

Bibliography

- Affi, A. A. and Elashoff, R. M. (1966). Missing observations in multivariate statistics. I: Review of the literature (ref: V62 p10-29; V64 p 337-385). *Journal of the American Statistical Association*, 61:595–604.
- Baker, S. G. (1994a). Composite linear models for incomplete multinomial data. *Statistics in Medicine*, 13:609–622.
- Baker, S. G. (1994b). Regression analysis of grouped survival data with incomplete covariates: Nonignorable missing-data and censoring mechanisms. *Biometrics*, 50:821–826.
- Baker, S. G. and Laird, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse (corr: V83 p1232). *Journal of the American Statistical Association*, 83:62–69.
- Baker, S. G., Rosenberger, W. F., and DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine*, 11:643–657.
- Begun, J. M., Hall, W. J., Huang, W.-M., and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, 11:432–452.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, Cambridge, MA, USA.

- Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A Guide to Simulation*. New York: Springer-Verlag, 2n edition.
- Bravo, R., Soriano, V., and Martín, R. *et al.* (1995). Cuantificación de la viremia por el virus de la inmunodeficiencia humana en pacientes con progresión rápida y lenta de la enfermedad (in spanish). *Med Clin (Barcelona)*, 104:530–534.
- Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, 2:437–453.
- Cao, Y., Qin, L., and Zhang, L. *et al.* (1995). Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. *N Engl J Med*, 332:201–208.
- Catchpole, E. A. and Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika*, 84:187–196.
- Caylà, J., Jansà, J., and Artacoz, L. *et al.* (1993a). Predictors of AIDS in a cohort of HIV-infected patients with pulmonary or pleural tuberculosis. *Tuberc Lung Dis*, 74:113–120.
- Caylà, J., Jansà, J., and Iglesias, B. *et al.* (1993b). Epidemiología del síndrome de inmunodeficiencia adquirida en Barcelona (1981–1991)(I). Estudio descriptivo y de tendencias temporales (in spanish). *Med Clin*, 101:286–293.
- Caylà, J., Marcos, A., and Bedoya, A. *et al.* (1995). Differential characteristics of AIDS patients with a history of imprisonment. *Int J Epidemiol*, 24:1188–1196.
- Centers for Disease Control (1993). Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR*, 41(No RR–17).
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34:305–334.
- Collett, D. (1994). *Modelling Survival Data in Medical Research*. Chapman & Hall.
- Conaway, M. R. (1993). Non-ignorable non-response models for time-ordered categorical variables. *Applied Statistics*, 42:105–115.

- Cook, N. R. (1997). An imputation method for non-ignorable missing data in studies of blood pressure. *Statistics in Medicine*, 16:2713–2728.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological*, 34:187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62:269–276.
- Curtis, R., Friedman, S., and Neaigus, A. *et al.* (1994). Implication of directly observed therapy in tuberculosis control measures among IDUs. *Public Health Rep*, 109:321–327.
- Darby, S., Ewart, and Giangrande, P. *et al.* (1996). Importance of age at infection with HIV-1 for survival and development of AIDS in UK haemophilia population. *Lancet*, 347:1573–1579.
- De Cock, K. (1995). Interaction HIV/TB. Conference on Global Lung Health and the 1995 Annual Meeting on the IUATLD/UICTMR, Paris. *Tuberc Lung Dis*, 76 (Suppl 2):5–6.
- de Maron, P. and García, A. (1993). La evolución de la infección VIH/SIDA en los países desarrollados. Impacto de la tuberculosis (in spanish). *Med Clin (Barcelona)*, 100:187–193.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (c/r: P22-37). *Journal of the Royal Statistical Society, Series B, Methodological*, 39:1–22.
- Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (disc: P73-93). *Applied Statistics*, 43:49–73.
- Diggle, P., Liang, K.-Y., and Zeger, S. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon.
- Diggle, P. J. (1989). Testing for random dropouts in repeated measurement data. *Biometrics*, 45:1255–1258.
- Efron, B. (1994). Missing data, imputation, and the bootstrap (disc: P475-479). *Journal of the American Statistical Association*, 89:463–475.

