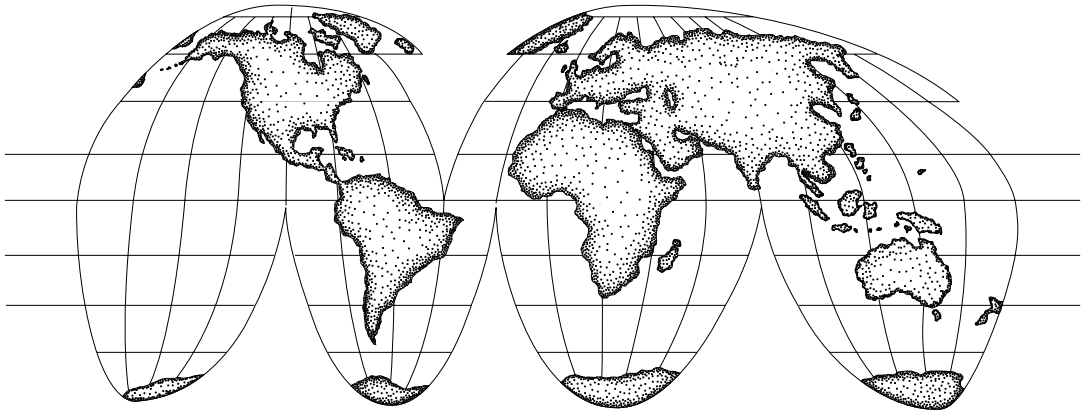
Proceedings of the International Collaborative Effort on Automating Mortality Statistics, Volume I

From the CENTERS FOR DISEASE CONTROL AND PREVENTION/National Center for Health Statistics



Kimberley Peters, Editor

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
July 1999
DHHS Publication No. (PHS) 99-1252

Library of Congress-in-Publication Data

International Collaborative Effort on Automating Mortality Statistics
(1st : 1996 : Washington, D.C.)
Proceedings of the International Collaborative Effort on Automating Mortality Statistics.
p. cm. — (DHHS publication ; no. (PHS) 99-1252)
"July, 1999."
"On November 12–15, 1996, the National Center for Health Statistics convened the first conference of the International Collaborative Effort (ICE) on Automating Mortality Statistics. The United States Agency for International Development provided support."
Includes bibliographical references.
ISBN 0-8406-0548-X
1. Mortality—Statistical methods—Data processing—Congresses.
I. National Center for Health Statistics (U.S.) II. Title. III. Series.
HB1321.I53 1996
614.4'2'0285—dc21 98-25577
CIP

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402-9328
Printed on acid-free paper.

National Center for Health Statistics

Edward J. Sondik, Ph.D., *Director*

Jack R. Anderson, *Deputy Director*

Jennifer H. Madans, Ph.D., *Associate Director for Science*

Jennifer H. Madans, Ph.D., *Acting Associate Director for Vital and Health Statistics Systems*

Edward L. Hunter, *Associate Director for Planning, Budget, and Legislation*

Jack R. Anderson, *Acting Associate Director for International Statistics*

Lester R. Curtin, Ph.D., *Acting Associate Director for Research and Methodology*

Jennifer H. Madans, Ph.D., *Acting Associate Director for Analysis, Epidemiology, and Health Promotion*

P. Douglas Williams, *Acting Associate Director for Data Standards, Program Development, and Extramural Programs*

Stephen E. Nieberding, *Associate Director for Management*

Charles J. Rothwell, *Associate Director for Data Processing and Services*

Division of Vital Statistics

Mary Anne Freedman, *Director*

James A. Weed, Ph.D., *Deputy Director*

Harry M. Rosenberg, Ph.D., *Chief, Mortality Statistics Branch*

Preface

On November 12-15, 1996, the Centers for Disease Control's National Center for Health Statistics (NCHS) convened the first conference of the International Collaborative Effort (ICE) on Automating Mortality Statistics. This conference was sponsored by NCHS, with support from the U.S. Agency for International Development.

The mission of the ICE on Automating Mortality Statistics is (1) to share knowledge and experience of automated systems for coding mortality information, (2) to develop and improve existing automated systems through collaboration, (3) to facilitate the transition to ICD-10 for mortality, and (4) to establish mechanisms for technical support of automated systems. The conference brought together researchers and vital statistics experts from nineteen countries to address the issues related to automation of mortality statistics.

The conference included five formal presentations that focused on descriptions of automation software, the advantages of available data output due to automation, improved data consistency, and international comparability. An additional eight papers were contributed. The core of the meeting centered around 20 facilitated small group discussions that concentrated on identifying issues and developing recommendations (see Chapter 11) related to the automation of mortality statistics. The background paper written by Harry Rosenberg, NCHS, (see Chapter 1) provided the framework for discussion and the focal topics.

The members of the planning committee for the ICE on Automating Mortality Statistics are from Statistics Canada: Gary Catlin, Janet Hagey, and Francois Nault; Institut National de la Sante et de la Recherche Medicale (INSERM), France: Gerard Pavillon; Office of National Statistics, Great Britain: Cleone Rooney; General Register Office of Scotland: Jack Arrundale and Graham Jackson; Generalitat de Catalunya, Spain: Gloria Perez-Albarracin; Statistics Sweden: Lars Age Johansson; and NCHS, United States: Mary Anne Freedman, Donna Glenn, Kenneth Kochanek, Francis Notzon, Kimberley Peters, Charles Rothwell, and Harry Rosenberg.

This volume contains the contributed and invited papers presented at the conference, meeting deliberations, and recommendations resulting from the group discussions. I am particularly pleased that a number of the recommendations of the first workshop have been adopted by the World Health Organization and that further progress is being made to confer the benefits of automated systems to health data. I am pleased to have this opportunity to thank and congratulate the participants for their contributions to improving the international statistical practices in the area of health.

Edward J. Sondik, Ph.D.
Director, National Center for Health Statistics

Acknowledgments

This conference was sponsored by the National Center for Health Statistics (NCHS). The United States Agency for International Development provided support for selected international travel. Coordination of international correspondence, hotel facilities, registration and meeting rooms, and reception arrangements was managed by Ginger Richards, NCHS. Arrangements for international travel were coordinated by Linda McCleary, NCHS.

Much of the credit for the success of the conference is due to Elizabeth Vasquez of Management Consulting Associates (Bethesda, Maryland) who guided the group and provided facilitator training, along with Glenn Pinder of NCHS, to NCHS staff members. Facilitators for the breakout sessions were: Robert Anderson, Linda Bean, Linda Biggar, Lisa Broitman, Susan Hawk, Ken Kochanek, Suzie Moberly, Kimberley Peters, Christine Plepys, and Michele Poulos.

Many thanks to the presenters and authors for their contributions to this volume. Individual comments should be addressed to them. Finally, thanks to all the participants, some traveling great distances, who were able to attend and share their questions, ideas, and experiences.

The editor for this volume was Kimberley Peters, and the publication manager was Gail Johnson.

Participants

Participants of the First International Collaborative Effort on Automating Mortality Statistics

Seated, left to right: Augusto Santo, Robert Israel, Iwao Moriyama, Mary Anne Freedman, Harry Rosenberg, Laura Garnica, Donna Hoyert, Pnina Zadke, Lois Fingerhut, Cleo Rooney, and Anna Iwanek. Second row: Li-ling Yu, Peter Jozan, Edna Roberts, Gloria Perez-Albarracin, Francoise Jean-Marie, Julia Raynor, Linda McCleary, Jack Arrundale, Mario Juarez Monroy, Barry Little, Susan Hawk, Patricia Wood, Michelle Poulos, Kimberely Peters, Brenda Green, and Lisa Broitman. Third row: Ginger Richards, Tanya Pitts, Charles Sirc, Joyce Bius,

International Collaborative Effort (ICE) on
Automating Mortality Statistics Meeting Participants
November 12-15, 1996

Mr. John Alexander
Manager, Health and Vitals
  National Project Centre
Australian Bureau of Statistics
GPO Box 9817
Brisbane, Queensland 4001
AUSTRALIA
Tel: 61-7-3222-6047
Fax: 61-7-3222-6038
E-Mail: jr.alexander@abs.gov.au

Dr. John Donovan (RETIREE)
Principal Medical Adviser
Australian Institute of Health & Welfare
GPO Box 570
Canberra ACT 2601
AUSTRALIA
Tel: 61-6-244-1103
Fax: 61-6-244-1111
E-Mail: john.donovan@aihw.gov.au

Mr. Celso Escobar Pinheiro
System Analyst
Ministério da Saude-FNS-Datasus
Rua Mena Barreto, 114-7th floor
Botafogo
22271-100 Rio De Janeiro-RJ
BRAZIL
Tel: 55-21-536-7117
Fax: 55-21-536-7147
E-Mail: celso@datasus.gov.br

Dr. Augusto Hasiak Santo
WHO Collaborating Centre for the
  Classification of Diseases in Portuguese
Faculdade de Saude Publica
Universidade de São Pãulo
Av. Dr Arnaldo 715
São Pãulo 01246-904
BRAZIL
Tel/Fax: 55-11-883-4246
E-Mail: auhsanto@usp.br

Ms. Françoise Jean-Marie
Production Manager
Statistics Canada
R.H. Coats Building, 18th Floor
Ottawa
K1A 0T6 Ontario
CANADA
Tel: 613-951-1648
Fax: 613-951-0792
E-Mail: jeanfra@statcan.ca

Ms. Patricia Wood
Mortality Classification Specialist
Health Statistics Division
Statistics Canada
R.H. Coats Building, 18th Floor
Ottawa
K1A 0T6 Ontario
CANADA
Tel: 613-951-1648
Fax: 613-951-0792
E-Mail: woodpat@statcan.ca

Mr. Tim Devis
Statistician, Population and Vital Statistics Division
Office for National Statistics (ONS)
Segensworth Road,
Titchfield
HANTS P015 5RR,
ENGLAND, UK
Tel: 44-132-981-3339
Fax: 44-132-981-3289
E-Mail: tim.devis@ons.gov.uk

Mr. Barry Little
Deputy, Data Services Division
Office for National Statistics (ONS)
Segensworth Road,
Titchfield
HANTS P015 5RR,
ENGLAND, UK
Tel: 44-132-981-3520
Fax: 44-132-981-3289
E-Mail: barry.little@ons.gov.uk

Dr. Cleone Rooney
Medical Statistician
Office for National Statistics (ONS)
Bessborough 6/ 12
1 Drummond Gate
London SW1V 2QQ,
ENGLAND, UK
Tel : 44-171-533 5254
Fax: 44-171-533-5252
E-Mail: cleo.rooney@ons.gov.uk

Mr. Eric Jougla
Service D'information Sur Les Causes Medicales De Deces
SC8-INSERM
44, chemin de Ronde
78110 Le Vésinet
FRANCE
Tel: 33-1-34-80-24-33
Fax: 33-1-34-80-24-29
E-Mail: jougla@vesinet.inserm.fr

Mr. Gérard Pavillon
Head, Centre Collaborateur OMS pour la Classification
Internationale des Maladies en Langue Française
SC8-INSERM
44, chemin de Ronde
78110 Le Vésinet
FRANCE
Tel: 33-1-34-80-24-62
Fax: 33-1-34-80-24-29
E-Mail: pavillon@vesinet.inserm.fr

Dr. Péter Józan
Head, Population and Health Statistics Department
Hungarian Central Statistical Office
5-7 Keleti Károly Street
H-1525
Budapest
HUNGARY
Tel: 36-1-345-6890
Fax: 36-1-345-6678
E-Mail: peter.jozan@ksh.x400gw.hu

Mr. Árpád Mészáros
Deputy Head, Population and Health Statistics Department
Hungarian Central Statistical Office
5-7 Keleti Károly Street
H-1525
Budapest
HUNGARY
Tel: 36-1-345-6950
Fax: 36-1-345-6678
E-Mail: arpad.meszaros/office@office.ksh.hu

Dr. János Weltner
Deputy Director
National Institute of Surgery
Üllöi ut 78
Budapest
H-1082
HUNGARY
Tel: 36-1-2100320
Fax: 36-1-2100321
E-Mail: wj@seb1.sote.hu

Mrs. Pnina Zadka
Head, Health Division
Central Bureau of Statistics
P.O. Box 13015
Hakirya-Romema
Jerusalem, 91130
ISRAEL
Tel: 972-2-6553370
Fax: 972-2-6553266
E-Mail: pnina@cbs.gov.il

Dr. Feola Giuseppe
Head of Classification and
Coding of Causes of Death Unit
National Statistical Institute (ISTAT)
Servizio SAN, Unit C
Viale Liegi 11
00198 Rome
ITALY
Tel: 39-6-8841341, ext. 7321
Fax: 39-6-85354401
E-Mail: feolag@pronet.it

Mr. Mustafa Mohammad
Head, Health & Vital Statistics Division
Ministry of Health, Kuwait
P.O. Box 5286
13053, Safat
KUWAIT
Tel: 965-243-7591
Fax: 965-240-2170/242-9723

Ms. Laura Garnica
Instituto Nacional de Estadistica
Geografia E Imformatica (INEGI)
Direccion De Estadisticas Dehograticas y Sociales
ACCESO 11, Segundo Nivel
Av. Heroe De Nacozari Sur 2301
Fracc. Jardines Del Parque
C.P. 20270, Aguascalientes. Ags.
MEXICO
Tel: 52-49-18-33-36
Fax: 52-49-18-24-18

Mr. Mario Juárez Monroy
Instituto Nacional de Estadistica
Geografia E Imformatica (INEGI)
Direccion De Estadisticas Dehograticas y Sociales
ACCESO 11, Segundo Nivel
Av. Heroe De Nacozari Sur 2301
Fracc. Jardines Del Parque
C.P. 20270, Aguascalientes. Ags.
MEXICO
Tel: 52-49-18-33-36
Fax: 52-49-18-24-18

