

---

# Record Linkage Techniques -- 1997

---

Proceedings of an International  
Workshop and Exposition

---

**March 20-21, 1997**  
**Arlington, VA**

---



**Federal Committee on Statistical Methodology**  
**Office of Management and Budget**  
**Washington, DC 1997**

---



---

# Record Linkage Techniques -- 1997

---

## Proceedings of an International Workshop and Exposition

---

**Compiled and Edited by  
Wendy Alvey and Bettye Jamerson**

**A Collaborative Effort by:**

Bureau of the Census  
Ernst and Young, LLP  
Federal Committee on Statistical Methodology (OMB)  
Internal Revenue Service  
National Agricultural Statistics Service  
National Cancer Institute  
National Center for Health Statistics  
National Research Council  
National Science Foundation  
Statistics Canada  
Washington Statistical Society

---

## Record Linkage Techniques – 1997

---

### Program Committee

FRITZ SCHEUREN (*Co-Chair*), Ernst and Young, LLP  
WILLIAM WINKLER (*Co-Chair*), Bureau of the Census  
MAXINE ANDERSON-BROWN (*Workshop Coordinator*), Bureau of the Census  
WENDY ALVEY, Internal Revenue Service  
JOHN ARMSTRONG, Statistics Canada  
CHARLES DAY, National Agricultural Statistics Service  
JOHN TUCKER, National Research Council  
LAURA ZAYATZ, Bureau of the Census

### Workshop Support

SHELBY FOLGER, Bureau of the Census  
ANNA HOLAUS, Bureau of the Census  
MICHAEL L. LUCAS, Bureau of the Census  
CAROL MCDANIEL, Bureau of the Census  
RUTH O'BRIEN, National Research Council  
BARBARA WRIGHT, National Research Council

### Acknowledgments

The Program Committee would like to thank the Bureau of the Census; Ernst and Young, LLP; the Office of Management and Budget's Federal Committee on Statistical Methodology; the Internal Revenue Service; the National Agricultural Statistics Service; the National Cancer Institute; the National Center for Health Statistics; the National Research Council's Theoretical Statistics; the National Science Foundation; Statistics Canada; and the Washington Statistical Society for supporting this Workshop project.

In addition, the Co-Chairs would like to thank John Tucker and Ruth O'Brien, of the National Research Council, for their support; Carol McDaniel, at Census, for overseeing the arrangements for the Workshop; Wendy Alvey, at the Internal Revenue Service, and Bettye Jamerson, of Jamie Productions, for compiling this compendium; and Nancy Kirkendall, at the Office of Management and Budget, and the Federal Committee on Statistical Methodology for printing this volume.

This publication will soon be available electronically through the Office of Management and Budget Web site. It can be reached through <http://www.fedstats.gov/>; click on *Policy*; click on *Statistical Policy Working Papers*; and go down to *Other Multi-Agency Papers*.

# Record Linkage Techniques – 1997

---

## Foreword

### Introduction

On March 20-21 of this year, a two-day International Record Linkage Workshop and Exposition was held in Arlington, Virginia. Over 200 hundred people were in attendance; and, because the facilities were limited, we had to turn away yet another 200 interested individuals.

The Workshop had two main goals: first, we wanted an occasion to celebrate **Howard Newcombe's** pioneering practical work on computerized record linkage, which began in the 1950's -- see, e.g., Newcombe, Kennedy, Axford, and James (1959), Automatic Linkage of Vital Records, which appeared in *Science* -- and the theoretical underpinnings of his work, which were formalized in the 1960's by **Ivan Fellegi** and **Alan Sunter** in their classic 1969 paper on A Theory for Record Linkage, published in the *Journal of the American Statistical Association*. Second, we wanted a way to broadly update the methodological and technological developments in record linkage research and their applications since the Workshop on Exact Matching Methodologies, which was held in Washington, DC, in March 1985. The Proceedings from that earlier conference -- *Record Linkage Techniques -- 1985* -- have been widely cited; but much new work has been done since then.

### Contents

The current volume is fortunate to present recent papers by two of the pioneers in record linkage research -- Fellegi and Newcombe -- as well as the work of many others who are exploring various aspects of exact matching techniques. Some of the new areas of related research involve increased privacy concerns due to record linkage; the growing interest in record linkage as a means for more efficient use of scarce statistical resources; the heightened importance of linkage technology, for such policy areas as health care reform; issues in physical security of data; and the measurement of nondisclosure and reidentification risks in public-use microdata files.

The format for this volume essentially follows that of the 1997 Workshop agenda -- with a section added to highlight key contributions made to the literature since the 1985 *Proceedings*. In those few cases where a paper was not available, the Conference abstract is provided. The report concludes with an Appendix, listing attendees who participated in the March 1997 Workshop and accompanying software expositions.

