.

WU, Y. and FULLER, W. A. (2005). Preliminary testing procedures for regression with survey samples. In *Proceedings of the Section on Survey Research Methods* 3683–3688. Amer. Statist. Assoc., Alexandria, VA.

Statistical Science
2017, Vol. 32, No. 2, 279–292
DOI: 10.1214/16-STS597

Options for Conducting Web Surveys

Matthias Schonlau and Mick P. Couper

Abstract. Web surveys can be conducted relatively fast and at relatively low cost. However, Web surveys are often conducted with nonprobability samples and, therefore, a major concern is generalizability. There are two main approaches to address this concern: One, find a way to conduct Web surveys on probability samples without losing most of the cost and speed advantages (e.g., by using mixed-mode approaches or probability-based panel surveys). Two, make adjustments (e.g., propensity scoring, post-stratification, GREG) to nonprobability samples using auxiliary variables. We review both of these approaches as well as lesser-known ones such as respondent-driven sampling. There are many different ways Web surveys can solve the challenge of generalizability. Rather than adopting a one-size-fits-all approach, we conclude that the choice of approach should be commensurate with the purpose of the study.

Key words and phrases: Convenience sample, Internet survey.

REFERENCES

AAPOR (2010). *AAPOR report on online panels*. American Association for Public Opinion Research, Deerfield, IL.

AAPOR (2012). Understanding a 'credibility interval' and how it differs from the 'margin of sampling error' in a public opinion poll. Available at http://www.aapor.org/AAPOR_Main/media/MainSiteFiles/DetailedAAPORstatementoncredibilityintervals.pdf.

AAPOR (2013). *Report of the AAPOR Task Force on Nonprobability Sampling*. American Association for Public Opinion Research, Deerfield, IL.

ANTOUN, C., ZHANG, C., CONRAD, F. G. and SCHOBER, M. F. (2015) Comparisons of online recruitment strategies for convenience samples Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk. *Field Methods*. DOI:10.1177/1525822X15603149.

BAKER-PREWITT, J. (2010). Looking beyond quality differences: How do consumer buying patterns differ by sample source? Paper presented at the CASRO panel conference, New Orleans, LA.

BAUERMEISTER, J. A., ZIMMERMAN, M. A., JOHNS, M. M., GLOWACKI, P., STODDARD, S. and VOLZ, E. (2012). Innovative recruitment using online networks: Lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (WebRDS) strategy. *J. Stud. Alcohol Drugs* **73** 834–838.

BENGTSSON, L., LU, X., NGUYEN, Q. C., CAMITZ, M., HOANG, N. L., NGUYEN, T. A., LILJEROS, F. and THORSON, A. (2012). Implementation of web-based respondent-driven sampling among men who have sex with men in Vietnam. *PLoS ONE* **7** e49417.

BERINSKY, A. J., HUBER, G. A. and LENZ, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Polit. Anal.* **20** 351–368.

BETHLEHEM, J. (2010). Selection bias in Web surveys. *Int. Stat. Rev.* **78** 161–188.

BETHLEHEM, J. (2016). Solving the nonresponse problem with sample matching? *Soc. Sci. Comput. Rev.* **34** 59–77.

BETHLEHEM, J. and BIFFIGNANDI, S. (2011). *Handbook of Web Surveys*. Wiley, New York.

BLOM, A. G., GATHMANN, C. and KRIEGER, U. (2015). Setting up an online panel representative of the general population the German Internet panel. *Field Methods* **27** 391–408.

BRICKMAN BHUTTA, C. (2012). Not by the book: Facebook as a sampling frame. *Sociol. Methods Res.* **41** 57–88.

BUHRMESTER, M., KWANG, T. and GOSLING, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* **6** 3–5.

CALLEGARO, M., BAKER, R. P., BETHLEHEM, J., GÖRITZ, A. S., KROSNICK, J. A. and LAVRAKAS, P. J. (2014). *Online Panel Research: A Data Quality Perspective*. Wiley, New York.

Matthias Schonlau is Professor, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. West, Bldg. M3, Waterloo, Ontario, N2L 3G1, Canada (e-mail: schonlau@uwaterloo.ca). Mick P. Couper is Professor, Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, Michigan 48106, USA (e-mail: mcouper@umich.edu).

CDC (2016). National HIV behavioral surveillance (NHBS). Available at http://www.cdc.gov/hiv/statistics/systems/nhbs/index.html.

CITIZEN PANEL (2015). The Citizen Panel at the University of Gothenburg. Available at http://lore.gu.se/surveys/citizen.

COCHRAN, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **25** 295–313.

COUPER, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opin. Q.* **64** 464–494.

COUPER, M. P. (2007). Issues of representation in eHealth research (with a focus on Web surveys). *Am. J. Prev. Med.* **32** S83–S89.

COUPER, M. P. (2012). Assessment of innovations in data collection technology for understanding society. Report to the Economic and Social Research Council, UK. Available at http://eprints.ncrm.ac.uk/2276/.

COUPER, M. P. and MILLER, P. V. (2008). Web survey methods: Introduction. *Public Opin. Q.* **72** 831–835.

CRAIG, B. M., HAYS, R. D., PICKARD, A. S., CELLA, D., REVICKI, D. A. and REEVE, B. B. (2013). Comparison of US panel vendors for online surveys. *J. Med. Internet Res.* **15** e260.

DAS, M., TOEPOEL, V. and VAN SOEST, A. (2011). Nonparametric tests of panel conditioning and attrition bias in panel surveys. *Sociol. Methods Res.* **40** 32–56. MR2758298

DENNIS, M. (2015). Technical overview of the AmeriSpeak panel—NORC's probability-based research panel. Technical report.

DEVER, J. A., RAFFERTY, A. and VALLIANT, R. (2008). Internet surveys: Can statistical adjustments eliminate coverage bias? *Surv. Res. Methods* **2** 47–62.

DISOGRA, C. (2008). River samples: A good catch for researchers? Knowledge Networks Newsletter. Available at http://www.knowledgenetworks.com/accuracy/fall-winter2008/disogra.html.

ELLIOTT, M. N. and HAVILAND, A. (2007). Use of a Web-based convenience sample to supplement a probability sample. *Surv. Methodol.* **33** 211–215.

ELLIOTT, M. R. and VALLIANT, R. (2017). Inference for nonprobability samples. *Statist. Sci.* **32** 249–264.

ERENS, B., BURKILL, S., COUPER, M. P., CONRAD, F., CLIFTON, S., TANTON, C., PHELPS, A., DATTA, J., MERCER, C. H., SONNENBERG, P. et al. (2014). Nonprobability Web surveys to measure sexual behaviors and attitudes in the general population: A comparison with a probability sample interview survey. *J. Med. Internet Res.* **16** e276.