Dr. Jan Kardaun
Project Manager Statistical Methods
Statistics Netherlands
428 Prinses Beatrixlaan
P.O. Box 4000
2270 JM Voorburg
THE NETHERLANDS
Tel: 31-70-377-4916
Fax: 31-70-377-5990
E-Mail: jkrn@cbs.nl

Ms. Anna Iwanek
Main Specialist
Demographic Statistics Division
Central Statistical Office of Poland
Al. Niepodleglosci 208.
00-925 Warsaw
POLAND
Tel: 48-22-608-32-07
Fax: 48-22-608-3181
E-Mail: aiwanek@gus.stsp.gov.pl

Dr. Vladimir V. Antonyuk
Senior Scientific Researcher
Department of Medical Demography
Research Public Health Institute (MedSocEconomInform)
11, Dobrolyubova str.
Moscow 127254
RUSSIA
Tel: 7-095-218-63-88
Fax: 7-095-219-38-40
E-Mail: 104633.1501@compuserve.com

Mr. Jack Arrundale
Statistician
General Register Office for Scotland
Ladywell House, Ladywell Road
Edinburgh EH12 7TF,
SCOTLAND, UK
Tel: 44-131-314-4229
Fax: 44-131-314-4344
E-Mail: gros@gtnet.gov.uk

Dr. Susan Cole
General Register Office for Scotland
Ladywell House, Ladywell Road
Edinburgh EH12 7TF,
SCOTLAND, UK
Tel: 44-131-667-3415
Fax: 44-131-314-4344
E-Mail: gros@gtnet.gov.uk

Dr. Glòria Pérez-Albarracín
Chief, Catalonian Mortality Register
Departament de Sanitat I Seguretat Social
Travessera de les Corts, 131-159
Pavelló Ave Maria
08028 - Barcelona
SPAIN
Tel: 34-3-227-2981
Fax: 34-3-227-2990
E-Mail: gperez@dsss.scs.es

Dr. Jaume Domènech
Technician, Catalonian Mortality Register
Departament de Sanitat I Seguretat Social
Travessera de les Corts, 131-159
Pavelló Ave Maria
08028 - Barcelona
SPAIN
Tel: 34-3-227-2900
Fax: 34-3-227-2990
E-Mail: gperez@dsss.scs.es

Dr. Núria Montellà
Technician, Catalonian Mortality Register
Departament de Sanitat I Seguretat Social
Pavelló Ave Maria
Travessera de les Corts, 131-159
08028 - Barcelona
SPAIN
Tel: 34-3-227-2900
Fax: 34-3-227-2990
E-Mail: gperez@dsss.scs.es

Mr. Lars Age Johansson
Senior Excutive Officer
Statistics Sweden
Statistical Programme for Health and Social Welfare
P.O. Box 24300
S-104 51 Stockholm
SWEDEN
Tel: 46-8-783-4498 or 46-8-82-63-14
Fax: 46-8-783-4652
E-Mail: lars.age@scb.se

Mr. Andre L'Hours
Technical Officer
Division of Health Situation and Trend Assessment
World Health Organization
Av. Appia
1211 . Geneva . 27
SWITZERLAND
Tel: 41-22-791-2843
Fax: 41-22-791-4194
E-Mail: volkans@who.ch

Ms. Li-ling Yu
Recommended Specialist
Office of Statistics
Department of Health
The Executive Yuan
10F, 100, Al Kuo E. Rd.
Taipei
TAIWAN R.O.C.
Tel: 886-2-3922643
Fax: 886-2-3224279
E-Mail: stliliyu@dohr6.doh.gov.tw

Dr. Tsung-Hsueh Lu
Department of Public Health and Family Medicine
Chung Shan Medical and Dental College Hospital
23, Section 1, Taichung Kang Rd.
40334 Taichung
TAIWAN R.O.C.
Tel: 886-4-261-4690
Fax: 886-4-262-1014
E-Mail: robertlu@ms1.hinet.net

Dr. Iwao Moriyama
President
International Institute for Vital Registration
    and Statistics (IIVRS)
9650 Rockville Pike
Bethesda, MD 20814
UNITED STATES
Tel: 301-530-7131
Fax: 301-571-1855

Ms. Elizabeth Vasquez
President
Management Consulting Associates
5208 Marlyn Drive
Bethesda, MD 208l6
UNITED STATES
Tel: 301- 229-1655
Fax: 301-229-0473

Mr. Alfred G. Zangri
Director, Alaska Center for Health Statistics
    and Registrar of Vital Statistics
National Association for Public Health Statistics
    and Information Systems
State of Alaska, Bureau of Vital Statistics
P.O. Box 110675
Juneau, Alaska 99811-0675
UNITED STATES
Tel: 907-465-8606
Fax: 907-465-3618
E-Mail: azangri@health.state.ak.us

Dr. Brenda K. Edwards
Associate Director, Cancer Control Research Program
National Cancer Institute
National Institutes of Health
6130 Executive Blvd., EPN Room 343
Bethesda, MD 20892-7350
UNITED STATES
Tel: 301-496-8506
Fax: 301-496-9949
E-Mail: edwardsb@dcpcepn.nci.nih.gov

Dr. Roberto Becker
Regional Advisor on Classification of Diseases
Health Situation Analysis Program
Health and Development Division
Pan American Health Organization
525 23rd Street, NW
Washington, DC 20037
UNITED STATES
Tel: 202-974-3131
Fax: 202-974-3674
E-Mail: beckerro@paho.org

Dr. Carlos Castillo-Salgado
Program Coordinator,
Health Situation Analysis Program
Health and Development Division
Pan American Health Organization
525 23rd Street, NW
Washington, DC 20037
UNITED STATES
Tel: 202-974-3327
Fax: 202-861-5230
E-Mail: castillc@paho.org

Ms. Edna Roberts
Statistician
Health Situation Analysis Program
Health and Development Division
Pan American Health Organization
525 23rd Street, NW
Washington, DC 20037
UNITED STATES
Tel: 202-974-3137
Fax: 202-861-5230
E-Mail: robertse@paho.org

Dr. Robert Hartford
18 Ranch Road
Cedar Crest, NM 87008
UNITED STATES
Tel/Fax: 505-286-2200
E-Mail: hartfordpr@aol.com

Mr. Robert Israel
12411 Sandal Lane
Bowie, MD 20715
UNITED STATES
Tel: 301-262-5083
E-Mail: 73125.627@compuserve.com

Dr. Robert Anderson
Statistician
Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8884, ext. 179
Fax: 301-436-7066
E-Mail: rca7@cdc.gov

Ms. Linda Bean
Chief
Data Dissemination Branch
Data Dissemination Services
National Center for Health Statistics
6525 Belcrest Road
Room 1064, Presidential Building
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-6154, ext. 179
Fax: 301-436-3797
E-Mail: llb3@cdc.gov

Ms. Linda Elias Biggar
Computer Specialist
Systems Programming and Statistical Resource Branch
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8900, ext. 187
Fax: 301-463-7066
E-Mail: lle1@cdc.gov

Ms. Joyce L. Bius
Medical Classification Specialist Assistant
Data Preparation Branch
Division of Data Processing
National Center for Health Statistics
12 Davis Drive, P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-3119
Fax: 919-541-4098
E-Mail: jlb9@cdc.gov

Ms. Lisa Broitman
Program Analyst
Office of Planning, Budget and Legislation
National Center for Health Statistics
6525 Belcrest Road
Room 1120, Presidential Building
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7142, ext. 179
Fax: 301-436-3705
E-Mail: lcm9@cdc.gov

Ms. Lois Fingerhut
Special Assistant for Injury Epidemiology
Office of Analysis, Epidemiology, and Health Promotion
National Center for Health Statistics
Room 750, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7032, ext. 111
Fax: 301-436-8459
E-Mail: laf4@cdc.gov

Ms. Mary Anne Freedman
Director, Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8951, ext. 112
Fax: 301-436-7066
E-Mail: maf7@cdc.gov

Ms. Donna E. Glenn
Survey Statistician
Technical Services Branch
Division of Vital Statistics
National Center for Health Statistics
12 Davis Drive, P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-0999
Fax: 919-541-4098
E-Mail: dgp2@cdc.gov

Ms. Brenda A. Green
Survey Statistician
Technical Services Branch
Division of Vital Statistics
National Center for Health Statistics
12 Davis Drive, P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-0987
Fax: 919-541-4098
E-Mail: bxj9@cdc.gov

Ms. Marjorie Greenberg
Program Analysis Officer
Office of Planning and Extramural Programs
National Center for Health Statistics
Room 1100, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7l42, ext. 133
Fax: 301-436-4233
E-Mail: msg1@cdc.gov

Mr. James Hart
Experts Systems Programming Manager
Data Preparation Branch
Division of Data Processing
National Center for Health Statistics
12 Davis Drive, P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-0101
Fax: 919-541-4098
E-Mail: bzh2@cdc.gov

Ms. Susan Hawk
Program Analyst
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Bluilding
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7464, ext. 159
Fax: 301-436-7066
E-Mail: sah4@cdc.gov

Dr. Donna Hoyert
Statistician
Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8884, ext. 168
Fax: 301-436-7066
E-Mail: dlh7@cdc.gov

Mr. Kenneth Kochanek
Statistician
Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8884, ext. 172
Fax: 301-436-7066
E-Mail: kdk2@cdc.gov

Ms. Linda S. McCleary
Staff Assistant
Office of International Statistics
National Center for Health Statistics
Room 701, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7039, ext. 169
Fax: 301-436-3568
E-Mail: lsm3@cdc.gov

Ms. Suzie Moberly
Human Resources Specialist
Office of Management
National Center for Health Statistics
6525 Belcrest Road
Room 1162, Presidential Building
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-6045, ext. 141
Fax: 301-436-6668
E-Mail: sbmo@cdc.gov

Dr. Francis C. (Sam) Notzon
Statistician
Office of International Statistics
National Center for Health Statistics
Room 701, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7039, ext. 171
Fax: 301-436-3568
E-Mail: fcn2@cdc.gov

Ms. Kimberley Peters
Statistician
Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8884, ext. 180
Fax: 301-436-7066
E-Mail: kcp6@cdc.gov

Ms. Donna Pickett
Medical Classification Administrator
Office of Planning and Extramural Programs
National Center for Health Statistics
Room 1100, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7050, ext. 142
Fax: 301-436-4233
E-Mail: dfp4@cdc.gov

Mr. Glenn D. Pinder
Special Assistant to the Director
Division of Health Examination Statistics
National Center for Health Statistics
Room 1000, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7068 ext. 205
Fax: 301-436-5431
E-mail: gdp1@cdc.gov

Ms. Tanya W. Pitts
Medical Classification Specialist
Data Preparation Branch
Division of Data Processing
National Center for Health Statistics
12 Davis Drive, P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-2728
Fax: 919-541-4098
E-Mail: twp1@cdc.gov

Ms. Christine Plepys
Statistician
Data Monitoring and Analysis Branch
Division of Health Promotion Statistics
National Center for Health Statistics
6525 Belcrest Road
Room 770, Presidential Building
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-3548, ext. 122
Fax: 301-436-5572
E-Mail: cmp0@cdc.gov

Ms. Michele Poulos
Project Manager
Office of the Associate Director
Office of Data Processing and Services
National Center for Health Statistics
6525 Belcrest Road
Room 1050, Presidential Building
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7135, ext. 164
FAX: 301-436-3797
E-Mail: rmp3@cdc.gov

Ms. Julia E. Raynor
Medical Classification Specialist
Data Preparation Branch
Division of Data Processing
National Center for Health Statistics
12 Davis Drive
P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-4408
Fax: 919-541-4098
E-Mail: jer3@cdc.gov

Ms. Ginger Richards
International Program Coordinator
Office of International Statistics
National Center for Health Statistics
Room 701, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7039, ext. 172
Fax: 301-436-3568
E-Mail: var2@cdc.gov

Dr. Harry M. Rosenberg
Chief, Mortality Statistics Branch
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8884, ext. 175
Fax: 301-436-7066
E-Mail: hmr1@cdc.gov

Mr. Charles Rothwell
Associate Director
Office of Data Processing and Services
National Center for Health Statistics
Room 1050, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7135, ext. 160
Fax: 301-436-3797
E-Mail: cjr4@cdc.gov

Mr. Charles E. Sirc
Acting Deputy Director
Division of Data Processing
National Center for Health Statistics
12 Davis Drive, P.O. Box 12214
Research Triangle Park, NC 27709
UNITED STATES
Tel: 919-541-0134
FAX: 919-541-4098
E-Mail: ces0@cdc.gov

Dr. Edward J. Sondik
Director
National Center for Health Statistics
Room 1140, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-7016
Fax: 301-436-5202
E-Mail: efs2@cdc.gov

Dr. James Weed
Deputy Director
Division of Vital Statistics
National Center for Health Statistics
Room 820, Presidential Building
6525 Belcrest Road
Hyattsville, MD 20782
UNITED STATES
Tel: 301-436-8951, ext. 112
Fax: 301-436-7066
E-Mail: jaw4@cdc.gov

Contents

International Collaborative Effort on Automating Mortality Statistics: Background and Issues

Harry M. Rosenberg, Ph.D., National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services

As in so many other aspects of modern life, computer technology is being rapidly and effectively applied to many aspects of vital statistics, including data collection, production, and dissemination. The benefits are demonstrably great in terms of improved data quality, improved timeliness, and reduced staffing resources. If technology and systems designs in automating vital statistics can be effectively shared at the international level, they have the promise of conferring the benefits of not only better and more timely data but also improved international data comparability. Thus, the coordinated application of computer technology has the potential of becoming a great facilitator, like the International Classification of Disease almost a century ago, in our common purpose to produce better statistical information in order to promote public health, prevent disease, and enhance scientific knowledge. Recognizing the potential of automation in vital statistics, the first "International Collaborative Effort on Automating Mortality Statistics" (ICE) was held in Washington, D.C., on November 12-15, 1996, as proposed at the WHO Center Heads meeting in Canberra, Australia, in 1995. The National Center for Health Statistics (NCHS) has previously sponsored ICEs in the areas of perinatal mortality, aging statistics, and injury statistics.