### Copy Preparation and Reviews

The contents of the papers included here are the responsibility of the authors -- any opinions, findings, conclusions, or recommendations expressed in these material are those of the authors and do not necessarily reflect the views of the sponsors. With the exception of selected previously published papers, which were simply reproduced (with permission) as is, all of the papers in this volume underwent only a limited editorial review. Since this did not constitute a formal referee process, authors were also encouraged to obtain their own

---

# Record Linkage Techniques – 1997

---

---

technical review. Corrections and changes were either made by the authors, themselves, or cleared through them by the editors. Final layout of the papers was done by the editors, Wendy Alvey and Bettye Jamerson. Minor changes of a cosmetic nature were considered the prerogative of the editors.

## Obtaining *Record Linkage Techniques* -- 1985

Ur  
wi  
for  
is :  
At  
on

Of

In  
mi  
Me

Al  
lin  
and  
Wi

Sc

Th  
Yc  
Int  
Na  
Ca

R

# Tribute



**Howard Newcombe**



This workshop is dedicated to three innovative pioneers in the field of record linkage—to Howard Newcombe, for his seminal work in the practice of record linkage, begun over 40 years ago; and to Ivan Fellegi and Alan Sunter, for their joint theoretical work 30 years ago. Together they established, in large measure, what we still do today. The future will see us continuing to build on their work.



**Ivan Fellegi**



---

# Record Linkage Techniques – 1997

---

## Table of Contents

	Page
<b>Foreword</b> .....	ix
■ <b>Chapter 1</b>	
<b>Keynote Address</b>	
Record Linkage and Public Policy – A Dynamic Evolution, <i>Ivan P. Fellegi</i> .....	3
■ <b>Chapter 2</b>	
<b>Invited Session on Record Linkage Applications for Epidemiological Research</b>	
OX-LINK: The Oxford Medical Record Linkage System, <i>Leicester E. Gill</i> .....	15
Complex Linkages Made Easy, <i>John R. H. Charlton and Judith D. Charlton</i> .....	34
Tips and Techniques for Linking Multiple Data Systems: The Illinois Department of Human Services Consolidation Project, <i>John Van Voorhis, David Koepke,     and David Yu [Abstract only]</i> .....	45
■ <b>Chapter 3</b>	
<b>Contributed Session on Applications of Record Linkage</b>	
Linking Administrative Records Over Time: Lessons from the Panels of Tax Returns, <i>John L. Czajka</i> .....	49
Record Linkage of Census and Routinely Collected Vital Events Data in the ONS Longitudinal Study, <i>Lin Hattersley</i> .....	57
Use of Probabilistic Linkage for an Analysis of the Effectiveness of Safety Belts and Helmets, <i>Dennis Utter</i> .....	67
Multiple Causes of Death for the National Health Interview Survey, <i>John Horm</i> .....	71



---

	<b>Page</b>
<b>■ Chapter 4</b>	
<b>Invited Session on Record Linkage Methodology</b>	
A Method for Calibrating False-Match Rates in Record Linkage, <i>Thomas R. Belin and Donald B. Rubin</i> .....	81
Modeling Issues and the Use of Experience in Record Linkage <i>Michael D. Larsen</i> .....	95
Regression Analysis of Data Files that are Computer Matched -- Part I, <i>Fritz Scheuren and William E. Winkler</i> .....	106
Regression Analysis of Data Files that are Computer Matched -- Part II, <i>Fritz Scheuren and William E. Winkler</i> .....	126
<b>■ Chapter 5</b>	
<b>Contributed Session on Confidentiality and Strategies for Record Linkage</b>	
Using Record Linkage to Thwart the Data Terrorist, <i>Robert Burton [Abstract only]</i> .....	141
$\mu$ - and $\tau$ -ARGUS: Software for Statistical Disclosure Control, <i>Anco J. Hundepool and Leon C. R. J. Willenborg</i> .....	142
Investigating Auto Injury Treatment in a No-Fault State: An Analysis of Linked Crash and Auto Insurer Data, <i>Lawrence H. Nitz and Karl E. Kim</i> .....	150
Quantitative Evaluation of the Linkage Operations of the 1996 Census Reverse Record Check, <i>Julie Bernier</i> .....	160
<b>■ Chapter 6</b>	
<b>Invited Session on Business and Miscellaneous Record Linkage Applications</b>	
Linking Federal Estate Tax Records, <i>Jenny B. Wahl</i> .....	171
Post-Matching Administrative Record Linkage Between Sole Proprietorship Tax Returns and the Standard Statistical Establishment List, <i>Philip M. Steel and Carl A. Konschnik</i> .....	179
Approximate String Comparison and its Effect in an Advanced Record Linkage System, <i>Edward H. Porter and William E. Winkler</i> .....	190