FAASSE, J. (2005). Panel proliferation and quality concerns. In *Proceedings of ESOMAR Conference on Worldwide Panel Research*: *Developments and Progress*, *Budapest*, *Hungary* 159–169, ESOMAR, Amsterdam. [CD].

GESIS PANEL (2015). GESIS panel study descriptions (version 11.0.0). Available at http://www.gesis.org/en/services/data-collection/gesis-panel/gesis-panel-data-usage/.

GHOSH-DASTIDAR, B., ELLIOTT, M. N., HAVILAND, A. M. and KAROLY, L. A. (2009). Composite estimates from incomplete and complete frames for minimum-MSE estimation in a rare population an application to families with young children. *Public Opin. Q.* **73** 761–784.

GILE, K. J. and HANDCOCK, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociol. Method.* **40** 285–327.

HAZIZA, D. and BEAUMONT, J.-F. (2017). Construction of weights in surveys: A review. *Statist. Sci.* **32** 206–226.

HECKATHORN, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Soc. Probl.* **44** 174–199.

HØGESTØL, A. and SKJERVHEIM, Ø. (2013). The Norwegian Citizen Panel: 2013 first wave. Available at http://www.uib.no/en/citizen.

HOLMBERG, A., LORENC, B. and WERNER, P. (2010). Contact strategies to improve participation via the Web in a mixed-mode mail and Web survey. *J. Off. Stat.* **26** 465–480.

HUGHES, T. and TANCRETO, J. (2015). Refining the Web response option in the multiple mode collection of the American Community Survey. Paper presented at the European Survey Research Association Conference, Reykjavik.

KEETER, S. and CHRISTIAN, L. (2012). *A Comparison of Results from Surveys by the Pew Research Center and Google Consumer Surveys*. Pew Research Center for the People & The Press, Washington, DC.

KEETER, S. and WEISEL, R. (2015). Building Pew Research Center's American Trends Panel. Pew Research Center report. Available at http://www.pewresearch.org/files/2015/04/2015-04-08_building-the-ATP_FINAL.pdf.

KLAUSCH, T., SCHOUTEN, B. and HOX, J. J. (2015). Evaluating bias of sequential mixed-mode designs against benchmark surveys. *Sociol. Methods Res.*. DOI:10.1177/0049124115585362.

LEE, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *J. Off. Stat.* **22** 329–349.

LEE, S. and VALLIANT, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociol. Methods Res.* **37** 319–343.

LITTLE, R. J. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.

LORCH, J., CAVALLARO, K. and VAN OSSENBRUGGEN, R. (2010). Sample blending: $1 + 1 > 2$. Paper presented at the CASRO Panel Conference, New Orleans, LA.

LU, X., BENGTSSON, L., BRITTON, T., CAMITZ, M., KIM, B. J., THORSON, A. and LILJEROS, F. (2012). The sensitivity of respondent-driven sampling. *J. Roy. Statist. Soc. Ser. A* **175** 191–216. MR2873802

MCDONALD, P., MOHEBBI, M. and SLATKIN, B. (2012). Comparing Google Consumer Surveys to existing probability and non-probability based Internet surveys. White paper. Google, Mountain View, CA. Available at http://www.google.com/insights/consumersurveys/static/358002174745700394/consumer_surveys_whitepaper.pdf.

MEDWAY, R. L. and FULTON, J. (2012). When more gets you less: A meta-analysis of the effect of concurrent Web options on mail survey response rates. *Public Opin. Q.* **76** 733–746.

NELSON, E. J., HUGHES, J., OAKES, J. M., PANKOW, J. S. and KULASINGAM, S. L. (2014). Estimation of geographic variation in human papillomavirus vaccine uptake in men and women: An online survey using Facebook recruitment. *J. Med. Internet Res.* **16** e198.

O'DONOVAN, G. and SHAVE, R. (2007). British adults' views on the health benefits of moderate and vigorous activity. *Prev. Med.* **45** 432–435.

PASEK, J. and KROSNICK, J. A. (2010). Measuring intent to participate and participation in the 2010 census and their correlates and trends: Comparisons of RDD telephone and non-probability

sample Internet survey data. Technical report 2010:15, Statistical Research Division of the US Census Bureau, Washington DC.

RIVERS, D. (2007). Sampling for Web surveys. Paper presented at the Joint Statistical Meetings in Salt Lake City, UT.

RIVERS, D. and BAILEY, D. (2009). Inference from matched samples in the 2008 US national elections. In *Proceedings of the Joint Statistical Meetings* 627–639.

SCHERPENZEEL, A. (2011). Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used. *BMS. Bull. Méthodol. Sociol.* **109** 56–61.

SCHNORF, S., SEDLEY, A., ORTLIEB, M. and WOODRUFF, A. (2014). A comparison of six sample providers regarding online privacy benchmarks. In *SOUPS Workshop on Privacy Personas and Segmentation*, Menlo Park, CA.

SCHONLAU, M., ASCH, B. J. and DU, C. (2003). Web surveys as part of a mixed-mode strategy for populations that cannot be contacted by e-mail. *Soc. Sci. Comput. Rev.* **21** 218–222.

SCHONLAU, M., FRICKER, R. and ELLIOTT, M. N. (2002). *Conducting Research Surveys via E-mail and the Web*. RAND Corporation, Santa Monica, CA.

SCHONLAU, M. and TOEPOEL, V. (2015). Straightlining in Web survey panels over time. *Surv. Res. Methods* **9** 125–137.

SCHONLAU, M., WEIDMER, B. and KAPTEYN, A. (2014). Recruiting an Internet panel using respondent-driven sampling. *J. Off. Stat.* **30** 291–310.

SCHONLAU, M., ZAPERT, K., PAYNE SIMON, L., HAYNES SANSTAD, K., MARCUS, S. M., ADAMS, J., SPRANCA, M., KAN, H., TURNER, R. and BERRY, S. H. (2004). A comparison between responses from a propensity-weighted Web survey and an identical RDD survey. *Soc. Sci. Comput. Rev.* **22** 128–138.

SCHONLAU, M., VAN SOEST, A., KAPTEYN, A. and COUPER, M. (2009). Selection bias in web surveys and the use of propensity scores. *Sociol. Methods Res.* **37** 291–318. MR2649462

SCHOUTEN, B., VAN DEN BRAKEL, J., BUELENS, B., VAN DER LAAN, J. and KLAUSCH, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Soc. Sci. Res.* **42** 1555–1570.