Purpose

The purposes of the ICE on automating mortality statistics were as follows: 1) to share knowledge and experience on automated systems for coding mortality information, 2) to develop and improve existing automated systems through collaboration, 3) to facilitate the transition to ICD-10 for mortality, and 4) to establish mechanisms for technical support of automated systems. The format of the initial meeting was largely one of discussions focusing on specific areas related to the application of automation to mortality coding, with the goal of developing practical recommendations related to automation. In addition, several formal presentations were made as follows: 1) a review of the results of a questionnaire about present coding practices and future plans regarding automation of ICE countries, 2) a description of the U.S. automated system, 3) a presentation on the advantages of using multiple cause-of-death statistics, which can be routinely produced by some automated systems, and 4) a presentation of the effect on data trends of implementing an automated system in one country. In addition, some countries also shared their experience in using automated systems. Contributed papers that highlight these experiences are included in this proceedings.

Planning Committee

The ICE was planned in January 1996 by a committee comprised of international users of automated mortality processing software including England, Canada, France, Sweden, Scotland, and the United States. Invitations to the ICE meeting were sent to selected countries known to have an interest

in automation for mortality data processing, as well as to the Heads of all the WHO Collaborating Centers, and to selected organizations including the World Health Organization, the World Bank, the American Medical Association, the National Association for Health Statistics and Information Systems, the U.S. National Cancer Institute, the U.S. National Heart, Lung, and Blood Institute, the International Institute for Vital Registration, and EUROSTAT. Support for some international travel was provided by the U.S. Agency for International Development and NCHS.

The ICE Planning Committee felt that the ICE is an important process by which the benefits of automating vital statistics can be widely shared, and thereby improved and strengthened, with an initial emphasis on developed countries with experience in automation. Participation by developing countries is envisioned for future meetings. It was further noted that the scope of automating vital statistics can go beyond coding to include data collection and data dissemination. Thus, in the area of data collection, the registration of births in the United States is being automated. For about 70 percent of the births in the United States, information is being collected electronically in hospitals and then transmitted electronically to the State. Parallel initiatives for mortality are likely to occur in the next few years. In the area of data dissemination, vital statistics are now being made available on CD-ROMs and are being put on the Internet. Future ICEs can include developments in electronic data collection and data dissemination.

Background

The stimulus for the ICE is the growing international interest in automation for processing mortality statistics, an interest motivated by a number of factors including: 1) the general success and continuing improvements in existing systems, 2) the loss without replacement of high level mortality coding specialists (nosologists) throughout the world, 3) the growing interest in containing costs in the public sector, 4) the continuous improvements and concurrent cost reductions in automation hardware and software, and 5) the growing use of automated systems in many countries including Australia, Brazil, Canada, England, Italy, Japan, Scotland, Spain (Catalonia), Sweden, and the United States, and the expressed interest of many other countries. By way of background, it may be instructive to describe the U.S. system as one model of automated coding of mortality medical data.

The U.S. Automated System

The United States has a decentralized vital statistics system in which the vital records functions are carried out under State, not Federal, law and in which coherence and coordination of the system derive from a cooperative relationship between NCHS and the States. The relationship is embedded in contractual agreements under which States deliver to NCHS vital statistics within specified standards and time schedules agreed to by all parties. In exchange, NCHS shares the State costs for producing and delivering vital statistics data. NCHS also provides data specifications and cause-of-death coding software to the States and conducts quality control on data received from the States.

The development of the U.S. automated software took place at NCHS over a period of three decades during which the software has been refined and expanded. Currently, the U.S. system includes four inter-related components known by acronyms ACME, TRANSAX, MICAR, and Super-MICAR. The original component of the U.S. system—ACME—was first used with the 1968 mortality data. ACME was designed with three goals: 1) to use software that embodies a set of explicit rules and relationships and could be modified and updated to select the underlying cause of death more consistently than could be done by manual coders, 2) to simplify data entry and thereby reduce the cost of training medical coders and nosologists, and 3) to produce multiple as well as underlying cause-of-death data. The ACME system did meet two of the three goals. The unmet goal was simplifying data entry and thereby reducing costs. In fact, instructions for ACME were at least as complex as underlying cause coding. Nevertheless, the benefits of consistently selecting underlying cause data and, for the first time, routine multiple cause-of-death data were considered to fully justify implementing the new system.

ACME

The ACME system applies the same rules for selecting underlying cause as would be applied manually by a nosologist, but the automated system with its explicit rules eliminates the intercoder variation in selecting the underlying cause that characterizes manual coding even with sample verification. All the medical conditions listed on the death certificate are coded using detailed instructions for data entry into the ACME system. The codes are then matched automatically against decision tables, which provide the comprehensive relationships among the conditions classified by the ICD when applying the rules for selection and modification of the underlying cause of death (1). The decision tables were developed by NCHS on the basis of experience in manually coding underlying cause of death and as a result of periodic independent validations. The tables are periodically updated to reflect new information on relationships among medical conditions and to accommodate new conditions such as HIV infection, which was introduced as a new category for deaths in 1987.

With the U.S.' decentralized vital statistics system, most of the States gradually adopted ACME and provided multiple cause and underlying cause codes to NCHS in electronic form. Those States that did not use ACME sent to NCHS copies of death certificates that were then processed by NCHS using ACME. Currently, all U.S. death certificates are processed through automated coding software to provide data on underlying and multiple-cause mortality. Coding procedures for data entry in the ACME system are documented in NCHS instruction manuals that are made available to all the States and to other countries and organizations with an interest in automated mortality medical coding (2–6).

TRANSAX

One goal in implementing automation in the United States was the routine production of multiple cause-of-death statistics, which, prior to automating mortality processing, had been produced only intermittently (7). While mortality statistics are normally tabulated and analyzed in terms of the underlying cause of death, often much additional valuable diagnostic information is provided by the medical certifier regarding other conditions usually reported on the death certificate as contributing to death. Together, the underlying and contributing (non-underlying causes) are referred to as

“multiple causes of death.” As a complement to ACME, in the late 1970's, a multiple cause-of-death system was developed called “TRANSAX,” (for translation of axes) (8). The TRANSAX program uses the same inputs as the ACME program, and then generates two types of outputs amenable to mortality analysis—“entity axis” data and “record axis” data. The former, which preserves diagnostic detail and placement on the death record (line and position), is particularly useful for analyzing medical certification practices; and the latter, which reflects relationships among all the reported conditions on the records and removal of contradictions and redundancies, is more suitable for mortality analysis. The TRANSAX program allows for preserving up to 20 entity axis codes and 20 record axis codes for each death.

MICAR

While one of the initial objectives of developing automated systems was not met—namely, that of reducing the costs of data entry—it was not lost sight of. Work continued to develop a system that would simplify and ultimately automate the multiple cause data entry to the ACME program. In 1990, another major milestone was reached with the implementation of MICAR, or the “Mortality Medical Indexing, Classification, and Retrieval” system (1). MICAR, which to a large extent automates multiple cause coding rules, allows data entry in the form of “sanitized” text descriptions or entity reference numbers (ERNs).

A sanitized text description is one that has been entered, not as reported on the death certificate, but as it is expected to appear in the MICAR dictionary. Thus, the certificate may report “Cancer of the lung,” but the dictionary will accept only “Lung Cancer.” Therefore, the data entry operator must enter the sanitized version, Lung Cancer, if he/she hopes to get a match.

An ERN is simply a six-digit numeric code for a cause-of-death entry. The ERNs are assigned by NCHS sequentially, without reference to any coding scheme. They contain much more detail than the ICD. Thus, the ERN for acute myocardial infarction is 000001. Nothing else in the MICAR dictionary has that ERN. Thus, ICD-9 No. 410 includes a large number of ERNs—all of which refer to heart conditions subsumed by, or synonymous with, acute myocardial infarction. The ERNs contain greater detail, which is necessary when relating terms using MICAR.

The use of the sanitized text entries and the ERNs can eliminate use of the ICD index, can reduce errors in recognizing terms, and can eliminate manual use of multiple cause coding rules, some of which are complex and rarely used. MICAR can also provide more detailed information on the conditions reported on death certificates, since terms are identified by reference numbers unique to each medical term rather than by the broader categories of the ICD, which often subsume a large number of synonymous, similar, or related medical terms (9).

Super-MICAR

In 1993 an enhancement of MICAR called “Super-MICAR” was implemented. Unlike the original MICAR system that required the coder to know or be able to look up the sanitized text or the ERN, Super-MICAR allows for total literal

entry of the multiple cause-of-death text as reported by the medical certifier. This serves as input to MICAR, which in turn is input into ACME. The use of MICAR and Super-MICAR has grown since 1990, when only about 5 percent of the U.S. death records were processed using MICAR. In the 1993 data year, MICAR could process about 88 percent of the records, while Super-MICAR could process about 70 percent. MICAR and Super-MICAR systems do not yet have the capability of processing all death certificates. Records that can not be processed with either system are manually multiple cause coded and then processed using ACME (1).

Quality control for data entry is maintained by having NCHS code a sample of 70 to 80 records per month for each State. Each sample record is independently coded by NCHS staff and compared with the State code assignments. Differences in code assignment are adjudicated to ascertain the source of the error and the need for corrective action.

With respect to ICD-10, the United States hopes to implement the new Revision effective with the 1999 data year. Planning for ICD-10 has had major ramifications for NCHS automated systems, which have to be converted to a classification system that is far more detailed than previous revisions of the ICD. The new automated systems are designed to run in a desktop computer (pc) environment rather than on the mainframe. The new system will be a WINDOWS application. The systems are programmed in the C language, with all the documentation on diskette. System requirements for the NCHS automated systems are shown in table 1.

Table 1. Minimum Configurations for the ICD-10 MICAR Software Systems Super-MICAR Processing, MICAR 100/200 PC-ACME/TRANSAX

Option	Minimum Required	Suggested
CPU	486/50	Pentium 75 or higher
RAM	4 Megabytes	16 Megabytes
HDD	340 Megabytes	850 Megabytes
Monitor	VGA	VGA
Operating System	MS Windows 3.1	MS Windows 3.1
Other	3.5" floppy, mouse	3.5" floppy, mouse

Note: These systems will also run under OS/2 2.1 or greater in a Win OS/2 session. They can also be run on the PowerMac but are not supported on this platform.

Issues Associated with Automating Cause-of-Death Coding

Implementing automation in cause-of-death processing has many implications at the national level, and even more at the international level. Among the issues identified by the ICE Planning Committee are the following: (a) nosology and the training of nosologists, (b) training of automation (pc) support managers and personnel, mechanisms for technical support, and training

users, (c) systems specifications, decision tables and mechanisms for updating them, and quality control, (d) bridge coding, (e) data editing and querying, (f) external causes of death, (g) language issues, and (h) implementation issues. Each of these issues is discussed below.

Nosology and the training of nosologists

Automation is changing the activities of nosologists and mortality medical coders who heretofore have been the foundation of producing cause-of-death statistics. Nosologists are mortality medical coders who have achieved high levels of expertise in the practice of medical coding; in the interpretation and application of the ICD rules; in the training, apprenticeship, and qualification of new medical coders; and in the implementation of special projects on cause of death. If current automated systems, whose "throughput" is being continuously improved, now are capable of processing at least 95 percent of all death certificates in the United States, what will be the future need for nosologists and medical coders? How many coders are required to code the remaining records that currently are "rejected" by the automated system for manual processing? These records, to be sure, are those that are more difficult than the ones amenable to automated processing, but the numbers of rejected records will surely diminish as automation is improved.

The advent of automation raises major questions about the future role of nosologists who have been central to keying mortality data, to developing cause-of-death coding specifications, to interpreting mortality data, and to developing the coding aspects of the ICD. The ICE Planning Committee expressed concern that reject processing alone will not keep nosologists' overall skills current enough to be able to modify decision tables, to interact with statisticians on research issues, and to effectively participate in processes that require an in-depth understanding of the ICD classification system and the rules for assigning underlying cause of death. Thus, automation has the potential to create systems that cannot be updated because the expertise to modify coding rules will be lost. Among the questions addressed by the ICE regarding nosologists were:

- How should skill levels for medical coding be maintained? Should statisticians be trained to have high level nosological skills? In the Planning Committee, considerable skepticism was expressed that statisticians and epidemiologists could effectively carry out nosological tasks because of a fundamental difference in perspective between these professions. For nosologists, strict adherence to specific rules is fundamental; for statisticians and epidemiologists a broader perspective and more flexibility is applied to the same problems. Thus, it is likely that statisticians and epidemiologists would change decision tables or code rejects in a much different way than nosologists. In developing and updating decision tables, the complementary skills and perspectives of both nosologists and epidemiologists are needed.

- There is need to determine the minimum number of coders needed to keep skill level up. Should there be a minimum of at least two people on staff who have these skills?

- Can job sharing be a possible solution? Can there be part-time mortality medical coding? Can nosologists' jobs be diversified? Could or should nosologists be moved into special projects to broaden their outlook and experiences, in such areas as mortality followback studies or reference studies? Could their job responsibilities be diversified? Job sharing has been implemented in several countries, where nosologists engage in other activities essential to statistical operations.

- The skills intrinsic to nosology are highly specialized, and in themselves offer limited opportunity for job advancement in a nonautomated environment, which even now makes recruitment and retention a problem. In an automated environment what will be the possibilities for career advancement of nosologists? What will be their career structure?

- In an automated environment, how can nosologists' skills be maintained? Should they code a sample of Super-MICAR records to keep up skills? It is believed that this would be difficult to justify.

- Is one solution to have a national or an international focal point for this expertise? It has been suggested that there could be an international quality control sample, but it is recognized that language problems would make this difficult. Can there be "virtual" assistance in the area of nosology from the United States, from other countries, from an international source? Is there danger in overly centralizing expertise, and, safety in having multiple sources?

- Should there be standards for coders? Should there be recertification? What qualifications are needed for mortality medical coders? For nosologists? It has been suggested that an international curriculum for training coders, and international certification, may be helpful to improve both international comparability of mortality statistics and the status of mortality medical coders and nosologists (10). Such certification might help coders, the complexity of whose jobs is often not fully appreciated, achieve the status of subject matter experts, possibly resulting in better pay.