- EGRET (1990). *Epidemiological Graphics, Estimation and Testing Package*. Seattle, Washington: Statistics and Epidemiology Research Corporation.
- Ekholm, A. and Skinner, C. (1998). The muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics*, 47:251–263.
- Ellner, J. (1990). Tuberculosis in the time of AIDS. *Chest*, 98:1051–1052.
- Fernández-Nogués, F., Garau, J., and Gudiol, F. *et al.* (1983). Incidencia y diagnóstico de la tuberculosis pulmonar en un servicio de medicina interna (in spanish). *Med Clin (Barcelona)*, 81:651–654.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley.
- Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, 77:270–278.
- Gill, R. D. and Robins, J. M. (1997). Sequential models for coarsening and missingness. In Lin, D. and Fleming, T., editors, *The First Seattle Symposium in Biostatistics: Survival Analysis*, pages 295–305. Springer-Verlag New York.
- Gill, R. D., Van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In Lin, D. and Fleming, T., editors, *The First Seattle Symposium in Biostatistics: Survival Analysis*, pages 255–294. Springer-Verlag New York.
- Glynn, R. J., Laird, N. M., and Rubin, D. B. (1986). Mixture modeling versus selection modeling with nonignorable nonresponse. In Wainer, H., editor, *Drawing Inferences from Self-Selected Samples*, pages 115–142. New York: Springer-Verlag.
- Glynn, R. J., Laird, N. M., and Rubin, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. *Journal of the American Statistical Association*, 88:984–993.
- Gómez, G. and Serrat, C. (1999). Estudios de supervivencia con datos no observados. Dificultades inherentes al enfoque paramétrico. *Qüestió*, 23(2):365–392.

- Gourevitch, M., Wasserman, W., and Panero, M. *et al.* (1996). Successful adherence to observed prophylaxis and treatment of tuberculosis among drug users in a methadone program. *J Addictive Dis*, 15:93–104.
- Grupo de Trabajo sobre Tuberculosis (FIS-MSD) (1992). Consenso nacional para el control de la tuberculosis en España (in spanish). *Med Clin (Barcelona)*, 98:24–31.
- Heitjan, D. F. (1993). Ignorability and coarse data: Some biomedical examples. *Biometrics*, 49:1099–1109.
- Heitjan, D. F. and Rubin, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, 85:304–314.
- Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics*, 19:2244–2253.
- Hogan, J. and Laird, N. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16:239–258.
- Jones, B., Young, S., and Antoniskis, D. *et al.* (1993). Relationship of the manifestations of tuberculosis to CD4 cell counts in patients with human immunodeficiency virus infection. *Am Rev Respir Dis*, 148:1292–1297.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.
- Kastner, C., Fieger, A., and Heumann, C. (1997). MAREG and WinMAREG. A tool for marginal regression models. *Computational Statistics and Data Analysis*, 24:235–241.
- Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statistics in Medicine*, 17:2723–2732.
- Kenward, M. G. and Molenberghs, G. (1999). Parametric model for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research*, 8:51–83.

- Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7:305–315.
- Laird, N. M. (1993). The EM algorithm. In *Handbook of Statistics Volume 9: Computational Statistics*, pages 509–520.
- Levy, J. (1993). HIV pathogenesis and long-term survival. *AIDS*, 7:1401–1410.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.
- Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, 88:1341–1349.
- Lipsitz, S. R. and Ibrahim, J. G. (1996a). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, 83:916–922.
- Lipsitz, S. R. and Ibrahim, J. G. (1996b). Using the EM-algorithm for survival data with incomplete categorical covariates. *Lifetime Data Analysis*, 2:5–14.
- Lipsitz, S. R. and Ibrahim, J. G. (1998). Estimating equations with incomplete categorical covariates in the Cox model. *Biometrics*, 54:1002–1013.
- Lipsitz, S. R., Ibrahim, J. G., and Zhao, L. P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association*, 94:1147–1160.
- Little, R. J. A. (1992). Regression with missing X 's: A review. *Journal of the American Statistical Association*, 87:1227–1237.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88:125–134.
- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81:471–483.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90:1112–1121.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley.

- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*, 44:226–233.
- Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. New York: Wiley.
- Macroft, A., Johnson, M., and Philips, A. (1996). Factors affecting survival in patients with the acquired immunodeficiency syndrome. *AIDS*, 10:1057–1067.
- Marco, A., Guerrero, R., and Rodríguez, A. *et al.* (1995). Características de los casos de sida detectados en una prisión de Barcelona (1991–1993) (in spanish). *Atención Primaria*, 15:487–490.
- McCullagh, P. and Nelder, J. (1983). *Generalized Linear Models*. London: Chapman & Hall.
- Meilijson, I. (1989). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society, Series B, Methodological*, 51:127–138.
- Meng, X.-L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, 86:899–909.
- Meng, X.-L. and Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79:103–111.
- Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80:267–278.
- Michiels, B., Molenberghs, G., and Lipsitz, S. R. (1999a). A pattern-mixture odds ratio model for incomplete categorical data. *Communications in Statistics, Theory Methods*, 28(12):2843–2869.
- Michiels, B., Molenberghs, G., and Lipsitz, S. R. (1999b). Selection models and pattern-mixture models for incomplete data with covariates. *Biometrics*, 55:978–983.
- Miller, R. (1980). *Simultaneous Statistical Inference*. New York: Springer-Verlag.

- Milles, G. and Jones, D. (1993). Relation between CD4 lymphocyte count and AIDS mortality, 1986–91. *AIDS*, 7:1383–1386.
- Molenberghs, G. and Goetghebeur, E. (1995). Incomplete categorical data: Some issues. Technical report, ECAS.
- Molenberghs, G., Goetghebeur, E. J., Lipsitz, S. R., and Kenward, M. G. (1999). Nonrandom missingness in categorical data: Strengths and limitations. *The American Statistician*, 53:110–118.
- Molenberghs, G., Kenward, M. G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics*, 50(1):15–29.
- Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika*, 84:33–44.
- Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52(2):153–161.
- Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5:99–135.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In McFadden, D. and Engler, R., editors, *Handbook of Econometrics*, volume IV, pages 2111–2245. Amsterdam: North-Holland.
- Nielsen, S. F. (1997). Inference and missing data: Asymptotic results. *Scandinavian Journal of Statistics*, 24:261–274.
- Nielsen, S. F. (1998). Semi-parametric regression with coarsely observed regressors. Technical report, Department of Theoretical Statistics. University of Copenhagen.
- Norussis, M. (1986). *SPSS/PV+ for the IBM PC/XT/AT*. Chicago: SPSS Inc.
- Ocaña, I., Ribera, E., and Martínez Vazquez, J. (1996). Diagnostic thoracocentesis and pleural biopsy. *Ann Intern Med*, 105:297–298.

- Paik, M. C. (1995). The EM algorithm for the quasi-likelihood regression model. *Communications in Statistics, Part A – Theory and Methods*, 24:1403–1430.
- Paik, M. C. (1996). Quasi-likelihood regression models with missing covariates. *Biometrika*, 83:825–834.
- Paik, M. C. and Tsai, W.-Y. (1997). On using the Cox proportional hazards model with missing covariates. *Biometrika*, 84:579–593.
- Pape, J., Jean, J., and Ho, L. *et al.* (1993). Effect of isoniazid prophylaxis on incidence of active tuberculosis and progression of HIV infection. *Lancet*, 342:268–272.
- Phair, J., Muñoz, A., and Detels, R. *et al.* (1990). The risk of *pneumocystis carinii* pneumonia among men infected with Human Immunodeficiency Virus Type 1. *N Engl J Med*, 322:161–165.
- Poznasky, M., Coker, R., and Skinner, C. *et al.* (1995). HIV positive patients first presenting with an AIDS defining illness: characteristics and survival. *Br Med J*, 311:156–158.
- Pugh, M. G. (1993). *Inference in the Cox Proportional Hazard Model with Missing Covariate Data*. PhD thesis, Harvard School of Public Health. Harvard University.
- Raistrick, D., Hay, A., and Wolff, K. (1996). Methadone maintenance and tuberculosis treatment. *Br Med J*, 313:925–926.
- Raviglione, M., Snider, D., and Kochi, A. (1995). Global epidemiology of tuberculosis. *JAMA*, 273:220–226.
- Reilly, M. and Pepe, M. (1995). A mean score method for missing and auxiliary covariate data in regression models. *Biometrika*, 82:299–314.
- Richter, C., Koelemay, M., and Swai, A. *et al.* (1995). Predictive markers of survival in HIV-seropositive and HIV-seronegative Tanzanian patients with extrapulmonary tuberculosis. *Tuberc Lung dis*, 76:510–517.