SEEMAN, N. (2015). Use data to challenge mental-health stigma. *Nature* **528** 309. Available at http://www.nature.com/news/use-data-to-challenge-mental-health-stigma-1.19033.

SEEMAN, N., TANG, S., BROWN, A. D. and ING, A. (2016). World survey of mental illness stigma. *J. Affective Disorders* **190** 115–121.

STEIN, M. L., VAN STEENBERGEN, J. E., CHANYASANHA, C., TIPAYAMONGKHOLGUL, M., BUSKENS, V., VAN DER HEIJDEN, P. G. M., SABAIWAN, W., BENGTSSON, L., LU, X., THORSON, A. E. et al. (2014). Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: A pilot study in Thailand. *PLoS ONE* **9** e85256.

STURGIS, P., BAKER, N., CALLEGARO, M., FISHER, S., GREEN, J., JENNINGS, W., KUHA, J., LAUDERDALE, B. and SMITH, P. (2016). Report of the inquiry into the 2015 British general election opinion polls. Available at http://eprints.ncrm.ac.uk/3789/1/Report_final_revised.pdf.

TAYLOR, H., BREMER, J., OVERMEYER, C., SIEGEL, J. W. and TERHANIAN, G. (2001). The record of Internet-based opinion polls in predicting the results of 72 races in the November 2000 US elections. *Int. J. Mark. Res.* **43** 127–136.

TOEPOEL, V., DAS, M. and VAN SOEST, A. (2008). Effects of design in Web surveys comparing trained and fresh respondents. *Public Opin. Q.* **72** 985–1007.

VALLIANT, R., DEVER, J. A. and KREUTER, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, Berlin. MR3088726

VAVRECK, L. and RIVERS, D. (2008). The 2006 cooperative congressional election study. *J. Elect. Publ. Opin. Part.* **18** 355–366.

VONK, T., VAN OSSENBRUGGEN, R. and WILLEMS, P. (2006). A comparison study across 19 online panels (NOPVO 2006). In *Access Panels and Online Research, Panacea or Pitfall?* (I. Stoop and M. Wittenberg, eds.). *DANS Symposium Publications* **4**. Aksant Academic Publishers, Amsterdam.

WANG, W., ROTHSCHILD, D., GOEL, S. and GELMAN, A. (2015). Forecasting elections with non-representative polls. *Int. J. Forecast.* **31** 980–991.

WEJNERT, C. and HECKATHORN, D. D. (2008). Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research. *Sociol. Methods Res.* **37** 105–134.

YEAGER, D. S., KROSNICK, J. A., CHANG, L., JAVITZ, H. S., LEVENDUSKY, M. S., SIMPSER, A. and WANG, R. (2011). Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. *Public Opin. Q.* **75** 709–747.

ZEWOLDI, Y. (2011). Introduction, 2011. Seminar on New Technologies in Population and Housing Censuses: Country Experiences, 42nd session of the United Nations Statistical Commission, New York. Available at http://unstats.un.org/unsd/statcom/statcom_2011/Seminars/NewTechnologies/default.html.

Statistical Science
2017, Vol. 32, No. 2, 293–312
DOI: 10.1214/16-STS584

Combining Survey Data with Other Data Sources

Sharon L. Lohr and Trivellore E. Raghunathan

Abstract. Collecting data using probability samples can be expensive, and response rates for many household surveys are decreasing. The increasing availability of large data sources opens new opportunities for statisticians to use the information in survey data more efficiently by combining survey data with information from these other sources. We review some of the work done to date on statistical methods for combining information from multiple data sources, discuss the limitations and challenges for different methods that have been proposed, and describe research that is needed for combining survey estimates.

Key words and phrases: Hierarchical models, imputation, multiple frame survey, probability sample, record linkage, small area estimation.

REFERENCES

ADES, A. E. and SUTTON, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: Current approaches. *J. Roy. Statist. Soc. Ser. A* **169** 5–35. MR2222010

AMERICAN ASSOCIATION OF PUBLIC OPINION RESEARCH (2015). Code of Professional Ethics and Practices. Available at https://www.aapor.org/Standards-Ethics/AAPOR-Code-of-Ethics.aspx.

ANDRIDGE, R. R. and LITTLE, R. J. A. (2010). A review of hot deck imputation for survey non-response. *Int. Stat. Rev.* **78** 40–64.

BAKER, R., BRICK, J. M., BATES, N. A., BATTAGLIA, M., COUPER, M. P., DEVER, J. A., GILE, K. J. and TOURANGEAU, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology* **1** 90–143.

BANCROFT, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Stat.* **15** 190–204. MR0010373

BANKIER, M. D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *J. Amer. Statist. Assoc.* **81** 1074–1079.

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey snd satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BERLIN, J. A. and RENNIE, D. (1999). Measuring the quality of trials: The quality of quality scales. *J. Amer. Med. Assoc.* **282** 1083–1085.

BHATT, S., WEISS, D. J., CAMERON, E., BISANZIO, D., MAPPIN, B., DALRYMPLE, U., BATTLE, K. E., MOYES, C. L., HENRY, A., ECKHOFF, P. A. et al. (2015). The effect of Malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526** 207–211.

BOHENSKY, M. A., JOLLEY, D., SUNDARARAJAN, V., EVANS, S., PILCHER, D. V., SCOTT, I. and BRAND, C. A. (2010). Data linkage: A powerful research tool with potential problems. *BMC Health Serv. Res.* **10** 1–7.

BRICK, J. M. (2013). Unit nonresponse and weighting adjustments: A critical review. *J. Off. Stat.* **29** 329–353.

BRICK, J. M. (2015). Compositional model inference. In *Proceedings of the Survey Research Methods Section* 299–307. Amer. Statist. Assoc., Alexandria, VA.

BRICK, J. M., CERVANTES, I. F., LEE, S. and NORMAN, G. (2011). Nonsampling errors in dual frame telephone surveys. *Surv. Methodol.* **37** 1–12.

CARPENTER, J. and KENWARD, M. (2012). *Multiple Imputation and Its Application*. Wiley, Hoboken, NJ.

CHAUVET, G. and DE MARSAC, G. T. (2014). Estimation methods on multiple sampling frames in two-stage sampling designs. *Surv. Methodol.* **40** 335–346.

CHEN, C., WAKEFIELD, J. and LUMELY, T. (2014). The use of sampling weights in Bayesian hierarchical models for small area estimation. *Spat. Spatiotemporal Epidemiol.* **11** 33–43.