Training of automation (pc) support, mechanisms for technical support, and training users

An automated environment requires a staff complement that differs substantially from one in which death certificates are manually coded. The shift is a structural one in which coding production skills of several staff members are, of necessity, replaced by programming and computer skills of fewer staff, but, as noted above, the retention of some high level nosological skills remains essential.

- Specifically, what types of management and computer skills are needed in this new environment?

- The ICE Planning Committee noted that more people need to be trained in all aspects of automated systems support, including the need for systems managers and automation (pc) managers. How many are required? What types of skills do they need?

- Once a country has adopted automated systems, what types of technical support are required to maintain the systems and to handle problems of software and hardware that can have a deleterious effect on production?

- How should the users of these new systems be trained? Should there be a central training site? How could it be organized and funded?

- Automated systems require training of data users, that is, of statisticians and analysts in a number of areas, including medical coding and multiple causes of death. Since medical coders will be far fewer in number in an automated environment, what types of courses should be developed so that statisticians and analysts can acquire a greater knowledge in mortality medical classification and coding? This training may need to include an understanding of the selection and modification rules, medical terminology, physiology and anatomy—topics that have traditionally been part of nosology training, not statistical training. In addition, since one of the important statistical byproducts of automated systems is multiple cause data, statisticians and analysts should familiarize themselves with the logic of TRANSAX, and with the analytical potential of multiple cause-of-death statistics.

Systems specifications, decision tables, mechanisms for updating decision tables, and quality control

Statistical operations including data entry, data processing, and data tabulation have always required explicit specifications in the form of written documentation. This is essential to training, trouble-shooting for problems, and for analysis and interpretation of data (inasmuch as changes in data handling can result in changes in trends and patterns). In an automated environment, clear and concise systems specifications are needed.

- What types of systems specifications are needed? Who maintains and updates the systems specifications? How is this information disseminated to users of the systems? In the automated systems developed by NCHS, the amount of documentation is copious, since the software is disseminated to all the States where the bulk of the data processing occurs. Nevertheless, it has been pointed out that the source code will also be needed by other countries wanting to use software developed in the United States or in other countries because of differences in platforms or interface requirements (10).

Central to automated processing of mortality statistics are the decision tables that embody the selection and modification coding rules of the ICD and that show explicitly the acceptable relationships between any two diagnostic entities. In an automated system used by many countries it is essential that the specifications for the decision tables be clear and concise, that they be standard, and that they be independent of the type of platform being used. The same is true of the edit specifications that ensure, for example, consistency between age and cause of death, sex and cause of death, and the proper range of entered data. Explicit specifications will allow non-English speaking countries and countries that are in disagreement over specific cause-of-death selection logic to make changes only where necessary, so that the resultant systems will be as comparable as possible. And, where differences are

preferred over standardization, written specifications ensure that those differences are identifiable and documented. For differences among countries, it may be advantageous to establish a forum for international quality control.

- Under what circumstances will changes in decision tables be made? Who will make the authoritative comparison between the decision tables and the ICD rules?

- Modifications in decision tables and edit specifications by a country should be communicated to other countries, as it may have an impact on international comparability. How can this best be done? Should there be a central clearinghouse or bulletin board or bulletin or an e-mail list?

- How will updates, new versions, and modifications to the systems be handled, monitored, and documented? If the systems are internationalized, there is a distinct need to know what version of the system is being used by which countries, or "version control." (In the United States, written specifications are readily available in printed form for receipt and control, for data entry, for ACME, MICAR, TRANSAX, for computer editing, and for querying; most of these are updated and published annually in NCHS Instruction Manuals that are sent to the States. Copies of the manuals are available upon request, and could be replicated in other languages.)

Provisions need to be made for quality control for data entry and systems design. Most countries have standard procedures for assessing and maintaining quality at the data entry level. However, procedures need to be implemented nationally to ensure consistency between manual and automated coding. An automated system should do the job an excellent coder would do.

Bridge coding

Bridge coding means coding a set of records using alternative coding schemes. Traditionally, bridge coding has been used to assess the quantitative impact of moving from one revision of the ICD to another. (In the United States, bridge coding studies are called "comparability studies.") The implementation of a new revision of the ICD can result in substantial changes in the number of deaths attributable to a disease category such as Ischemic heart disease, because of changes in classification or changes in the rules for selecting the underlying cause of death (or the scope of the rules). Thus, prior to the implementation of a new revision of the ICD many countries take a large sample of death records that have been coded by the prevailing system (say the Ninth Revision) and then code the same sample of records by the new system (the Tenth Revision). The double-coded records are used to compare the statistical impact of the new revision. The effect of coding changes between revisions of the ICD (e.g., the introduction of HIV infection as a category) can also be assessed using bridge coding to determine if they introduce systematic changes in trends and patterns of mortality.

Accordingly, bridge coding has a role to play in converting from manual to automated systems. Changes from manual to automated systems, especially automated systems developed in other countries, can have a major impact on trends. In England, adoption of the ACME system resulted in discontinuities in mortality trends from a number of causes of death, because the ACME system incorporates a different interpretation of the international rules than the

English manual system in use earlier. Automation, then, raises questions with regard to bridge coding such as:

! Should there be any standards for the conduct of bridge coding studies? How should the results of such studies be disseminated?

Data editing and querying

Data edits ensure that codes entered manually or automatically are acceptable for further processing. These include, for example, range edits. Other edits ensure that consistency is achieved between the assigned cause-of-death code and the sex or the age of a decedent. Thus, deaths from prostate cancer should occur only to males, and a death from Alzheimer's disease should be questioned if reported as occurring to a child. Such edits to ensure consistency among variables are routine in any coding system. These edits may be conditional or absolute. A conditional edit implies that medical certification is dubious, and should be examined closely; whereas an absolute edit indicates that the demographic and medical information is completely incompatible. In the latter case, procedures must be available to reconcile the incompatibility by either going back to the source of the information, or by arbitrarily changing the coded information to produce a consistent combination of medical and demographic information. (In the United States, data edits are embodied in vital statistics instruction manuals that are sent annually to the States. The edits are explicitly embodied in the automated coding systems.)

In an international context, edit information must be made widely, explicitly, and concisely available. Updates need to be routinely shared, and the consequences of major changes fully assessed on a statistical basis using bridge coding where deemed appropriate. It will be important for quality control to measure discontinuities when changes are made in edits, for example, in the criteria for rejecting death certificates as in maternal deaths, or sudden deaths to persons under the age of 60 years, or certain infectious diseases.

Querying is the act of questioning the physician where the medical certification of death is ambiguous, incomplete, or questionable in some other way. In both manual and automated systems, it is important to ensure that information on the death certificate is not only codable, but is sufficiently detailed to make it useful for public health and medical research. While a death reported as due to cancer can be assigned an ICD code, such a certification must be considered incomplete unless additional information is provided as to the nature and primary site of the cancer. Obtaining such additional information, called "querying," is integral to vital statistics systems not only for quality control but also to promote good certification practices by physicians and to communicate to the certifying physicians that their certifications are being used for statistical purposes and for research. Querying can be partially automated by printing letters to physicians in those instances where queryable deaths are encountered. (Querying guidelines are spelled out explicitly in an NCHS vital statistics instruction manual. In the United States a minimum level of demographic and medical querying is required of all States. Some years ago NCHS randomly queried physicians about their medical certification practices, but discontinued the practice because of costs).

External causes of death

External causes are causes that are not the result of disease processes. They include homicides, suicides, and accidents, as well as deaths where it could not be determined whether accidentally or purposely inflicted. The medical certification of such deaths usually also indicates the physiological consequences of such deaths such as fractures, puncture wounds, lacerations, etc. Included among these deaths are those that occur in connection with health care such as surgeries. External causes have always presented challenges to medical coders, because sometimes the medical certification of death is ambiguous with respect to the circumstances of the death. For these deaths, information from Part I of the medical certification is sometimes insufficient to assign a definitive code so the medical coder has to turn to other information on the death certificate such as a narrative on "how the injury occurred" (or in the United States the checkbox item on "manner of death") to assist in code assignment.

As external causes have presented challenges for manual coding, similarly, they have posed particular problems in the United States in developing MICAR and Super-MICAR. Johansson has pointed out that external causes are much less amenable to automated processing than natural causes (10): natural causes of death lend themselves to automated processing, because scientific medical terminology consists of a comparatively limited set of basic words and phrases, and therefore can be matched against a dictionary of standard medical words and phrases. In contrast, external causes—including accidents, homicides, suicides, and other forms of trauma—are often described in ordinary, nonscientific language that is not readily amenable to direct matching.

The already difficult problems of automating external causes have been exacerbated by the introduction of the Tenth Revision with its requirements for far greater detail in coding and classification than previous revisions. While automated coding may help improve international comparability of mortality from external causes, problems of comparability may remain because of international variations in the medical-legal context in which the external cause is reported on the death certificate. In some countries, for example, the results of a coroner's inquest and a verdict determine how these deaths are medically certified, while in other countries, the statement of a medical-legal officer (such as a coroner or medical examiner in the United States) is sufficient basis for the cause-of-death report.

To date, external causes remain the least amenable to fully automated coding. In the United States, external causes account for a disproportionate share of the records rejected by MICAR and Super-MICAR for manual coding. Eventually, these causes of death will also be largely processed by automated systems, but capabilities to handle rejects must be planned for.

Language issues

Language issues bear on moving from manual to automated systems for processing cause-of-death information, because some diagnostic terms may have different denotations and connotations in different languages. The relevance of language considerations is greater for MICAR and Super-MICAR than for ACME and TRANSAX. MICAR and Super-MICAR have dictionaries that are sensitive to the language context. In contrast, edits and decision tables are couched entirely

in terms of ICD codes and therefore are less likely to be sensitive to language context. Thus, there is the possibility of entering the automated systems at different points in order to make the system compatible with national needs. (England has developed its own front end system called "Tracer," and its own data base file.)

While some language issues are more or less obvious *a priori*, others will arise out of the experience of using the United States and other automated systems in different national contexts. The use of automated systems may uncover national variations in coding practices, in diagnostic terminology, in editing, and in reporting practices that were not heretofore known widely or known at all. Among the language issues raised by the ICE Planning Committee are the following:

- How will special national terminology be handled?
- Who does national code assignments?
- What should be the role of the WHO Collaborating Centers?
- How can we coordinate and add expressions that are not currently found in the dictionary of the system?
- How will the need for keeping the classification system current be handled? There will also be opportunities for sharing translated versions; but there is a need for coordination among countries with similar languages to develop the language-specific front-end software. Questions arise with regard to what portion of Super-MICAR will be appropriate for countries to use, and what portion will need to be changed for the French, Swedish, etc. versions.

Automated coding, with a high degree of international standards and coordination, can greatly reduce problems of comparability in mortality statistics among countries. Residual and possibly very important differences may remain due to variations in diagnostic technology and terminology, in death certificates, in querying, and in the legal context of vital records completion (particularly for external causes).

Implementation issues

One of the main reasons for convening the first International Collaborative Effort on Automating Mortality Statistics is finding ways to better solve problems with respect to mortality coding in the real world. It is fervently hoped that these discussions are more than a paper and pencil exercise, or an isolated intellectual dialogue. The key to the success of the ICE will be the way in which ideas are translated into meaningful actions, in a word, identifying implementation issues that are amenable to practical solutions.

Most immediately, institutional mechanisms need to be identified to provide technical assistance in installing, maintaining, enhancing, and updating automated systems in an international context. The Planning Committee identified some questions that need to be addressed with respect to implementation, including the following:

- What options are there to sharing the burden of internationalizing the automated systems?
- Who will take the lead in mobilizing them?
- What resources are required?
- What, if any, role should the ICE play?
- Should there be an e-mail network? Should there be newsletters?
- Should there be additional staff at NCHS to support this function? If so, could an additional position at NCHS be financed through a licensing fee to support a reimbursable position?
- Within NCHS can there be additional administrative support for this function, for example, contracting out disc copying and mailing discs?
- Updating systems is an essential part of maintaining automated systems; the shift from ICD-9 to ICD-10 is the most extreme example of the need for updating. However, minor updating will be needed to adjust for errors that are found, and for accommodating the classification needs of advances in medical knowledge. How will these be handled on an international scale?

Conclusion

The first international collaborative effort on automating mortality statistics attempted to address some of the issues associated with the rapid diffusion of electronic applications to vital statistics processing. It is hoped that recommendations emanating from this meeting will move the international community toward establishing standards in this area, and thereby promote the international comparability of mortality statistics.

Prepared for the first meeting of the International Collaborative Effort (ICE) on Automating Mortality Statistics, held in Washington, D.C., November 12-15, 1996. Contributing to the paper were the other members of the ICE Planning Committee: *Jack Arrundale*, General Registers Office of Scotland; *Janet Hagey*, Statistics Canada; *Lars Age Johansson*, Statistics Sweden; *Gerard Pavillon*, Institute National de la Sante et de la Recherche Medicale (INSERM), France; *Cleone Rooney*, Office of National Statistics, England; *Mary Anne Freedman*, *Donna Glenn*, *Kenneth Kochanek*, *Francis Notzon*, and *Charles Rothwell*, National Center for Health Statistics, United States.

References

1. National Center for Health Statistics. Vital statistics of the United States, Volume II, Mortality, 1993; technical appendix. Hyattsville, Maryland: Public Health Service.

2. National Center for Health Statistics. Instructions for classifying multiple causes of death, 1991. NCHS instruction manual; part 2b. Hyattsville, Maryland: Public Health Service. Published annually.

3. National Center for Health Statistics. Vital statistics, ICD-9 ACME decision tables for classifying underlying causes of death, 1991. NCHS instruction manual; part 2c. Hyattsville, Maryland: Public Health Service. Published annually.

4. National Center for Health Statistics. Nonindexed terms, standard abbreviations, and State geographic codes used in mortality data classification, 1991. NCHS instruction manual; part 2e. Hyattsville, Maryland: Public Health Service. Published annually.

