STATISTICAL POLICY
WORKING PAPER 25

Data Editing Workshop and Exposition

Prepared by the
Organizing Committee for the
Data Editing Workshop and Exposition
Federal Committee on Statistical Methodology

Statistical Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget

DECEMBER 1996

Members of the Federal Committee on Statistical Methodology

(MARCH 1996)

Maria Elena Gonzalez, Chair
Office of Management and Budget

M. Denice McCormick Myers, Secretary
National Agricultural Statistics Service

Susan Ahmed
National Center for Education Statistics

Yvonne M. Bishop
Energy Information Administration

Cynthia Z. F. Clark
Bureau of the Census

Steven Cohen
Agency for Health Care Policy and Research

Lawrence H. Cox
Environmental Protection Agency

Zahava D. Doering
Smithsonian Institution

Daniel Kasprzyk
National Center for Education Statistics

Nancy Kirkendall
Energy Information Administration

Daniel Melnick
National Science Foundation

Robert P. Parker
Bureau of Economic Analysis

Charles P. Pautler, Jr.
Bureau of the Census

David A. Pierce
Federal Reserve Board

Thomas J. Plewes
Bureau of Labor Statistics

Wesley L. Schaible
Bureau of Labor Statistics

Rolf Schmitt
Bureau of Transportation Statistics

Monroe Sirken
National Center for Health Statistics

Alan R. Tupek
National Science Foundation

Denton Vaughan
Social Security Administration

Robert Warren
Immigration and Naturalization Service

G. David Williamson
Centers for Disease Control and Prevention

Foreword

This volume, No. 25 in the Federal Committee on Statistical Methodology (FCSM)'s Working Paper series, is the written record of the Data Editing Workshop and Exposition, held on March 22, 1996, at the Bureau of Labor Statistics (BLS) Conference and Training Center. This conference was over a year in planning, by an Organizing Committee that was an outgrowth of the FCSM's Subcommittee on Data Editing. From an initial plan of ten or 20 papers and computer software demonstrations, and perhaps100 attendees, the registrations and submissions kept growing until the final program consisted of 44 oral presentations and 19 software demonstrations on data editing, with over 500 conference attendees. This success is probably due to several causes, not the least of which were the many outstanding contributions by the authors whose work appears in this volume. Perhaps the high participation level also suggests that data editing has been an overlooked area in the work of Federal, state, and international statistical agencies, as well as private-sector organizations.

From the start it was our intention to plan and produce this conference on as close to a zero budget as possible. Our holding this goal seemed to foster an atmosphere of cooperation in which contributions and offers of assistance came forth from numerous sources and at the times they were most needed. From the early publicity provided by several agencies and collaborating organizations to the preparation by IRS of the works published in this volume, donations of time and effort were most generous. The BLS staff was truly outstanding in anticipating and handling the many physical arrangements for the Workshop. And the agencies listed as affiliations of the Organizing Committee members all contributed varying degrees of staff time towards ensuring the success of this conference.

A Dedication in Memory of
Maria Elena Gonzalez

We began the planning of the Data Editing Workshop and Exposition under the guidance of the FCSM and its founding chairperson, Maria Elena Gonzalez. After an illness, Maria passed away earlier this year and, while the FCSM continued its sponsorship of the conference and these *Proceedings*, Maria Gonzalez' departure is a deep personal and professional loss to all of us. Her career as a Federal government statistician spanned a quarter century, during which she made many contributions to improving the quality of Federal (and international) statistics. She did this both directly and as an outstanding leader in bringing forth and leveraging the talents of others for the many valuable statistical projects and conferences that she initiated. The editors would like to dedicate this *Proceedings* volume to Maria Gonzalez' memory, as was done by the Organizing Committee for the conference itself.

The next few pages contain the table of contents, followed by the conference contributions themselves. The conference program, including the list of sponsors and additional acknowledgments, is reproduced in an appendix.

David Pierce and Mark Pierzchala, Chairs
Organizing Committee for the Data
Editing Workshop and Exposition

Wendy Alvey and Bettye Jamerson, Editors
Proceedings of the Data Editing Workshop and Exposition

DECEMBER 1996

Data Editing Workshop and Exposition: Organizing Committee

David A. Pierce, Chair
Federal Reserve Board

Mark Pierzchala, Co-Chair
National Agricultural Statistics Service

Yahia Ahmed
U. S. Internal Revenue Service

Frances Chevarley
National Center for Health Statistics

Charles Day
National Agricultural Statistics Service

Rich Esposito
U. S. Bureau of Labor Statistics

Sylvia Kay Fisher
U. S. Bureau of Labor Statistics

Laura Bauer Gillis
Federal Reserve Board

Maria Elena Gonzalez
U. S. Office of Management and Budget

Robert Groves
Joint Program in Survey Methodology

Ken Harris
National Center for Health Statistics

David McDonell
National Agricultural Statistics Service

Renee Miller
U. S. Energy Information Administration

M. Denice McCormick Myers
National Agricultural Statistics Service

Jeff Owings
National Center for Education Statistics

Thomas B. Petska
U. S. Internal Revenue Service

Linda Stinson
U. S. Bureau of Labor Statistics

Paula Weir
U. S. Energy Information Administration

William E. Winkler
U. S. Bureau of the Census

TABLE OF CONTENTS

Page

FORWARD i

1 ▾ OVERVIEWS

A Paradigm Shift for Data Editing, *Linda M. Ball* 3
The New View on Editing, *Leopold Granquist* 16
Data Editing at the National Center for Health Statistics, *Kenneth W. Harris* 24

2 ▾ FELLEGI-HOLT SYSTEMS

DISCRETE: A Fellegi-Holt Edit System for Demographic Data, *William E. Winkler and Thomas F. Petkunas* 39
Generalized Edit and Imputation System for Numeric Data [ABSTRACT ONLY], *Joel Bissonnette* 49
The New SPEER Edit System, *William E. Winkler and Lisa R. Draper* 50