CHRISTEN, P. (2012). *Data Matching*: *Concepts and Techniques for Record Linkage*, *Entity Resolution*, *and Duplicate Detection*. Springer Science & Business Media, New York.

CITRO, C. F. (2014). From multiple modes for surveys to multiple data sources for estimates. *Surv. Methodol.* **40** 137–161.

Sharon L. Lohr is Vice President, Westat, 1600 Research Boulevard, Rockville, Maryland 20850, USA (e-mail: sharonlohr@westat.com). Trivellore E. Raghunathan is Director of Survey Research Center, Institute for Social Research, and Professor of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48106, USA (e-mail: teraghu@umich.edu).

CITRO, C. F. and STRAF, M. L., EDS. (2013). *Principles and Practices for a Federal Statistical Agency*, 5th ed. National Academies Press, Washington, DC.

CRUZE, N. (2015). Integrating survey data with auxiliary sources of information to estimate crop yields. In *Proceedings of the Survey Research Methods Section* 565–578. Amer. Statist. Assoc., Alexandria, VA.

DAAS, P. J. H., PUTS, M. J., BUELENS, B. and VAN DEN HURK, P. A. (2015). Big data as a source for official statistics. *J. Off. Stat.* **31** 249–262.

DATTA, G. S., GHOSH, M., STEORTS, R. and MAPLES, J. (2011). Bayesian benchmarking with applications to small area estimation. *TEST* **20** 574–588.

DEMING, W. E. (1950). *Some Theory of Sampling*. Wiley, New York.

DEVILLE, J.-C., SÄRNDAL, C.-E. and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *J. Amer. Statist. Assoc.* **88** 1013–1020.

DONG, Q., ELLIOTT, M. R. and RAGHUNATHAN, T. E. (2014a). A nonparametric method to generate synthetic populations to adjust for complex sampling design features. *Surv. Methodol.* **40** 29–46.

DONG, Q., ELLIOTT, M. R. and RAGHUNATHAN, T. E. (2014b). Combining information from multiple complex surveys. *Surv. Methodol.* **40** 347–354.

DUGOFF, E. H., SCHULER, M. and STUART, E. A. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Serv. Res.* **49** 284–303.

DUNCAN, G. T., JABINE, T. B. and DE WOLF, V. A. (1993). *Private Lives and Public Policies*: *Confidentiality and Accessibility of Government Statistics*. National Academies Press, Washington, DC.

DUNCAN, J. W. and SHELTON, W. C. (1992). U.S. Government contributions to probability sampling and statistical analysis. *Statist. Sci.* **7** 320–338. MR1181415

DURRANT, G. B. (2009). Imputation methods for handling item-nonresponse in practice: Methodological issues and recent debates. *International Journal of Social Research Methodology* **12** 293–304.

DWORK, C. (2011). A firm foundation for private data analysis. *Commun. ACM* **54** 86–95.

ELLIOTT, M. R. and DAVIS, W. W. (2005). Obtaining cancer risk factor prevalence estimates in small areas: Combining data from two surveys. *J. Roy. Statist. Soc. Ser. C* **54** 595–609. MR2137256

FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019

FELLEGI, I. P. (1999). Record linkage and public policy: A dynamic evolution. In *Record Linkage Techniques*—1997: *Proceedings of an International Workshop and Exposition* 1–12. National Academy Press, Washington, DC.

FELLEGI, I. P. and SUNTER, A. B. (1969). A theory of record linkage. *J. Amer. Statist. Assoc.* **64** 1183–1210.

FINUCANE, M. M., PACIOREK, C. J., DANAEI, G. and EZZATI, M. (2014). Bayesian estimation of population-level trends in measures of health status. *Statist. Sci.* **29** 18–25. MR3201842

FINUCANE, M. M., PACIOREK, C. J., STEVENS, G. A. and EZZATI, M. (2015). Semiparametric Bayesian density estimation with disparate data sources: A meta-analysis of global childhood undernutrition. *J. Amer. Statist. Assoc.* **110** 889–901. MR3420668

GELMAN, A., KING, G. and LIU, C. (1998). Not asked and not answered: Multiple imputation for multiple surveys. *J. Amer. Statist. Assoc.* **93** 846–857.

GOLDSTEIN, H., HARRON, K. and WADE, A. (2012). The analysis of record-linked data using multiple imputation with data value priors. *Stat. Med.* **31** 3481–3493. MR3041825

GREENLAND, S. (2005). Multiple-bias modelling for analysis of observational data. *J. Roy. Statist. Soc. Ser. A* **168** 267–306. MR2119402

GROVES, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opin. Q.* **70** 646–675.

GROVES, R. M. and HEERINGA, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *J. Roy. Statist. Soc. Ser. A* **169** 439–457. MR2236915

HARRON, K., GOLDSTEIN, H. and DIBBEN, C. (2016). *Methodological Developments in Data Linkage*. Wiley, Hoboken, NJ.

HARTLEY, H. O. (1962). Multiple Frame Surveys. In *Proceedings of the Social Statistics Section*, *American Statistical Association* 203–206. Amer. Statist. Assoc., Alexandria, VA.

HARTLEY, H. O. (1974). Multiple frame methodology and selected applications. *Sankhyā*, *Ser. C* **36** 99–118.

HE, Y., LANDRUM, M. B. and ZASLAVSKY, A. M. (2014). Combining information from two data sources with misreporting and incompleteness to assess hospice-use among cancer patients: A multiple imputation approach. *Stat. Med.* **33** 3710–3724.

HERZOG, T. N., SCHEUREN, F. J. and WINKLER, W. E. (2007). *Data Quality and Record Linkage Techniques*. Springer Science & Business Media, New York.

HURST, B. (2015). Big Data and Agriculture: Innovations and Implications. Statement of the American Farm Bureau Federation to the House Committee on Agriculture, available at http://agriculture.house.gov/uploadedfiles/10.28.15_hurst_testimony.pdf.

HYNDMAN, R. J., LEE, A. J. and WANG, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Comput. Statist. Data Anal.* **97** 16–32.

JACKSON, C., BEST, N. and RICHARDSON, S. (2008). Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *J. Roy. Statist. Soc. Ser. A* **171** 159–178. MR2412651

JONES, K. M., THOMSON, J. C. and ARNOLD, K. (2014). Questions of data ownership on campus. *EDUCAUSE Review*, *August* 1–10.

KALTON, G. and ANDERSON, D. W. (1986). Sampling rare populations. *J. Roy. Statist. Soc. Ser. A* **149** 65–82.