5. National Center for Health Statistics. Vital statistics, data entry instructions for the mortality medical indexing, classification, and retrieval system (MICAR). NCHS instruction manual; part 2g. Hyattsville, Maryland: Public Health Service. Published annually.

6. National Center for Health Statistics. Vital statistics, dictionary of valid terms for the mortality medical indexing, classification, and retrieval system (MICAR). NCHS instruction manual; part 2h. Hyattsville, Maryland: Public Health Service. Published annually.

7. Israel RA, Rosenberg HM, Curtin LR. "Analytical potential for multiple cause-of-death data," *American Journal of Epidemiology*, Vol. 124, No. 2, August 1986, pp. 161-179.

8. Chamblee RF, Evans MC. TRANSAX: the NCHS system for producing multiple cause-of-death statistics, 1968-78, Vital and Health Statistics, Series 1, No. 20, Hyattsville, Maryland: National Center for Health Statistics, Public Health Service. June 1986.

9. Harris KW, Rosenberg HM, Kochanek KD, Chamblee RF, and Glenn DE. "Evaluation of an automated multiple cause-of-death coding system." In: *American Statistical Association. 1993 Proceedings of the social statistics section*. Alexandria, Virginia: American Statistical Association, 262-5. 1993.

10. Personal communication with Lars Age Johansson, March 11, 1996.

Plenary Sessions

Welcome

Harry M. Rosenberg, Ph.D., National Center for Health Statistics, Centers for Disease Control, U.S. Department of Health and Human Services

Good morning. We are happy you are here on this nice, crisp Washington morning. My name is Harry Rosenberg, and you are attending the first International Collaborative Effort on Automating Mortality Statistics.

I want to welcome you on behalf of NCHS, the National Center for Health Statistics. We have about 40 visitors from 19 countries, plus representatives from the United States at this meeting. We want to thank you for making your trips, which we hope were not too arduous or painful.

We have folks here from Australia, Eastern Europe, South America, North America, pretty much all over the world, and it is very exciting and pleasing to have you here. We also have representatives from a number of private organizations, and as I walk you through the folder this morning, you can find out the names of everybody and their affiliations. Of course, you will see their names on their badges.

I want to ask your indulgence for any typos that you might see in the program or on the signs. You know the classic typographical error for mortality, don't you? Morality. I think that in this meeting you may see morality mentioned in a few places. It should be mortality though, not morality. They are related, and I guess that is the subject of Dr. Sondik's presentation this morning.

This is the first ICE meeting, and we would like to encourage informality. The first ICE meeting has several purposes. One is for all of us to share knowledge and experience on automated systems for coding and processing mortality information. The second is to develop and improve our data systems. The third is to attempt to do what we can to facilitate the transition to ICD-10; and fourth, we would like to explore mechanisms for institutionalizing the type of dialogue that we hope to have over the next few days and ways of providing technical support for automation on an international level.

The meeting is going to have two different approaches; there will be a few formal presentations, but the heart of the meeting is going to be small discussion groups, and we are going to have four discussion groups on each topic concurrently.

The areas that we are going to discuss are Nosology and the training of nosologists in an era of automation, PC support, Mechanisms for technical support, Maintaining and updating decision tables, Bridge coding, Data editing and querying, and Issues related to external causes of death (which are quite difficult to handle using automated systems compared to nonexternal causes and language issues.)

Our focus is going to be on the developed countries that are actually making progress in implementing these systems. We hope that in future ICEs we will be able to focus on developing countries.

If you will bear with me, I have a lot of people to thank. There are so many people who have made this meeting possible. The U.S. Agency for International Development (AID) provided funding for some of the travel, and I want to recognize and express my appreciation to them.

Very importantly is Ms. Ginger Richards. Ms. Richards, of the Office of International Statistics has been invaluable in providing assistance, and Dr. Sam Notzon, also in the Office of International Statistics, is with me a co-organizer and convener of this meeting. I thank Sam for his support.

Elizabeth Vasquez has trained the facilitators. Elizabeth is from Management Consulting Associates of Bethesda, Maryland. Glen Pinder of the National Center for Health Statistics worked with her in training the facilitators. Finally, I would like to thank the Planning Committee that made this meeting possible, organized the meeting, and recommended its structure. The names of the members of the Planning Committee are in the program.

For those of you who are concerned about your return travel, your return reservations can be confirmed with Linda McCleary of our Office of International Statistics, who is sitting at the registration desk. She will be able to help you.

There will be demonstrations of NCHS software during the coffee breaks, and you are invited to discuss the software with Jim Hart and Donna Glenn. I understand that a number of countries have also brought software, and we hope there will be opportunities for you to demonstrate as well.

Eleven countries contributed papers. We did not ask for papers, but we are very pleased that they brought papers of their own volition. Everyone will have a copy of those papers, and if in the discussion groups there are opportunities for them to speak about their work, we would appreciate hearing from them. Unfortunately, the time constraints do not allow us to have a contributed paper session.

I want to call your attention to a special session on Friday afternoon on training for ICD-10. It is not a formal part of our meeting. Chuck Sirc of our Research Triangle Park, N.C., facility is organizing this session. Anyone who wishes to attend is welcome, and Chuck may say a word or two about that.

What I would like to do very briefly is just review with you the contents of your packets: you have a welcome letter and an invitation to a social on Wednesday night from 6:30 to 9:00, in the hotel. We invite you to come. There will be food and refreshments. Get acquainted. Tonight we are having a cash bar at 6 o'clock, which I believe will be in this room.

You have the formal program, which is called the "final" program. The final program is a little different from the program that we sent to you in the mail. There has been a little reshuffling of sessions, so throw out your old one and keep the one that you have in your package.

You also have a background paper, which we are going to use, in effect, as a textbook. So, I hope you will keep it close at hand. The background paper has a discussion of each of the topics for the focal group meetings. Please bring this to the focal group meetings. You will each be assigned to a

focal group, and later today we will be passing out the groups in which you will be participating.

Let me just ask you to take a look at today's agenda so you know what we will be doing. In the final program, please turn to page 4. We expect to have a coffee break at 10:30 and resume the morning session at 11 o'clock, when Donna Glenn will talk about the NCHS software systems for processing mortality data. Donna's presentation will be a broad overview and relatively nontechnical.

On Thursday morning, we shall have a very technical session about our software, and Jim Hart will give that presentation and will talk about such esoteric topics as source codes. Those of you who are very technically attuned, please come to the Thursday session. There will also be ample time on Thursday for discussion.

At 2 o'clock, we will have our first breakout session devoted to nosology. Then there will be a break, followed by a report to the group on nosology issues.

I have the great pleasure now of introducing to you the Director, the new Director of the National Center for Health Statistics, Dr. Edward Sondik, who will present some welcoming remarks.

Welcome

Edward J. Sondik, Ph.D., Director, National Center for Health Statistics, Centers for Disease Control, U.S. Department of Health and Human Services

Thank you very much, Harry. It is a great pleasure to be here and to welcome you all. I would like to apologize personally for the cold weather, and to thank the Committee for not having this meeting in Cleveland, which I think has 20 inches of snow!

Many of you have traveled quite a distance to get here. Welcome to the first ICE on Automating Mortality Statistics.

I cannot emphasize enough how important I think international efforts are. All of us gain so much from comparing experiences of other countries with our own. I know from my own efforts at the National Cancer Institute that comparing cancer data across the world and comparing the experiences from one area to another gives enormous insight into the problem of cancer, specifically the etiology of the disease and ideas on how to prevent and control it.

None of that is possible without a common language, and in effect, that is part of what we are dealing with here today. A few weeks ago I met in Tokyo with the World Health Organization (WHO) Collaborating Centers on the International Classification of Diseases (ICD), where the focus there was on developing and perfecting, if you will, the latest version of this common language. A common language is absolutely critical if we are going to be able to use international data and communicate with each other. It is also critical that we use current technology to enable us to deal effectively with this information. That is why I think this effort is so important. It is one in which all of us can share our experiences and move toward the four meeting goals that Harry talked about.

Let me just mention a couple of experiences I have had recently, including the Tokyo meeting—which I thought was really quite an impressive experience—especially in understanding the importance of the ICD effort, in particular the ICD-10 effort. Here in the United States we recently published the latest vital statistics on mortality and natality, and as we were briefing Secretary Shalala, she became very interested in the data. As we talked about the recent drop in infant mortality she immediately asked, "What portion of that is due to a change in deaths from Sudden infant death syndrome, SIDS?" Harry was there and said, "Well, it looks like about one-third of this drop was due to a change in SIDS deaths." There has been, as many of you from the United States know and other countries may not know, a considerable effort here to promote having infants sleep on their backs. That intervention is the only significant change over the past few years, and it correlates very well with the decline in SIDS. It is one of those small points, but a very important one, in which the mortality data—the timely effective production of the mortality data—can tell us something about how effective our public health programs are.

This example reinforces the importance of mortality as a key—the key—public health measure. I believe all countries need to do as much as we can to process this information as quickly and thoroughly as possible. To do

that, it is clear we have to use technology, not only in the coding of the information; but also, as is mentioned in this excellent background paper that Harry and others put together, in the collection and the dissemination of the information. I hope that can be added as a goal to the other four objectives. The first one being to share information about automation, then to work to develop these systems, to support and establish the ICD-10 for mortality, and finally to develop the support for these systems.

I think it is very important that we focus on the process not only in the middle, so to speak, but also at the source, where the data are actually collected, and automate that to the greatest extent possible, and then disseminate the information.

Another topic of particular current interest is smoking and lung cancer. The effort to reduce smoking and prevent lung cancer and heart disease has been ongoing since before the first Surgeon General's Report in 1965; and the results in smoking reduction in this country have been quite dramatic. We were dealing with a smoking prevalence of more than 60 percent in the sixties; it is now down to well below 30 percent, and, hopefully, heading toward 20 percent by the end of the century.

This reduction is an enormous public health feat, and it was accomplished in significant part by the stimulus provided by the availability of solid, accurate mortality statistics. It is only within the past week or so that a causal pathway relating smoking to lung cancer has been illuminated.

Sometimes in public health we make decisions not necessarily based on understanding the most detailed pathways, but on evidence that has to be considered suggestive. The evidence linking smoking to lung cancer was built on the mortality data. And therefore the reduction in smoking in this country is, in a sense, a victory for the vital statistics system.

The United States has made significant changes in the vital statistics system in this country. Those responsible for the changes are both at the Federal level and in the States. As Harry points out in the background paper, our system is very highly decentralized. What happens nationally in vital statistics is very much a function of what happens at the State level, which is the administrative source of the national data.

Through these efforts, we have been able to speed the processing of this information. For birth certificates, particularly, over 70 percent of that process is now completely automated. I see no reason why we cannot achieve even higher statistics for an automated death registration, as well.

I want to thank the organizers of this meeting and in particular I, too, want to thank the Agency for International Development (AID) for their generous financial support in mounting this effort. I also want to thank WHO for their continuing support and coordination of these international efforts.

I look forward to the results from this meeting, and I also look forward to what I know will be a continuing part of this ICE Program: translating these results to developing countries. I would imagine that technology may have to change a bit as we move to developing countries, but it is fascinating that technology is moving so quickly that in some places it is beginning to skip stages.

For example, in my trips to Eastern Europe I have been surprised at the prevalence of cellular phones. It is fascinating that telephone technology is, in effect, skipping a stage, skipping the wiring stage. Today in developing countries, networks and satellite transmissions of information are commonplace. Distance learning is becoming a reality, and I encourage you to put this in the back of your mind and be aware that the developing countries may not follow exactly the same path that we are following here, and we may need to take some shortcuts. We may need to use different technology, but I think it is very important that the results you develop be translated and applied as quickly as possible.

Unfortunately, I am going to have to leave because there is another ICE meeting, sort of an ice jam. I will return for the wrap-up. I am sure this will be a very productive meeting and, again, I thank those who have done so much to make this meeting possible.

Thank you.

Welcome

Mary Anne Freedman, M.A., Director, Division of Vital Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services

I would like to add my welcome to those of Harry Rosenberg and Ed Sondik. The Division of Vital Statistics is very excited about this meeting and proud to be sponsoring it in conjunction with our colleagues in the NCHS Offices of International Statistics and Data Processing and Services.

Dr. Sondik talked about the many uses of mortality statistics, such as research, surveillance, setting public health priorities, and evaluation. These are important functions. Those of us involved in the production of vital statistics data must provide our users with the best possible product.

We believe that automation offers an opportunity to enhance the quality and timeliness of vital statistics data. We are especially committed to automation improvements related to the cause of death, which is the focus of this meeting.

Let me begin with a few remarks about the United States' vital statistics system. Our automated systems were developed to meet the unique needs of vital registration in this country, and our registration system may differ from yours.

Vital registration in the United States is a decentralized process. It is a function and responsibility of the States, governed solely by State laws and regulations. We have 57 registration areas in the United States: the 50 States, the District of Columbia, New York City, and five territories.

The responsibility for registering and filing vital certificates in the U.S. is with the providers of service. This is different from the system in many other countries where the family is responsible for registering vital events. For death registration, the funeral director fills out and files the death certificate after he or she obtains the cause of death certification from the attending physician or the medical examiner or coroner.

While there is no Federal responsibility for vital registration, the National Center for Health Statistics is required by Federal law to produce national vital statistics by collecting data on births, deaths, marriages, and divorces from the vital records of the States. Our ability to do that well depends on the availability of quality, uniform, and timely data in every State. Thus, we work collaboratively with our State partners to promote uniformity throughout the system. Examples of these collaborative efforts include model legislation, data standards, and system specifications. We also provide fiscal and technical support to State vital statistics offices.

NCHS’s support of automated software to code cause of death is an example of our efforts to promote uniformity. This is not a new activity. NCHS began development of automated entry, classification, and retrieval of cause of death information in the late 1960's with the first edition of the ACME system to select the underlying cause of death. Since that time we have

continuously updated and refined the automated medical coding systems. Later this morning, Donna Glenn will describe the current versions of our software and our plans for ICD-10.