---

■ **Chapter 7**

**Contributed Session on More Applications of Probabilistic Record Linkage**

A Comparison of Direct Match and Probabilistic Linkage in the Death Clearance of the Canadian Cancer Registry, <i>Tony LaBillois, Marek Wysocki, and Frank J. Grabowiecki</i> .....	203
Improvements in Record Linkage Processes for the Bureau of Labor Statistics' Business Establishment List, <i>Kenneth Robertson, Larry Huff, Gordon Mikkelson, Timothy Pivetz, and Alice Winkler</i> .....	212
Technical Issues Related to the Probabilistic Linkage of Population-Based Crash and Injury Data, <i>Sandra Johnson</i> .....	222
Record Linkage of Progress Towards Meeting the New Jersey High School Proficiency Testing Requirements, <i>Eva Miller</i> .....	227

■ **Chapter 8**

**Invited Session on Confidentiality**

Public Attitudes Toward Data Sharing by Federal Agencies, <i>Eleanor Singer, John VanHoewyk, and Stanley Presser</i> .....	237
Multiple Imputation and Disclosure Protection: The Case of the 1995 Survey of Consumer Finances, <i>Arthur B. Kennickell</i> .....	248
Sharing Statistical Information for Statistical Purposes, <i>Katherine K. Wallman and Jerry L. Coffey</i> .....	268

■ **Chapter 9**

**Contributed Session on Methods and Plans for Record Linkage**

Linking Records on Federal Housing Administration Single-Family Mortgages, <i>Thomas N. Herzog and William J. Eilerman [Abstract only]</i> .....	279
Using Microsoft Access to Perform Exact Record Linkages, <i>Scott Meyer</i> .....	280
Analysis of Immigration Data: 1980-1994, <i>Adam Probert, Robert Semenciw, Yang Mao, and Jane F. Gentleman</i> .....	287

---

---

**■ Chapter 10**
**Invited Session on More Record Linkage Applications in Epidemiology**

Record Linkage Methods Applied to Health Related Administrative Data Sets Containing Racial and Ethnic Descriptors, <i>Selma C. Kunitz, Clara Lee, Rene C. Kozloff, and Harvey Schwartz</i> .....	295
Matching Census Database and Manitoba Health Care Files, <i>Christian Houle, Jean-Marie Berthelot, Pierre David, Michael C. Wolfson, Cam Mustard, and Leslie Roos</i> .....	305
The Development of Record Linkage in Scotland: The Responsive Application of Probability Matching, <i>Steve Kendrick</i> .....	319

**■ Chapter 11****Selected Related Papers, 1986-1997**

The Use of Names for Linking Personal Records, <i>Howard B. Newcombe, Martha E. Fair, and Pierre Lalonde</i> .....	335
Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida, <i>Matthew A. Jaro</i> .....	351
Record Linkage and Genealogical Files, <i>Nancy P. NeSmith</i> .....	358
A Review of the Statistics of Record Linkage for Genealogical Research, <i>David White</i> .....	362
Matching and Record Linkage, <i>William E. Winkler</i> .....	374
Linking Health Records: Human Rights Concerns, <i>Fritz Scheuren</i> .....	404
Record Linkage in an Information Age Society, <i>Martha E. Fair</i> .....	427
Computational Disclosure Control for Medical Microdata: The Datafly System, <i>Latanya Sweeney</i> .....	442

**■ Chapter 12****Tutorial on Record Linkage**

<i>Martha E. Fair and Patricia Whitridge</i> .....	457
--	-----

---

	<b>Page</b>
<b>■ Chapter 13</b>	
<b>A Checklist for Evaluating Record Linkage Software</b>	
<i>Charles Day</i> .....	483
<b>Software Demonstrations:</b>	
MatchWare Product Overview, <i>Matthew A. Jaro</i> .....	489
$\mu$ - and $\tau$ -ARGUS: Software Packages for Statistical Disclosure Control, <i>Anco J. Hundepool, Agnes Wessels, and Lars van Gemerden</i> .....	490
OXLINK: The Oxford Medical Record Linkage System Demonstration of the PC Version, <i>Leicester E. Gill</i> .....	491
Software for Record Linkage of Primary Care Data, <i>John R. H. Charlton</i> .....	491
GRLS – Record Linkage, <i>Kathy Zilahi</i> .....	492
<b>Appendix</b>	
List of Attendees .....	495