KIM, J. K. and RAO, J. N. K. (2012). Combining data from two independent surveys: A model-assisted approach. *Biometrika* **99** 85–100.

KISH, L. J. and TOPOL, E. J. (2015). Unpatients—Why patients should own their medical data. *Nat. Biotechnol.* **33** 921–924.

KOHUT, A., KEETER, S., DOHERTY, C., DIMOCK, M. and CHRISTIAN, L. (2012). *Assessing the Representativeness of Public Opinion Surveys*. Pew Research Center, Washington DC. Available at http://www.people-press.org/files/legacy-pdf/Assessing%20the%20Representativeness%20of%20Public%20Opinion%20Surveys.pdf.

KORN, E. L. and GRAUBARD, B. I. (1999). *Analysis of Health Surveys*. Wiley, New York.

KOSTKOVA, P., BREWER, H., DE LUSIGNAN, S., FOTTRELL, E., GOLDACRE, B., HART, G., KOCZAN, P., KNIGHT, P., MARSOLIER, C., MCKENDRY, R. A. et al. (2016). Who owns the data? Open data for healthcare. *Frontiers in Public Health* **4** 1–6.

LEE, S. and VALLIANT, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociol. Methods Res.* **37** 319–343.

LESSER, V. M., NEWTON, L. and YANG, D. (2008). Evaluating Frames and Modes of Contact in a Study of Individuals with Disabilities. Paper presented at the Joint Statistical Meetings, Denver, Colorado.

LOHR, S. L. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Surv. Methodol.* **37** 197–213.

LOHR, S. L. and BRICK, J. M. (2012). Blending domain estimates from two victimization surveys with possible bias. *Canad. J. Statist.* **40** 679–696. MR2998856

LOHR, S. L. and BRICK, J. M. (2014). Allocation for dual frame telephone surveys with nonresponse. *Journal of Survey Statistics and Methodology* **2** 388–409.

LOHR, S. L. and RAO, J. N. K. (2006). Estimation in multiple-frame surveys. *J. Amer. Statist. Assoc.* **101** 1019–1030. MR2324141

MACHANAVAJJHALA, A. and KIFER, D. (2015). Designing statistical privacy for your data. *Commun. ACM* **58** 58–67.

MANZI, G., SPIEGELHALTER, D. J., TURNER, R. M., FLOWERS, J. and THOMPSON, S. G. (2011). Modelling bias in combining small area prevalence estimates from multiple surveys. *J. Roy. Statist. Soc. Ser. A* **174** 31–50. MR2758280

MECATTI, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Surv. Methodol.* **33** 151–157.

MERCER, L., WAKEFIELD, J., CHEN, C. and LUMLEY, T. (2014). A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spat. Stat.* **8** 69–85. MR3326822

MERKOURIS, T. (2004). Combining independent regression estimators from multiple surveys. *J. Amer. Statist. Assoc.* **99** 1131–1139. MR2109501

MERKOURIS, T. (2010). Combining information from multiple surveys by using regression for efficient small domain estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 27–48. MR2751242

METCALF, P. and SCOTT, A. (2009). Using multiple frames in health surveys. *Stat. Med.* **28** 1512–1523. MR2649709

MORIARITY, C. and SCHEUREN, F. (2001). Statistical matching: A paradigm for assessing the uncertainty in the procedure. *J. Off. Stat.* **17** 407–422.

MOSTELLER, F. (1948). On pooling data. *J. Amer. Statist. Assoc.* **43** 231–242.

NACHMAN, K. E. and PARKER, J. D. (2012). Exposures to fine particulate air pollution and respiratory outcomes in adults using two national datasets: A cross-sectional study. *Environ. Health* **11** 1–12.

NANDRAM, B., BERG, E. and BARBOZA, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Environ. Ecol. Stat.* **21** 507–530. MR3248537

NATIONAL CENTER FOR HEALTH STATISTICS (2016). Survey Description, National Health Interview Survey, 2014. Centers for Disease Control and Prevention, Hyattsville, MD. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2015/srvydesc.pdf.

NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* **97** 558–625. MR0121942

PFEFFERMANN, D. and TILLER, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *J. Amer. Statist. Assoc.* **101** 1387–1397. MR2307572

POCOCK, S. J. (1976). The combination of randomized and historical controls in clinical trials. *J. Chronic. Dis.* **29** 175–188.

PRENTICE, R. L., SMYTHE, R. T., KREWSKI, D. and MASON, M. (1992). On the use of historical control data to estimate dose response trends in quantal bioassay. *Biometrics* **48** 459–478. MR1173492

RAGHUNATHAN, T. E. (1991). Pooling controls from different studies. *Stat. Med.* **10** 1417–1426.

RAGHUNATHAN, T. E. (2006). Combining information from multiple surveys for assessing health disparities. *Allg. Stat. Arch.* **90** 515–526.

RAGHUNATHAN, T. E., XIE, D., SCHENKER, N., PARSONS, V. L., DAVIS, W. W., DODD, K. W. and FEUER, E. J. (2007). Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *J. Amer. Statist. Assoc.* **102** 474–486. MR2370848

RANALLI, M. G., ARCOS, A., RUEDA, M. D. M. and TEODORO, A. (2016). Calibration estimation in dual-frame surveys. *Stat. Methods Appl.* **25** 321–349. MR3539496

RAO, J. N. K. and MOLINA, I. (2015). *Small Area Estimation*, 2nd ed. Wiley, Hoboken, NJ. MR3380626

RAO, J. N. K. and WU, C. (2010). Pseudo-empirical likelihood inference for multiple frame surveys. *J. Amer. Statist. Assoc.* **105** 1494–1503. MR2796566

RAO, S. R., GRAUBARD, B. I., SCHMID, C. H., MORTON, S. C., LOUIS, T. A., ZASLAVSKY, A. M. and FINKELSTEIN, D. M. (2008). Meta-analysis of survey data: Application to health services research. *Health Serv. Outcomes Res. Methodol.* **8** 98–114.

RÄSSLER, S. (2002). *Statistical Matching*: *A Frequentist Theory*, *Practical Applications*, *and Alternative Bayesian Approaches*. *Lecture Notes in Statistics* **168**. Springer, New York. MR1996879

RENSSEN, R. H. and NIEUWENBROEK, N. J. (1997). Aligning estimates for common variables in two or more sample surveys. *J. Amer. Statist. Assoc.* **92** 368–374. MR1436124

RODGERS, W. L. (1984). An evaluation of statistical matching. *J. Bus. Econom. Statist.* **2** 91–102.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

SÄRNDAL, C.-E. (2007). The calibration approach in survey theory and practice. *Surv. Methodol.* **33** 99–119.