Currently, all deaths in the United States are coded using NCHS software, whether that coding occurs in a State vital statistics office, at NCHS, or elsewhere. Since our vital statistics system is so decentralized, a major concern for us is assuring that the software products are amenable to different user situations, different computer platforms, different user skills, etc. This means that our systems have to be fairly well documented, a factor that has facilitated the export of these systems to other countries.

What does the future hold? Dr. Sondik mentioned the rapid expansion of automation in general, and particularly in the area of vital registration. Approximately 70 percent of births in the United States are now registered electronically, using software that allows the hospital to fill out the birth certificate and send it electronically to the State or local vital registrar. The full automation of death registration has been slower since the process is more complex, and more providers are involved. However, automation of death registration is now under development in a number of our local areas. We believe it will expand rapidly once its feasibility and cost effectiveness are demonstrated.

The electronic death certificate, or EDC, will bring together medical and demographic information in one record. An important component of the EDC is the integration of on-line interactive tutorials, edit checks, and queries in the software so that the physician filling out the cause of death will get feedback and help as he records the information. We see this as a major advancement in assuring that proper certification procedures are used, and we anticipate that it will lead to significant improvements in the quality of cause of death data.

We have other vital statistics' automation efforts under way. For example, we have introduced some new product lines. In addition to our traditional publications and data tapes, we are now releasing vital statistics data on CD-ROM and through the Internet. We are also working with our States to expedite the transfer of vital registration electronically between the State offices and NCHS.

We see this week's conference as a first step in expanding the application of automation to vital statistics to provide better and more timely data. While we will be focusing on the automation of cause of death coding this week, we see this focus expanding into other automation areas over time.

This is an important symposium for NCHS. It is timely and its theme is a major priority area for our program. I am very pleased to be here with you and look forward to some very lively discussions and interesting and informative results over the next several days.

Thank you.

Results of the ICE Questionnaire on Registration and Coding Practices

Kimberley D. Peters, Robert N. Anderson, Ph.D., Harry M. Rosenberg, Ph.D., and Kenneth D. Kochanek, M.A., National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services

Introduction

Computer technology is becoming an increasingly important part of the collection, production, and dissemination of vital statistics worldwide. One of the primary purposes of the International Collaborative Effort (ICE) on Automating Mortality Statistics is to provide a forum where knowledge and experience of automated computer systems for coding mortality data can be shared. Ideally, this effort will serve to improve and standardize the automated systems already in use and will facilitate the transition to automation for countries who will be implementing computerized coding systems in the future. Improvements in the way mortality statistics are processed may result in more accurate and timely data as well as improved international comparability of mortality data. These benefits promise to substantially improve our ability to promote public health, prevent disease, and facilitate scientific research on a broader scale.

As a means of providing background data for the forum, the ICE Planning Committee recommended that a questionnaire be developed and distributed to countries participating in the first ICE conference in order to gather cross-national information on death registration and coding practices (see Appendix for a copy of the questionnaire). The purpose of this paper is to review the results of the questionnaire and to discuss potential implications for improving or implementing current or future automated coding systems.

Background

The goal of the survey was to provide information on the current and future automated mortality coding systems of the countries participating in the ICE conference. The seven sections of the questionnaire deal with:

1. death certificates
2. the death certification process
3. cause-of-death coding
4. querying and validation
5. coding certification and training
6. automated cause-of-death coding
7. training for ICD-10

Questions in the first three sections focus on the format and procedure for death certification and coding. Section 1 involves questions related to the format of the death certificates, with the first several questions addressing deviations from the World Health Organization (WHO) International Form of Medical Certificate of Death. These questions ask for specific details such as the number of lines in the cause-of-death section of the death certificate and whether a section on the interval between onset of condition and death is included. The remaining questions are concerned with the format

of the death certificate for different circumstances of death, such as whether death certificates are different for deaths certified by medical examiners or coroners, and whether there are special death certificates for neonatal deaths and stillbirths.

Section 2 focuses on procedural issues related to certification of death. For example, who completes the medical portion of a death certificate and are there situations that require special certifiers (i.e., medical examiners and coroners). The questions in section 3 relate to cause-of-death coding, including questions regarding the use of the International Classification of Diseases (ICD) and plans to begin using ICD-10. In addition, there are also questions concerning where cause-of-death coding is processed, how many nosologists there are, where they work, and whether or not multiple causes of death are coded.

The fourth and fifth sections of the questionnaire relate to data accuracy and quality control. Querying and validation of cause of death is the focus of section 4, with questions oriented toward enquiry letters to doctors, amendments to the statistical files based on these queries, and recoding of the cause-of-death data for quality control. In section 5, countries were asked about coder training and certification as well as cause-of-death coding instructions.

The questions in section 6 relate to the automation of cause-of-death coding for descriptions of systems currently in use as well as for plans of future automated systems. Each of the countries was asked whether they would like help with automated coding systems and about available computing facilities. The final section, 7, focuses on the introduction of ICD-10 and interest areas for the development of training materials.

Results

Eighteen countries responded to the questionnaire on death certification and coding. The questionnaire was mailed to countries invited to the ICE conference on Automating Mortality Statistics. The countries who responded were France, Italy, Scotland, Sweden, Denmark, England, Poland, Hungary, Catalonia (Spain), the Netherlands, Australia, Brazil, Taiwan, Kuwait, Israel, Japan, Canada, and the United States.

Death certificates

Most of the countries who participated in the survey currently use the WHO International Form of Medical Certificate of Cause of Death (figure 1), which recommends the use of three lines in Part I (the causal chain) and two lines in Part II (other significant conditions). For part I of the death certificate cause-of-death section, 2 countries use two lines, 10 countries use three lines, and 5 countries use four lines. In 1997, Australia and Denmark moved from three to four lines (see table 1). For part II, most of the countries have only one line in contrast to the two recommended by WHO. Others range from two to four lines in Part II. Italy uses a format similar to that proposed by WHO but has four distinct parts in the cause-of-death section. Thirteen of the eighteen countries also require information on the approximate interval between the onset of conditions and death, as recommended by WHO.

Figure 1. International Form for Medical Certificate of Cause of Death

CAUSE OF DEATH		Approximate interval between onset and death
I		
Disease or condition directly leading to death *	(*a*) . . . . . due to (or as a consequence of)	. . . . .
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	(*b*) . . . . . due to (or as a consequence of)	. . . . .
	(*c*)	. . . . .
II		
Other significant conditions contributing to the death, but not related to the disease or condition causing it	. . . . .	. . . . .
	. . . . .	. . . . .

* This does not mean the mode of dying, e.g., heart failure, asthenia, etc. It means the disease, injury, or complication which caused death.

Source: WHO, *International Classification of Diseases*, 1977

For most countries, death certificates differ little from the form recommended by WHO. On the whole, countries include more information rather than less. The most common deletions from the WHO form are instructions to the certifier and the interval between the onset of conditions and death. Sweden comments that physicians are more confused than helped by the instructions to the certifier. The most common additions to the death certificate are sections on whether an autopsy had been performed and the manner of death. Also asked by some countries are questions on whether the deceased was pregnant prior to or at the time of death. This is recommended in ICD-10, even though a formatting example is not provided. There is some variation in the organization of the cause-of-death section. Catalonia (Spain) organizes this section as a series of steps with immediate cause, intermediate cause, initial cause, and other processes. Italy does not use the WHO format, but has four areas under natural cause: first, the initial cause, *a;* then the intermediate cause, *b*; the terminal cause, *c*; and finally, other conditions. Poland lists the originating or external cause, followed by the direct cause, and then the intervening cause.

The majority of the countries use the same death certificate for traumatic and unexplained deaths as for nontraumatic deaths; but Taiwan, England, Catalonia (Spain), and some provinces of Canada use separate forms. For neonatal deaths, Australia, England, Italy, and Hungary use a specially designed certificate. Twelve countries also use a specially designed

certificate for stillbirths and late fetal deaths. Poland and Italy report stillbirths and late fetal deaths on the birth certificate; and Brazil, Denmark, Sweden, and the Netherlands use the death certificate. Italy is the

only country to use separate death certificates for males and females. Males have blue or green forms; females have pink or red.

Table 1. Death Certificate Characteristics, by Country

Country	Number of lines								Interval between condition onset and death	Special certificate				
	Part I			Part II								Stillbirth/Late fetal deaths		
	2	3	4	0	1	2	3	4		Traumatic deaths	Neonatal deaths	Separate form	On birth certificate	On death certificate
Australia[1]			✓				✓		✓		✓	✓		
Brazil		✓				✓			✓					✓
Canada		✓				✓			✓	✓		✓		
Denmark[1]			✓		✓				✓					✓
England		✓				✓			✓	✓	✓	✓		
France	✓				✓							✓		
Hungary		✓						✓	✓		✓	✓		
Israel		✓			✓							✓		
Italy[2]									✓		✓		✓	
Japan			✓		✓				✓			✓		
Kuwait		✓			✓							✓		
Netherlands		✓					✓		✓					
Poland		✓		✓									✓	
Scotland		✓			✓				✓			✓		
Catalonia, Spain	✓							✓		✓		✓		
Sweden			✓					✓	✓					✓
Taiwan		✓			✓				✓	✓		✓		
United States			✓		✓				✓			✓		

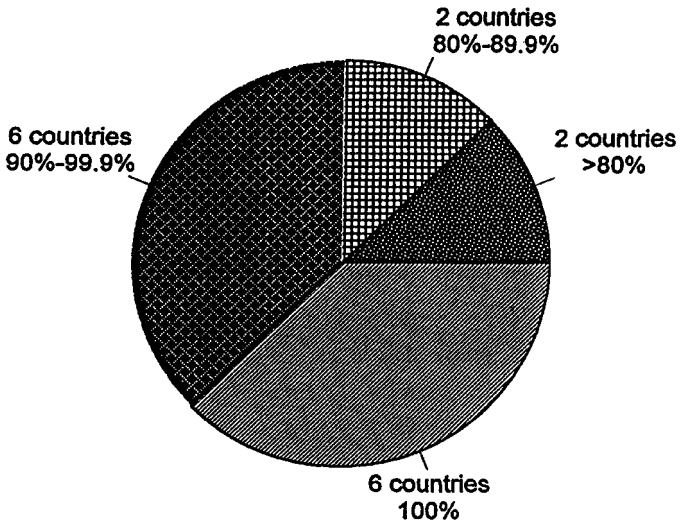
[1] Beginning in 1997, Australia and Denmark moved to four lines in Part I, and Australia moved to three lines in Part II.
[2] Italy uses a different format from that proposed by WHO. The medical section contains four distinct areas, rather than two parts.

Certification of death

Typically, only a qualified physician (usually the attending physician) or a medical examiner may complete the medical (cause-of-death) portion of the death certificate, as recommended by WHO. In Sweden, the physician responsible for treatment during the last illness certifies death. If the deceased did not receive recent medical care, and cause of death is evident and not suspicious, the physician responsible for primary care in the region the deceased lived certifies death. Some countries also allow coroners to complete the certificate. For example, in England they complete 25 percent of the certificates, but must be legally qualified and postmortem must also be conducted by a pathologist. In some countries, a coroner is a medical doctor, while in parts of the United States a coroner may be an elected county official. Scotland, England, and Poland allow licensed midwives to certify stillbirths. In Israel, a Bedouin chieftain may complete the certificate for wandering tribes; and in Australia, ship captains can complete the certificate for deaths at sea. Six of the countries report that all of the death certificates are completed by qualified medical practitioners, another eight report that at least 80 percent of the death certificates are completed by medical practitioners, and only two countries report that less than 80 percent of the death certificates are completed by medical practitioners (see figure 2). All the countries report that the following deaths are certified by medical examiners, coroners, or equivalents: unattended deaths; those with unknown causes of death; deaths from injuries or poisonings; and suspicious deaths, including suicides, homicides, and some accidents. For registration

of stillbirths and fetal deaths, most countries report that the minimum gestation required is between 20 and 28 weeks, but Japan reports fetal deaths at 12 weeks' gestation. The minimum weight criterion reported on the questionnaire varied between 400 grams in Australia to 1,000 grams in Brazil.

Figure 2. Percent of Death Certificates Completed by Medical Practioners



While the majority of countries use only one official language for death registration, Canada, Spain, and Kuwait use two languages. Denmark uses three languages, and Israel uses four. Five countries report English as their official language, and it is also occasionally used by Denmark, Sweden, and Israel. Kuwait uses English for reporting cause of death only. French is used in France and in the Quebec province of Canada. Spanish is used in Spain and in the U.S. Commonwealth of Puerto Rico. (The United States later translates the Puerto Rican text into English.) Israel, Denmark, and the Netherlands use Latin for medical reporting, and Israel, along with Kuwait, also use Arabic. The other major languages are Portuguese (Brazil), Chinese (Taiwan), Dutch (the Netherlands), Polish (Poland), Hebrew (Israel), Swedish (Sweden), Italian (Italy), Danish (Denmark), Japanese (Japan), Catalan (Catalonia, Spain), and Hungarian (Hungary).

Coding cause of death

All the responding countries use the International Classification of Diseases (ICD) for coding and classifying causes of death. The majority (12) are still using ICD-9, but are planning to make the transition to ICD-10 (see table 2). Kuwait, Brazil, Denmark, Hungary, Japan, and the Netherlands are already using ICD-10. Israel, Poland, and Sweden expect to implement it by 1997; the United States expects to use it by 1999; and all other countries expect to have made the transition by 1998. Most of the cause-of-death coding is done centrally, at the national level. Brazil and Poland code cause of death regionally, and Canada and the United States code cause of death both at the national level and at the regional level. Catalonia codes their own deaths regionally, but Spain also codes centrally. All of the countries

except Italy and Hungary centrally collate coded causes of death, and all of the countries except Italy and Canada centrally validate cause of death as well.