- Robins, J. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In Jewell, N. P., Dietz, K., and Farewell, V. T., editors, *AIDS Epidemiology. Methodological Issues*, pages 297–331. Boston: Birkhäuser.
- Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *ASA Proceedings of the Biopharmaceutical Section*, pages 24–33.
- Robins, J. M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine*, 16:21–37.
- Robins, J. M. and Gill, R. D. (1997). Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine*, 16:39–56.
- Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90:122–129.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90:106–121.
- Rotnitzky, A. (1996). Curso sobre inferencias paramétricas y no paramétricas con datos no observados (in spanish). Barcelona: Universitat Politècnica de Catalunya.
- Rotnitzky, A. and Robins, J. M. (1995a). Semi-parametric estimation of models for means and covariances in the presence of missing data. *Scandinavian Journal of Statistics*, 22:323–333.
- Rotnitzky, A. and Robins, J. M. (1995b). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika*, 82:805–820.

- Rotnitzky, A. and Robins, J. M. (1997). Analysis of semi-parametric regression models with non-ignorable non-response. *Statistics in Medicine*, 16:81–102.
- Rotnitzky, A., Robins, J. M., and Scharfstein, D. (1998). Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association*, 93:1321–1339.
- Rotnitzky, A. and Wypij, D. (1994). A note on the bias of estimators with missing data. *Biometrics*, 50:1163–1170.
- Rousseeuw, P. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63:581–590.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.
- Rubin, D. B. (1991). EM and beyond. *Psychometrika*, 56:241–254.
- Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, 81:366–374.
- Rubin, D. B., Stern, H. S., and Vehovar, V. (1995). Handling “Don’t Know” survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association*, 90:822–828.
- Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Monographs on Statistics and Applied Probability (72). Chapman & Hall.
- Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (c/r: P1121–1146). *Journal of the American Statistical Association*, 94:1096–1120.
- Schervish, M. J. (1995). *Theory of Statistics*. Springer Series in Statistics. New York: Springer-Verlag.
- Schluchter, M. D. and Jackson, K. L. (1989). Log-linear analysis of censored survival data with partially observed covariates. *Journal of the American Statistical Association*, 84:42–52.

- Serrat, C. and Gómez, G. (1995). Estimació de paràmetres amb dades absents en un estudi d'anàlisi de supervivència (in catalan). Technical Report 95/10, Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya, Spain.
- Serrat, C., Gómez, G., García, P., and Caylà, J. (1998). CD4+ lymphocytes and tuberculin skin test as survival predictors in pulmonary tuberculosis HIV-infected patients. *International Journal of Epidemiology*, 27:703–712.
- Skinner, C. J. and Coker, O. (1996). Regression analysis for complex survey data with missing values of a covariate. *Journal of the Royal Statistical Society, Series A, General*, 159:265–274.
- Stoneburrer, R., Laroche, E., and Prevots, R. *et al.* (1992). Survival in a cohort of human immunodeficiency virus infected tuberculosis patients in New York City. Implications for expansion of the AIDS case definition. *Arch Intern Med*, 152:2033–2037.
- Van Buuren, S. and Oudshoorn, K. (2000). Multivariate Imputation by Chained Equations: MICE V1.0 User's manual. Technical Report TNO Report PG/VGZ/00.038, TNO Prevention and Health. Leiden. The Netherlands.
- Vella, S., Giuliano, M., and Florida, M. *et al.* (1995). Effect of sex, age and transmission category on the progression to AIDS and survival of zidovudine-treated symptomatic patients. *AIDS*, 9:51–55.
- Whalen, C., Horsburch, C., and Hom, D. *et al.* (1995). Accelerated course of human immunodeficiency virus infection after tuberculosis. *Am J Respir Crit Care Med*, 151:129–135.
- Whalen, C., Okwera, A., and Johnson, J. *et al.* (1994). Prognosis and survival in HIV-infected individuals tuberculosis patients: comparison of US and Uganda cohorts. *Tuberc Lung Dis*, 75 (Suppl 1):66.
- World Health Organization (1989). Tuberculosis and AIDS. Statement on AIDS and tuberculosis. Geneve 1989. *Bull Int Union Tuberc Lung Dis*, 64:8–11.
- World Health Organization (1994). *TB A Global Emergency*. Geneve: WHO/TB/94 177.

- World Health Organization (1996). *TB Groups at Risk*. Geneva: WHO.
- Wu, C. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, 11:95–103.
- Zhao, L. P., Lipsitz, S. R., and Lew, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics*, 52:1165–1182.