SCHENKER, N., RAGHUNATHAN, T. E. and BONDARENKO, I. (2010). Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey. *Stat. Med.* **29** 533–545. MR2758451

SKINNER, C. J. and RAO, J. N. K. (1996). Estimation in dual frame surveys with complex designs. *J. Amer. Statist. Assoc.* **91** 349–356. MR1394091

SMITH, T. W. (2011). The report of the international workshop on using multi-level data from sample frames, auxiliary databases, paradata and related sources to detect and adjust for nonresponse bias in surveys. *Int. J. Public Opin. Res.* **23** 389–402.

STATISTICS CANADA (2014). Note to Users of Data from the 2012 Canadian Income Survey, available at http://www.statcan.gc.ca/pub/75-513-x/75-513-x2014001-eng.htm.

STEORTS, R. C., HALL, R. and FIENBERG, S. E. (2016). A Bayesian approach to graphical record linkage and de-duplication. *J. Amer. Statist. Assoc.* **111** 1660–1672. MR3601725

STOKES, L. and LIN, D. (2015). Measurement error in dual frame designs. Paper presented at the Joint Statistical Meetings, Seattle WA.

STRAUSS, W. J., CARROLL, R. J., BORTNICK, S. M., MENKEDICK, J. R. and SCHULTZ, B. D. (2001). Combining datasets to predict the effects of regulation of environmental lead exposure in housing stock. *Biometrics* **57** 203–210. MR1833308

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812

SUTTON, A. J. and HIGGINS, J. (2008). Recent developments in meta-analysis. *Stat. Med.* **27** 625–650. MR2418504

SWEETING, M. J., DE ANGELIS, D., HICKMAN, M. and ADES, A. E. (2008). Estimating hepatitis C prevalence in England and Wales by synthesizing evidence from multiple data sources. Assessing data conflict and model fit. *Biostatistics* **9** 715–734.

TOURANGEAU, R., BRICK, J. M., LOHR, S. and LI, J. (2017). Adaptive and responsive survey designs: A review and assessment. *J. Roy. Statist. Soc. Ser. A.* **180** 203–223. MR3600507

TURNER, R. M., OMAR, R. Z., YANG, M., GOLDSTEIN, H. and THOMPSON, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat. Med.* **19** 3417–3432.

TURNER, R. M., SPIEGELHALTER, D. J., SMITH, G. C. S. and THOMPSON, S. G. (2009). Bias modelling in evidence synthesis. *J. Roy. Statist. Soc. Ser. A* **172** 21–47. MR2655603

UNITED STATES CENSUS BUREAU (2016). Model-Based Small Area Income & Poverty Estimates (SAIPE) for School Districts, Counties, and States. Available at http://www.census.gov/did/www/saipe/.

UNITED STATES GENERAL ACCOUNTING OFFICE (1992). Cross-Design Synthesis: A New Strategy for Medical Effectiveness Research. U.S. General Accounting Office, Washington, DC. Available at archive.gao.gov/d31t10/145906.pdf.

VALLIANT, R. and DEVER, J. A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociol. Methods Res.* **40** 105–137. MR2758301

VALLIANT, R., DORFMAN, A. H. and ROYALL, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley, New York.

VOS, T., BARBER, R. M., BELL, B., BERTOZZI-VILLA, A., BIRYUKOV, S., BOLLIGER, I., CHARLSON, F., DAVIS, A., DEGENHARDT, L., DICKER, D. et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386** 743–800.

WAGNER, J. and RAGHUNATHAN, T. (2007). Bayesian approaches to sequential selection of survey design protocols. In *Proceedings of the Survey Research Methods Section* 3333–3340. Amer. Statist. Assoc., Alexandria, VA.

WAGNER, J., WEST, B. T., KIRGIS, N., LEPKOWSKI, J. M., AXINN, W. G. and NDIAYE, S. K. (2012). Use of paradata in a responsive design framework to manage a field data collection. *J. Off. Stat.* **28** 477.

WAKEFIELD, J. (2004). Ecological inference for 2×2 tables (with discussion). *J. Roy. Statist. Soc. Ser. A* **167** 385–445. MR2082057

WAKEFIELD, J. and SALWAY, R. (2001). A statistical framework for ecological and aggregate studies. *J. Roy. Statist. Soc. Ser. A* **164** 119–137. MR1819026

WANG, J. C., HOLAN, S. H., NANDRAM, B., BARBOZA, W., TOTO, C. and ANDERSON, E. (2012). A Bayesian approach to estimating agricultural yield based on multiple repeated surveys. *J. Agric. Biol. Environ. Stat.* **17** 84–106.

WANG, H., WOLOCK, T. M., CARTER, A., NGUYEN, G., KYU, H. H., GAKIDOU, E., HAY, S. I., MILLS, E. J., TRICKEY, A., MSEMBURI, W. et al. (2016). Estimates of global, regional, and national incidence, prevalence, and mortality of HIV, 1980–2015: The Global Burden of Disease Study 2015. *The Lancet. HIV* **3** e361–e387.

WHELDON, M. C., RAFTERY, A. E., CLARK, S. J. and GERLAND, P. (2016). Bayesian population reconstruction of female populations for less developed and more developed countries. *Popul. Stud. (Camb.)* **70** 21–37.

WINKLER, W. E. (2014). Matching and record linkage. *Wiley Interdiscip. Rev.: Comput. Stat.* **6** 313–325.

YBARRA, L. M. and LOHR, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika* **95** 919–931. MR2461220

YEAGER, D. S., KROSNICK, J. A., CHANG, L., JAVITZ, H. S., LEVENDUSKY, M. S., SIMPSER, A. and WANG, R. (2011). Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. *Public Opin. Q.* **75** 709–747.

YOU, J., DATTA, G. S. and MAPLES, J. J. (2014). Modeling disability in small areas: An area-level approach of combining two surveys. In *Proceedings of the Survey Research Methods Section* 3770–3784. Amer. Statist. Assoc., Alexandria, VA.

ZHOU, H., ELLIOTT, M. R. and RAGHUNATHAN, T. E. (2015). A two-step semiparametric method to accommodate sampling weights in multiple imputation. *Biometrics* **72** 242–252. MR3500593

ZOLAS, N., GOLDSCHLAG, N., JARMIN, R., STEPHAN, P., OWEN-SMITH, J., ROSEN, R. F., ALLEN, B. M., WEINBERG, B. A. and LANE, J. I. (2015). Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients. *Science* **350** 1367–1371.