Table 2. Cause-of-Death Coding, by Country

Country	ICD Version		Coding location		Collating location		Validating location		Code multiple cause
	9	10	Regional	Central	Regional	Central	Regional	Central	
Australia[1]	✓			✓		✓		✓	✓
Brazil		✓	✓			✓		✓	✓
Canada	✓		✓	✓		✓	✓	✓	✓
Denmark[2]		✓		✓		✓		✓	✓
England	✓			✓		✓		✓	✓
France	✓			✓		✓		✓	✓
Hungary		✓	✓		✓			✓	✓
Israel	✓			✓		✓		✓	
Italy	✓			✓	✓		✓	✓	
Japan		✓				✓		✓	
Kuwait		✓		✓		✓		✓	
Netherlands		✓		✓		✓		✓	✓
Poland	✓		✓			✓		✓	
Scotland	✓			✓		✓		✓	✓
Catalonia, Spain[3]	✓		✓			✓		✓	✓
Sweden	✓			✓		✓		✓	✓
Taiwan	✓			✓		✓		✓	
United States	✓		✓	✓		✓		✓	✓

[1] Australia will be coding multiple cause of death with 1997 data.

[2] In 1997, Denmark began centrally collating cause of death.

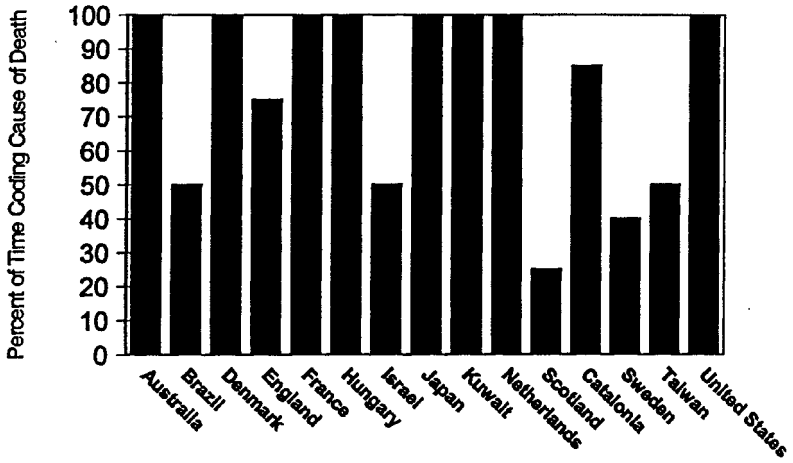
[3] Catalonia, Spain codes their own certificates regionally, but Spain uses a central location.

The number of nosologists or medical coders used to process death certificates varies widely among countries. While some of the variation may be due to factors such as geographic size of the country or whether cause-of-death coding takes place regionally or centrally, part of the variation in the number of reported nosologists may reflect the country's definition of nosologist. In the United States, the definition of nosologist refers to a person who is trained in the science of classification of disease, whereas a medical coder generally codes medical diagnoses and does not examine the relationships between medical diagnoses and the rules associated with underlying cause of death.[1] Australia does not report having any nosologists, but has highly experienced medical coders. In the near future, however, Australia will have two nosologists who will provide advice and assist in solving complex issues. Kuwait and Denmark each have one nosologist and Israel has two. Conversely, Japan reports 34 medical coders, Brazil reports 60 nosologists, Italy reports 230 nosologists, and the United States reports 110 nosologists. Brazil, Italy, and the United States may report larger numbers of nosologists because cause of death is coded regionally. For example, Italy has 230 local health agencies. Additionally, they process their records at the local level and the national level. Brazil and the United States are also geographically large countries.

[1] At the 1997 WHO conference in Copenhagen, NCHS proposed that the following definition be accepted as the standard definition of nosologist: *A nosologist is someone who can select the underlying cause of death and understand the concepts and principles behind the selection rules.* (Please see the Recommendations in Chapter 11.)

Over two-thirds of the countries have medical coders who perform tasks other than cause-of-death coding. In countries with only part-time coders, medical coders spend 25 to 85 percent of their time coding cause of death (see figure 3). France, Kuwait, Denmark, Hungary, Japan, the United States, and the Netherlands all have full-time medical coders. All of the countries code underlying cause of death, and all but Taiwan, Italy, Kuwait, Japan, Poland, and Israel code multiple causes of death as well. (At the time of this survey, Australia reported that they were not yet coding multiple causes, but would be with 1997 data.) The number of causes coded range from three for Denmark to as many as required for Brazil, Canada, and the United States.

Figure 3. Percent of Time Medical Coders Use Coding Cause of Death, by Country



Note: Data not available for Canada, Italy and Poland.

Querying and validation

To ensure precise cause-of-death coding, certifying physicians should be queried to clarify incomplete or ambiguous medical certification. The individual who certified the cause of death will often receive a letter or telephone call requesting clarification. All but four countries have a routine procedure for querying causes of death. The number of query letters ranges from 161 for Catalonia (Spain) to more than 13,000 annually for Australia. The number of query letters each country sends out is partially reflective of the population size of the country. Three-quarters of the countries who send out query letters send them from a central facility, while Brazil, the United States, Japan, Catalonia (Spain), and parts of Canada send letters out from individual regions, provinces, and States. Of the countries who responded to this question, all but Poland and Hungary amend their statistical files based on these queries. For quality control, all of the countries except France, Japan, Hungary, Taiwan, Kuwait, and Scotland check for consistency by performing some sort of independent recoding or another supervised form of verification. For those who conduct independent recoding, the outgoing error rate for the cause-of-death file is usually about 1 to 3 percent, but in one country it reached as high as 12 percent.

Coding certification and training

About half of the countries have a process for qualifying or certifying mortality medical coders. Of those countries that require certification, most offer a certification course where trainees are introduced to the coding rules and receive guidance regarding the ICD. On-the-job training by experienced medical coders was emphasized by almost all countries. Israel, Sweden, and the Netherlands also require medical training. In Israel, coders either are medical doctors or have a degree in biology. Sweden and the Netherlands require medical training in such areas as nursing, laboratory technology, or medical registration. At a minimum, the Netherlands requires courses in medical terminology, anatomy, and nosology/pathology. In the United States, courses on coding cause of death include components of anatomy, physiology and medical terminology. Canada and the United States also emphasize computer training for using their automated processing systems—MICAR and ACME (see Chapter 6 for a description of NCHS software systems). Of those who provide training, all countries train their medical coders at a central location and the United States also trains coders regionally. About half the countries continue coder training, with recurrent training occurring every few weeks to every other year. Of the countries who responded to this question, all except Poland, Denmark, and France have written instructions available for coding cause of death, and six of the countries use the United States instructions for coding cause of death.

Automated coding of cause of death

A total of nine countries currently use some form of automated coding for underlying and for multiple causes of death. Among those who have automated coding, almost all (90-100 percent) of the country's deaths are processed using the automated system. Italy has translated the MICAR dictionary into Italian. At the time this survey was conducted, about 82 percent of Italian death certificates were processed automatically, and about 95 percent of the conditions on the death certificate were automatically coded. The United States and Canada are using a U.S. automated system; and Sweden, Scotland, Italy, Brazil, England, Japan, and Catalonia (Spain) are using a combination of the U.S. system and a specially developed system. For example, Brazil began using ACME in 1983, but because it required a larger computer than many of their States had, they developed their own system that incorporates the logic of the mortality coding rules and the ACME decision tables to select underlying cause. This system is also currently in use in Colombia, Argentina, Uruguay, and Cuba. Differences between the combination systems and the U.S. systems appear to be minor. Denmark reported using a specially developed system, which includes electronically scanning the death certificate.

Among the countries who are already using automated coding, the United States began automating cause of death for deaths occurring in 1968, and began using the MICAR system for 1990 deaths and SuperMICAR for 1993 deaths. Brazil began automating cause-of-death coding in 1983 with ACME. Canada and Japan began automating cause-of-death coding for 1989 deaths, Sweden and England began with the 1993 data year, and the remaining countries began with the 1994 and 1995 data years. Among those countries who are not currently using an automated system, all plan to use automated coding in the future. Israel, Taiwan, and Australia will be using the U.S. system; the Netherlands will be

using a specially designed system; France, Italy, and Hungary will be using a combination system; and Poland and Kuwait are unsure which system they will be using.

A little over half of the countries, including Israel, Taiwan, Scotland, Italy, Brazil, Japan, Hungary, Canada, England, the United States, and Catalonia (Spain) would like help or advice with their present coding system. Scotland would like continuing support and a user group to discuss issues and problems. Italy would like help with the PC software, and Brazil and Catalonia (Spain) would like help developing and adapting decision tables for ICD-10. Some countries would also like help or advice with the introduction of automatic coding. All of the countries except Brazil have a computer environment that will support the U.S. system. In Brazil, some states have computer environments that will support the U.S. system, but others do not.

Training for ICD-10

Thirteen of the eighteen countries reported an interest in helping to develop training materials for ICD-10. Of those expressing a preference, Brazil offered assistance with the multiple cause coding; France with the rules' application; Italy with death certificate terminology; Hungary with the rules for edit matrices; Sweden with the rules for selection and modification of underlying causes; Taiwan with differentiating between cause of death and mechanism of death; Scotland with all areas; Canada with developing training materials, decision tables, tabulation lists, and edits; and Poland with practical problems of ICD-10, implementation, methods of ensuring quality and temporal comparability of mortality data.

Discussion

The results of the ICE questionnaire are informative with regard to the general implications of the transition to automated mortality coding systems. These include potential impediments to automation, possible difficulties associated with the transition to ICD-10, and other implications such as the future of nosologists in the automated environment.

Impediments to automation

One of the most significant impediments to automation involves the high costs associated with software development and programming, especially for those countries designing and implementing specially-designed systems. Examination of the ICE questionnaire reveals that many of the ICE countries are developing or intend to develop systems that are at least to some extent specially designed. As a result, these countries can expect to make large initial investments in terms of time and money to cover startup costs. One way of reducing costs is by adapting an already existing system (the U.S. system, for example). However, this presents another problem as parts of an existing system developed in one country may be sensitive to language context and may not be entirely compatible in countries using a different language. This is true of the MICAR and SuperMICAR components of the U.S. system. To minimize startup costs and to ensure that new systems are sensitive to language context, the best strategy for those yet to implement automated systems may be to use a modified version of an existing system. Several

countries currently using automated systems are successfully using a modification of the U.S. system.

There are also significant costs associated with computer equipment. However, it is evident from the survey that, with one exception, each of the countries participating in the ICE conference already has computer resources equal to or greater than that required to run an automated coding system. As a result, computer resources do not appear to be an important impediment to automation.

Another significant obstacle to the automation of mortality coding is a decentralized system for coding cause of death. While most of the countries participating in the ICE survey code cause of death centrally, some also note that they code cause of death regionally. A decentralized system for mortality coding may create difficulties for the transition to automation. If coding is to remain at the regional level, then provisions must be made to automate each region. This can entail significant costs in terms of equipment and training to ensure comparability across regions. Despite the apparent difficulties of implementing major changes in a decentralized mortality coding system, Canada and the United States have shown that it is still possible to make a successful transition to automation on a nationwide scale.

Automation and the change to ICD-10

The change from ICD-9 to ICD-10 has the potential to create some problems or at least delays in implementing automated mortality coding systems for many of the ICE countries. The transitions to ICD-10 and automation are inextricably bound. It should be questioned whether one should implement an automated system for ICD-9 when the transition to ICD-10 is imminent. Thus, countries beginning the development of automated systems should consider tailoring their automated coding systems to ICD-10. The shift to ICD-10 for those already using automated systems involves a great deal of effort in developing and reprogramming key systems including decision tables and data edits. This has significant implications for countries whose automated systems are dependent on others (i.e., those using modified versions of the U.S. system) as well as those in the process of developing systems based on the U.S. system. For these countries, the complexity of adapting automated systems to ICD-10 can affect the timing of the transition to ICD-10 as well as the transition to automation. International collaboration in developing ICD-10 components such as decision tables and data edits could facilitate the transition to ICD-10. Once the automated systems for ICD-10 are in place, countries planning to make the transition to automated mortality coding are likely to find the transition much easier.

Implications of the change to automation for nosologists

The change from manual to automated coding entails a significant change in staff requirements needed to maintain the system. Automated systems place a much greater emphasis on computer programming and data entry skills, combined with a supplemental nosological capability. This shift has important implications for the role of nosologists in automated mortality coding systems. Some nosological expertise is needed to code those death records that are not yet amenable to processing by the automated systems, and are, therefore, rejected. Nosologists are also needed for decisions about system

modifications and for independent quality control. Thus, while fewer nosologists are needed, their level of nosological skill must be high. A major concern is that by decreasing the number of nosologists, these skills will become scarce and eventually lost. The paradoxical situation is that automation may result in processing that will be difficult to update since the expertise to maintain and update coding rules will be unavailable.

Nosological expertise is critical to maintaining automated coding systems in the long run even though these skills may be used less in production. One solution to this potential problem is by diversifying nosologists' jobs; that is, using nosological skills for medical coding on a part-time basis. From the ICE questionnaire, it is apparent that in many countries this is already a reality: in most of the ICE countries, nosologists routinely engage in activities other than medical coding. As nosologists spend less time in mortality coding, their participation in training and quality control becomes increasingly critical for maintaining nosological skills. Another option for maintaining nosological skills is by establishing a certification procedure for nosologists and a continuous training program for medical coders.

Acknowledgment: We are grateful for the assistance of JoAnn Wiley (NCHS) in preparing some of the figures used in this report.

Appendix.

QUESTIONNAIRE ON DEATH CERTIFICATION AND CODING
FOR THE
INTERNATIONAL COLLABORATIVE EFFORT ON AUTOMATING MORTALITY STATISTICS (ICE)
Washington, D.C., November 12-15, 1996

Results of this questionnaire will be tabulated for presentation at the ICE meeting. Accordingly, we would appreciate your reply no later than September 13. Please send your completed questionnaire *along with* a copy of each of the types of death certificate used in your country, i.e. for fetal death/stillbirth, infant death/neonatal, perinatal death, death (also copies of certificates used for medical examiner/coroner -- unattended, unknown cause, injury and poisoning, suspicious, etc.). Please also *send* a copy of your coding instructions, if available in written form.