3 ▾ ON-SITE DATA CAPTURE

Electronic Data Interchange for Statistical Data Collection, *Wouter J. Keller and W. F. H. Ypma* 61
PERQS (Personalized Electronic Reporting Questionnaire System), *Janet Sear* 73
Electronic Data Collection: The Virginia Uniform Reporting System, *Anne Rhodes, Kishau Smith, and Peter Goldstein* 78

4 ▾ CASE STUDIES -- I

Toward A Unified System of Editing International Data [ABSTRACT ONLY], *Glen Ferri and Tom Ondra* 85
Data Editing Software for NSF Surveys, *Richard J. Bennof, M. Marge Machen, and Ronald L. Meeks* 86

5 ▾ CENSUSES

Automated Record Linkage and Editing: Essential Supporting Components in Data Capture Process, *Olivia Blum and Eliahu Ben-Moshe* 93
Editing and Imputation Research for the 2001 Census in the United Kingdom, *Jan Thomas and David Thorogood* 99
A Priority Index for Macro-Editing the Netherlands Foreign Trade Survey, *Frank van de Pol and Bert Diederen* 109

6 ▾ GRAPHICAL/INTERACTIVE SYSTEMS

Experiences on Changing to PC-based Visual Editing in Current Employment Statistics Program, *Bill Goodman, Laura Freeman, Mike Murphy, and Richard Esposito* 121
Graphical Editing Analysis Query System (GEAQS), *Paula Weir* 125
Integrated Data Capture: A System for All Office of Compensation and Working Conditions Surveys, *Mary Kelly* 137

Page

7 ▾ CATI-CAPI TECHNICAL

Questionnaire Programming Language (QPL) [ABSTRACT ONLY], *Kevin Dooley* 145
Using A Parallel CASES Instrument to Edit Call Record Information and Removal of Incorrect Interview Data, *Timothy Triplett and Beth Webb* 146
A Computer-Assisted Coding and Editing System for Non-Numeric Educational Transcript Data [ABSTRACT ONLY], *Stanley E. Legum* 153

8 ▾ STATISTICAL TECHNIQUES -- I

Rethinking the Editing Algorithm for the Survey of Employment Payrolls and Hours [ABSTRACT ONLY], *Michael Scrim* 157
A Statistical Edit for Livestock Slaughter Data [ABSTRACT ONLY], *Linda Simpson, Henry Chiang, and Cathy Tomczak* 158
A CSFII Data User's Principal Components Analysis for Outlier Detection, *Adeline J. Wilcox* 159

9 ▾ CASE STUDIES -- II

The Thin Yellow Line: Editing and Imputation in a World of Third-Party Artifacts, *Clifford Adelman* 173
Sampling Design and Estimation Properties of a Study of Perinatal Substance Exposure in California, *Jimmy Hwang, Bo Kolody, and William A. Vega* 182
The Processing and Editing System of the National Health Interview Survey: The Old and New [ABSTRACT ONLY], *Susan S. Jack* 195

10 ▾ NEURAL NETWORKS

Data Editing Using Neural Networks, *L. H. Roddick* 199
Editing Monthly Survey Data Using Neural Networks, *L. H. Roddick* 206
Editing and Imputation by Means of Neural Networks, *Svein Nordbotten* 213

11 ▾ CATI-CAPI CONCEPTUAL

Statistics Canada's Experience in Moving to CAI from Paper and Pencil [ABSTRACT ONLY], *R. Jamieson* 227
A Feasibility Test of On-Line Editing for Touchtone Data Entry Collection, *David O'Connell* 228
CAI and Interactive Editing in One System for a Survey in a Multi-mode Environment, *Mark Pierzchala* 234

12 ▾ STATISTICAL TECHNIQUES -- II

Time Series and Cross Section Edits, *David A. Pierce and Laura Bauer Gillis* 245
Inflation Factors for Stratified Samples with Control Information, *Peter Ochshorn* 259
Empirical Data Review: Objective Detection of Unusual Patterns of Data [ABSTRACT ONLY], *James Kennedy* 266

Page

13 ▾ CASE STUDIES -- III

Time-Series Editing of Quarterly Deposits Data, *Anusha Fernando Dharmasena* 269
Experiences in Re-Engineering the Approach to Editing and Imputing Canadian Imports Data, [ABSTRACT ONLY], *Clancy Barrett and Francois Laflamme* 283
Data Editing in an Automated Environment: A Practical Retrospective -- The CPS Experience [ABSTRACT ONLY], *Gregory D. Weyland* 284

14 ▾ STATISTICAL TECHNIQUES -- III

Statistical Analysis of Textual Information [ABSTRACT ONLY], *Thierry Delbecque, Sid Laxson, and Nathalie Millot* 287
The Impact of Ratio Weighting, *Jai Choi* 288
Fitting Square Text Into Round Computer Holes — An Approach to Standardizing Textual Responses Using Computer-Assisted Data Entry, *Richard Wendt, Irene Hall, Patricia Price-Green, V. Ramana Dhara, and Wendy E. Kaye* 296

15 ▾ EDIT AUTHORING TECHNIQUES

Methods of Reusing Edit Specifications Across Collection and Capture Modes and Systems, *Shirley Dolan* 303
CDC Edits: Tools for Writing Portable Edits, *J. Tebbel and T. Rawson* 313
Skip Patterns and Response Bases: Graph Manipulation in Survey Processing [ABSTRACT ONLY], *Robert F. Teitel* 318

16 ▾ SOFTWARE DEMONSTRATIONS 319

APPENDIX 331
Program 333
List of Attendees 339