Statistical Science
2017, Vol. 32, No. 2, 313–329
DOI: 10.1214/16-STS599

J. B. S. Haldane's Contribution to the Bayes Factor Hypothesis Test

Alexander Etz and Eric-Jan Wagenmakers

Abstract. This article brings attention to some historical developments that gave rise to the Bayes factor for testing a point null hypothesis against a composite alternative. In line with current thinking, we find that the conceptual innovation—to assign prior mass to a general law—is due to a series of three articles by Dorothy Wrinch and Sir Harold Jeffreys (1919, 1921, 1923a). However, our historical investigation also suggests that in 1932, J. B. S. Haldane made an important contribution to the development of the Bayes factor by proposing the use of a mixture prior comprising a point mass and a continuous probability density. Jeffreys was aware of Haldane's work and it may have inspired him to pursue a more concrete statistical implementation for his conceptual ideas. It thus appears that Haldane may have played a much bigger role in the statistical development of the Bayes factor than has hitherto been assumed.

Key words and phrases: History of statistics, induction, evidence, Sir Harold Jeffreys.

REFERENCES

ALDRICH, J. (2005). The statistical education of Harold Jeffreys. *Int. Stat. Rev.* **73** 289–307.

BANKS, D. L. (1996). A conversation with I. J. Good. *Statist. Sci.* **11** 1–19. MR1437124

BARNARD, G. A. (1967). The Bayesian controversy in statistical inference. *J. Inst. Actuar.* **93** 229–269.

BAYARRI, M. J., BERGER, J. O., FORTE, A. and GARCÍA-DONATO, G. (2012). Criteria for Bayesian model choice with application to variable selection. *Ann. Statist.* **40** 1550–1577. MR3015035

BENNETT, J. H. (1990). *Statistical Inference and Analysis*: *Selected Correspondence of R. A. Fisher*. The Clarendon Press, Oxford Univ. Press, New York.

BERGER, J. O. (2006). Bayes factors. In *Encyclopedia of Statistical Sciences*, *Vol.* 1, 2nd ed. (S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic and N. L. Johnson, eds.) 378–386. Wiley, Hoboken, NJ.

BROAD, C. D. (1918). On the relation between induction and probability (Part I). *Mind* **27** 389–404.

CLARK, R. W. (1968). *JBS*: *The Life and Work of J. B. S. Haldane*. Coward-McCann, Inc., New York.

CROW, J. F. (1992). Centennial: JBS Haldane, 1892–1964. *Genetics* **130** 1.

CUTHILL, J. H. and CHARLESTON, M. (in press). Wing patterning genes and coevolution of Müllerian mimicry in *Heliconius* butterflies: Support from phylogeography, co-phylogeny and divergence times. *Evolution*.

DIENES, Z. (2014). Using Bayes to get the most out of non-significant results. *Front. Psychol.* **5**, 781.

EDWARDS, A. W. F. (1974). The history of likelihood. *Int. Stat. Rev.* **42** 9–15. MR0353514

FIENBERG, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Anal.* **1** 1–40. MR2227361

FISHER, R. A. (1932). Inverse probability and the use of likelihood. *Math. Proc. Cambridge Philos. Soc.* **28** 257–261.

FISHER, R. A. (1934). Probability likelihood and quantity of information in the logic of uncertain inference. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **146** 1–8.

FOUSKAKIS, D., NTZOUFRAS, I. and DRAPER, D. (2015). Power-expected-posterior priors for variable selection in Gaussian linear models. *Bayesian Anal.* **10** 75–107. MR3420898

GOOD, I. J. (1958). Significance tests in parallel and in series. *J. Amer. Statist. Assoc.* **53** 799–813. MR0103560

GOOD, I. J. (1979). Studies in the history of probability and statistics. XXXVII. A. M. Turing's statistical work in World War II. *Biometrika* **66** 393–396. MR0548210

GOOD, I. J. (1980). The contributions of Jeffreys to Bayesian statistics. In *Bayesian Analysis in Econometrics and Statistics*:

Alexander Etz is a graduate research fellow in the Department of Cognitive Sciences, University of California, Irvine, Irvine, California 92617, USA (e-mail: Etz.Alexander@gmail.com). Eric-Jan Wagenmakers is Professor at the Methodology Unit of the Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 VZ Amsterdam, The Netherlands (e-mail: EJ.Wagenmakers@gmail.com).

Essays in Honor of Harold Jeffreys (A. Zellner, ed.) 21–34. North-Holland, Amsterdam.

GOOD, I. J. (1988). The interface between statistics and philosophy of science. *Statist. Sci.* **3** 386–412. MR0996411

HALDANE, J. B. S. (1919). The probable errors of calculated linkage values, and the most accurate method of determining gametic from certain zygotic series. *J. Genet.* **8** 291–297.

HALDANE, J. B. S. (1927). *Possible Worlds*: *And Other Essays* **52**. Chatto & Windus.

HALDANE, J. B. S. (1932). A note on inverse probability. *Math. Proc. Cambridge Philos. Soc.* **28** 55–61.

HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small. *Biometrika* **29** 133–143. MR0478438

HALDANE, J. B. S. (1938). The approximate normalization of a class of frequency distributions. *Biometrika* **29** 392–404.

HALDANE, J. B. S. (1945). On a method of estimating frequencies. *Biometrika* **33** 222–225. MR0019900

HALDANE, J. B. S. (1948). The precision of observed values of small frequencies. *Biometrika* **35** 297–300.

HALDANE, J. B. S. (1951). A class of efficient estimates of a parameter. *Bull. Inst. Internat. Statist.* **23** 231–248. MR0088845

HALDANE, J. B. S. (1966). An autobiography in brief. *Perspect. Biol. Med.* **9** 476–481.

HALDANE, J. B. S. and SMITH, C. A. B. (1947). A simple exact test for birth-order effect. *Ann. Eugen.* **14** 117–124.

HODGKIN, D. C. and JEFFREYS, H. (1976). Obituary. *Nature* **260** 564.

HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14** 382–417. MR1765176

HOLMES, C. C., CARON, F., GRIFFIN, J. E. and STEPHENS, D. A. (2015). Two-sample Bayesian nonparametric hypothesis testing. *Bayesian Anal.* **10** 297–320. MR3420884

HOWIE, D. (2002). *Interpreting Probability*: *Controversies and Developments in the Early Twentieth Century*. Cambridge Univ. Press, Cambridge. MR2362916

JEFFREYS, H. (1931). *Scientific Inference*, 1st ed. Cambridge Univ. Press, Cambridge.

JEFFREYS, H. (1933). On the prior probability in the theory of sampling. *Math. Proc. Cambridge Philos. Soc.* **29** 83–87.

JEFFREYS, H. (1935). Some tests of significance, treated by the theory of probability. *Math. Proc. Cambridge Philos. Soc.* **31** 203–222.

JEFFREYS, H. (1936a). Further significance tests. *Math. Proc. Cambridge Philos. Soc.* **32** 416–445.

JEFFREYS, H. (1936b). XXVIII. On some criticisms of the theory of probability. *The London*, *Edinburgh*, *and Dublin Philosophical Magazine and Journal of Science* **22** 337–359.

JEFFREYS, H. (1939). *Theory of Probability*. Oxford Univ. Press, Oxford. MR0000924

JEFFREYS, H. (1977). Probability theory in geophysics. *IMA J. Appl. Math.* **19** 87–96. MR0434330

JEFFREYS, H. (1980). Some general points in probability theory. In *Bayesian Analysis in Econometrics and Statistics*: *Essays in Honor of Harold Jeffreys* (A. Zellner, ed.) 451–453. North-Holland, Amsterdam.

KASS, R. E. (2009). Comment: The importance of Jeffreys's legacy. *Statist. Sci.* **24** 179–182. MR2655844

KASS, R. E. and RAFTERY, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90** 773–795. MR3363402

KENDALL, M. G., BERNOULLI, D., ALLEN, C. G. and EULER, L. (1961). Studies in the history of probability and statistics: XI. Daniel Bernoulli on maximum likelihood. *Biometrika* **48**.

LAI, D. C. (1998). John Burdon Sanderson Haldane (1892–1964) polymath beneath the firedamp: The story of JBS Haldane. *Bull. Anesth. History* **16** 3–7.

LAMBERT, J. H. and DILAURA, D. L. (2001). *Photometry*, *or*, *on the Measure and Gradations of Light*, *Colors*, *and Shade*: *Translation from the Latin of Photometria*, *Sive*, *de Mensura et Gradibus Luminis*, *Colorum et Umbrae*. Illuminating Engineering Society of North America, New York.

LANE, D. A. (1980). Fisher, Jeffreys, and the nature of probability. In *R. A. Fisher*: *An Appreciation*. *Lecture Notes in Statist.* **1** 148–160. Springer, New York. MR0578901

LATTIMER, J. M. and STEINER, A. W. (2014). Neutron star masses and radii from quiescent low-mass X-ray binaries. *Astrophys. J.* **784**, 123.

LINDLEY, D. V. (1957). Binomial sampling schemes and the concept of information. *Biometrika* **44** 179–186. MR0087273

LINDLEY, D. (2009). Comment [MR2655841]. *Statist. Sci.* **24** 183–184. MR2655845

LY, A., VERHAGEN, J. and WAGENMAKERS, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *J. Math. Psych.* **72** 19–32. MR3506022

MALIKOV, E., KUMBHAKAR, S. C. and TSIONAS, E. G. (2015). Bayesian approach to disentangling technical and environmental productivity. *Econometrics* **3** 443–465.

OSTROGRADSKI, M. V. (1848). On a problem concerning probabilities. *St. Petersburg Acad. Sci.* **6** 321–346.

PEARSON, K. (1892). *The Grammar of Science* **17**. Walter Scott.

PEIRCE, C. S. (1878). The probability of induction. *Pop. Sci. Mon.* **12** 705–718.

PIRIE, N. W. (1966). John Burdon Sanderson Haldane. 1892–1964. *Biogr. Mem. Fellows R. Soc.* **12** 219–249.

PREVOST, P. and LHUILIER, S. A. (1799). Sur les probabilités. *Mémoires de L'Academie Royale de Berlin* **1796** 117–142.

ROBERT, C. P. (2016). The expected demise of the Bayes factor. *J. Math. Psych.* **72** 33–37. MR3506023

ROBERT, C., CHOPIN, N. and ROUSSEAU, J. (2009). Harold Jeffreys's theory of probability revisited. *Statist. Sci.* **24** 141–172. MR2655841

SARKAR, S. (1992). A centenary reassessment of JBS Haldane, 1892–1964. *BioScience* **42** 777–785.

SENECHAL, M. (2012). *I Died for Beauty*: *Dorothy Wrinch and the Cultures of Science*. Oxford Univ. Press, Oxford.

SENN, S. (2009). Comment on "Harold Jeffreys's theory of probability revisited". *Statist. Sci.* **24** 185–186.

SMITH, J. M. (1992). JBS Haldane. In *The Founders of Evolutionary Genetics* 37–51. Springer, New York.

SPARKS, D. K., KHARE, K. and GHOSH, M. (2015). Necessary and sufficient conditions for high-dimensional posterior consistency under g-priors. *Bayesian Anal.* **10** 627–664. MR3420818

TARONI, F., MARQUIS, R., SCHMITTBUHL, M., BIEDERMANN, A., THIÉRY, A. and BOZZA, S. (2014). Bayes factor for investigative assessment of selected handwriting features. *Forensic Sci. Int.* **242** 266–273.

TERROT, C. (1853). Summation of a compound series, and its application to a problem in probabilities. *Trans. R. Soc. Edinb.* **20** 541–545.

TODHUNTER, I. (1858). *Algebra for the Use of Colleges and Schools*: *With Numerous Examples*. Cambridge Univ. Press, Cambridge.

TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability*: *From the Time of Pascal to That of Laplace/by I. Todhunter*. Macmillan and Company.

TURING, A. M. (1941/2012). The Applications of Probability to Cryptography. UK National Archives, HW 25/37.

VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theorestischen Physik*. Franz Deuticke.

WHITE, M. J. D. (1965). JBS Haldane. *Genetics* **52** 1.

WRINCH, D. and JEFFREYS, H. (1919). On some aspects of the theory of probability. *Philos. Mag.* **38** 715–731.

WRINCH, D. and JEFFREYS, H. (1921). On certain fundamental principles of scientific inquiry. *Philos. Mag.* **42** 369–390.

WRINCH, D. and JEFFREYS, H. (1923a). On certain fundamental principles of scientific inquiry (second paper). *Philos. Mag.* **45** 368–374.

WRINCH, D. and JEFFREYS, H. (1923b). The theory of mensuration. *Philos. Mag.* **46**.

ZABELL, S. (1989). The rule of succession. *Erkenntnis* **31** 283–321.

ZABELL, S. (2012). Commentary on Alan M. Turing: The applications of probability to cryptography. *Cryptologia* **36** 191–214.