Death certificates

1. Do you use the International Form of Medical Certificate of Cause of Death, as recommended by the World Health Organization (WHO)?
_____Yes _____No **(If your answer is "No," please go to Question No. 2)**

   If you use the International Form of Medical Certificate of Death, please answer the following questions.

1a. How many lines are provided in Part I?
*Specify number*_____

1b. How many lines are provided in Part II?
*Specify number*_____

1c. Do you include the question regarding interval between onset and death?
_____Yes _____No

2. Does your certificate differ in any significant way from the International Form, for example, in asking a question about surgery, or about pregnancy, or about whether the death was a homicide, suicide, accident, could not be determined, under investigation?
_____ Yes _____ No

ICE Questionnaire on Death Certification and Coding

Death certificates (continued)

2a. If your certificate does differ significantly from the International Form, please describe how your certificate departs from the International Form.
Use additional sheets if necessary

_____
_____
_____
_____
_____
_____

3. Is the same death certificate used for deaths certificates completed by a medical examiner, coroner or equivalent?
_____ Yes _____ No

3a. If a different certificate is used by for deaths certificate by a medical examiner, coroner, or equivalent, please *provide* a copy.

4. Do you use a specially-designed certificate for neonatal deaths?
____Yes ____No

4a. If a specially-designed certificate is used for neonatal deaths, please *provide* a copy.

5. Do you use a specially-designed certificate for stillbirths/late fetal deaths?
_____ Yes ____ No.

5a. If a specially-designed certificate is used for stillbirths/late fetal deaths, please *provide* a copy.

Certification of death

6. Specify all of the types of persons who complete the medical portion (cause of death) of death certificate, that is, whether they are a qualified medical practitioner, medical examiner, coroner, registered nurse, etc.
Specify all of the types _____
_____
_____

ICE Questionnaire on Death Certification and Coding

Certification of death (continued)

7. Approximately what percent of your deaths are completed by qualified medical practitioner?
Specify percent _____

8. If some deaths are certified by a medical examiner, coroner, or equivalent, what types of deaths are covered?
*Specify type, for example, unattended, unknown cause of death, injury and poisoning, suspicious, etc.*_____
_____
_____

9. For stillbirths/late fetal deaths, for what minimum gestation/weight is registration required?
Specify minimum gestation/weight _____

10. How many languages are officially used in your death certificate?
Specify number _____

10a. What languages are used?
Specify languages used _____
_____

Coding cause of death

11. Does your country use the International Classification of Diseases (ICD) for coding and classifying causes of death?
_____ Yes _____ No

11a. If you do use the ICD, which Revision are you using, for example, ICD 8, 9, 10?
Specify which revision of the ICD you are using _____

11b. If you do not use the ICD, what coding and classification system do you use?
Specify coding and classification system _____
_____

11c. If you are not currently using ICD-10, do you intend to begin using ICD-10?
_____ Yes _____ No

Coding cause of death (continued)

11d. If you intend to change to ICD-10, beginning with what data year?
Specify data year _____

12. Are causes of death coded centrally (that is, at the national level), regionally, or some mixture of central and regional?
Specify by checking one
Centrally _____ Regionally _____ Mixture of central and regional _____

13. Are coded causes of death centrally collated?
____Yes ____ No

14. Are coded causes of death centrally validated?
____Yes ____ No

15. How many nosologists are used to process all the death certificates in your country?
Specify number _____

15a. How many of these nosologists work in a central processing facility?
*Specify number*_____

15b. How many of these nosologists work in regional or decentralized facilities?
Specify number _____

16. Do your medical coders perform work other than cause-of-death coding?
_____Yes _____ No

16a. If your medical coders perform work <u>other</u> than cause-of-death coding, what percent of their time do they spend coding cause of death?
Specify percent _____

17. Do you code underlying cause of death?
_____Yes _____No

18. Do you routinely code causes of death other than the underlying cause, that is, *multiple causes of death*?
_____Yes _____ No

Coding cause of death (continued)

18a. If you code multiple causes of death, how many causes are coded in all, including the underlying cause?
Specify number _____

Querying and Validation

19. Enquiry or *query* letters to doctors to obtain further information can ensure a more valid or more specific cause of death. This can not only aid in nosological coding but can provide a more informative cause of death. Does your country have a routine procedure for querying causes of death?
____ Yes ____ No

19a. If your country does send out letters of enquiry, approximately how many such letters are issued per year?
Specify number _____

19b. If your country does send out letters of enquiry, are the letters sent out regionally or centrally?
Specify by checking one
Centrally _____ Regionally _____ A mixture of central and regional _____

19c. Does your country amend the statistical file based on these queries and other information?
Yes _____ No _____

20. For quality control, do you perform any type of independent recoding, or other form of supervision of coding consistency?
_____Yes _____ No

20a. If you conduct independent recoding, please indicate the outgoing error rate of the mortality medical (cause of death) file as a percent of the records processed.
Specify percent error rate _____

Coding certification and training

21. Is a qualification or certification process used for mortality medical coders?
_____Yes _____ No.

Coding certification and training (continued)

21a. If a qualification or certification process is used for mortality medical coders, please describe the qualification or certification process.
*Use separate sheets if necessary*_____
_____

22. Are your coders trained centrally or regionally?
Specify by checking one
Centrally _____ Regionally _____ Mixture of central and regional _____

23. Is there any formal continuing training of coders?
_____Yes _____No

23a. If there is recurrent training of coders, how frequent is it?
Specify frequency in terms of number of times per year _____
_____

24. Are written instructions available for coding cause of death?
_____ Yes _____ No

24a. Does your country use the U.S. instructions for coding cause of death?
_____ Yes _____ No

24b. If written instructions are available for coding cause of death, please *provide* a copy if you do not use the U.S. instructions.

Automated coding of cause of death

25. Do you use any form of automated selection of the underlying cause of death?
_____ Yes _____ No

26. Do you use any form of automated processing of multiple causes of death?
____Yes ____ No

Automated coding of cause of death (continued)

27. If you use any form of automated coding, please respond to the following questions: **(If you do not use automated coding, go to Question No. 29)**

27a. What percent of your country's deaths are coded using an automated system?
Specify percent _____

27b. Indicate whether your system is entirely based on the U.S. system, or is specially-developed, or is a combination of the U.S. and a specially-designed system.
Specify by checking one
U.S. system based _____
Specially-developed _____
Combination of U.S. and specially-developed _____

27c. If your system is other than the U.S. system, please describe it.
*Use separate sheets if necessary*_____
_____
_____
_____
_____
_____

28. For what data year did automated coding begin in your country? Specify _____ **(Go to Question No. 31)**

29. If you do not use an automated system, do you have plans to use such coding in the future?
_____ Yes _____ No

30. If you plan to use an automated system for coding cause of death, what type of system will it be?
Specify by checking one
U.S. system based _____
Specially-developed _____
Combination of U.S. and specially-developed _____

31. Do you require any help or advice with your present coding system?
_____Yes ____No

Automated coding of cause of death (continued)

31a. If you require help or advice with your present coding system, please specify type of help or advice.
Specify type of help _____
_____
_____
_____

32. Do you require any help or advice with the introduction of automatic coding?
____ Yes ____No

33. Do you have a computer environment that will support the U.S. system as described in Item 34?
_____Yes _____ No

34. Minimum Configurations for the ICD-10 MICAR Software Systems

Super-MICAR Processing, MICAR 100/200 PC-ACME/TRANSAX

Option	Minimum Required	Suggested
CPU	486/50	Pentium 75 or higher
RAM	4 Megabytes	16 Megabytes
HDD	340 Megabytes	850 Megabytes
Monitor	VGA	VGA
Operating System	MS Windows 3.1	MS Windows 3.1
Other	3.5" floppy, mouse	3.5" floppy, mouse

Note: These systems will also run under OS/2 2.1 or greater in a Win OS/2 session. They can also be run on the PowerMac, but they are not supported on this platform.

Automated coding of cause of death (continued)

34a. If you do not have a computer environment that will support the U.S. system, please describe your computing environment in terms of its hardware, operating system, network, and other.
Use separate sheets if necessary _____
_____
_____
_____

Training for ICD-10

35. Would you be interested in helping develop training materials for ICD-10?
____Yes ____No

36. If you are interested in helping develop training materials for ICD-10, in what particular areas?
Specify areas _____
_____
_____
_____

ICE Questionnaire on Death Certification and Coding

Questions about the survey may be directed to Harry M. Rosenberg. Please return the completed questionnaire to the following address, no later than August 30.

Harry M. Rosenberg, Ph.D.
Chief, Mortality Statistics Branch
Division of Vital Statistics
Room 840
6525 Belcrest Road
Hyattsville, Maryland 20782
Telephone: 301-436-8884, extension 175
FAX: 301-436-7066
e-mail: hmr1@nch08a.em.cdc.gov

Please provide the name, affiliation, and address (mailing, telephone, FAX, e-mail) of the persons who completed this form:

_____

_____

_____

_____

_____

_____

_____

Description of the National Center for Health Statistics Software Systems and Demonstrations

Donna E. Glenn, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services

Introduction

Since the introduction of ACME effective with deaths occurring in 1968, the U.S. National Center for Health Statistics (NCHS) has been interested in a complete automated system for coding causes of death. This morning I am going to discuss the development and implementation of ACME, TRANSAX, and MICAR. Dr. Rosenberg's paper contained a brief description of our system in chronological order of development. I will be discussing each system in detail; but first, I want to review the acronyms:

MICAR: Mortality Medical
Indexing
Classification
And
Retrieval

MICAR generates the multiple cause ICD codes that are input to ACME from the literal text. We have two methods of data entry for MICAR. One is called PC-MICAR and the other is SuperMICAR data entry. I will demonstrate both systems so you can see the differences. Both accept text as input, but PC-MICAR requires translation of the text into a standard format and terminology. SuperMICAR accepts everything exactly as it is reported on the U.S. certificates.

ACME: Automated
Classification of
Medical
Entities

ACME selects the underlying cause of death from the multiple cause codes.

The final program is TRANSAX. We ran out of nice little acronyms. The name comes from TRANSlation of AXis. TRANSAX generates two different sets of multiple cause codes for tabulation. One we refer to as the "entity-axis," and the second we refer to as the "record-axis."

Since the beginning of software development we have tried to keep up with computer technology. As you heard this morning, the original program was written in PL1 and ran on an IBM mainframe. This definitely limited the number of States that could use the system, and also, the number of foreign countries that could use it. When we began to develop MICAR, we decided that we would provide a data entry system. Although the processing program was still on the mainframe and still written in PL1, we wanted to provide a mechanism for more States to have access to the actual program. The original data entry program was written in dBASE III, but we soon learned that dBASE was limiting our ability to provide high-quality software. So we converted to the programming language C. This choice was beneficial because it allowed us to use a wide variety of platforms at the mainframe and personal computer (PC) level. It provided us with the flexibility of using either the mainframe or

the PC. There is still a version of TRANSAX written in C running on the mainframe. However, as we were developing MICAR, NCHS began to move away from mainframe computers to downsized systems. For ICD-9 data, NCHS still runs our mainframe programs, but many States use our PC systems. For ICD-10, our entire system will be PC-based. The system will be written in Windows 3.1, but it should run in Windows 95 and Windows NT. I will be using the DOS version of our software for this morning's demonstration. This also means that we will discuss ICD-9 codes. We have the preliminary Windows version of the data entry programs that you may see during breaks.

For data entry, we have two options: PC-MICAR Data Entry or SuperMICAR Data Entry. With the PC-MICAR system, you have to translate the reported causes of death to generate input to MICAR. This system requires the user to understand medical terminology and some anatomy. With this knowledge, users can translate or change the order of the terms. Diseases, injuries, and external causes reported on the death certificate are entered in almost literal text. In addition, medical entities can be entered as abbreviations or entity reference numbers.

An entity reference number (ERN) is a 6-digit number that we have assigned to every term in our dictionary. There is no relationship between entity reference numbers and ICD codes. The numbers are totally independent. When we first started the system, we reserved the first 200 numbers for frequently occurring conditions because at the time our coders were used to entering numeric values and wanted to use code numbers. For example, ERN 99 is pneumonia and ERN 1 is acute myocardial infarction. Although we do not encourage entering numbers, the system has this option. While PC-MICAR reduces the complexity of multiple cause coding, it still requires formal classroom training.

The other option for data entry into MICAR is SuperMICAR. While SuperMICAR's primary function is to allow entry of literal information from the death certificate, it has additional functions that make it more than just a data entry package. SuperMICAR accepts almost everything in literal text. There is no coding involved. Even the interval between onset of the disease and death, the duration field, is entered in full text.

PC-MICAR

I am now going to describe how to use the systems. I plan to start with PC-MICAR so you can appreciate the power of SuperMICAR. The interface used for all our systems is alike and is similar to a Windows program interface. In order to activate the menu, the ALT key is used with another key. We have built in some safeguards. When we originally implemented MICAR at NCHS and in some of the States, we had such small computers that we had to provide a warning as to how many records could be entered. The first screen shows the number of records that can be entered, based upon the amount of free space on the PC. At the time we implemented PC-MICAR at NCHS, we could only enter 500 records. The coders had to close each batch, copy it to a diskette, then enter another 500 records. Of course, that was several years ago; we now have much more powerful computers.

The certificate number is a six-digit code. A new batch will initially display certificate number 000001. This can be changed and the system will increment the number by one with each new certificate screen. Sex is a coded variable. The user can enter numeric values 1, 2, or 9 or alpha variables M, F, or U. If the alpha values are entered, the system automatically converts them to numeric values. The next variable is the month and day of death, entered in that order, as numeric values. Age must be entered as a three-digit coded variable. The first digit represents the units of age (years, months, hours, etc.), the last two digits represent the number of units. For example, if the certificate is for the death of a 3-month-old child, the coders enter 203 (2 representing months and 03 indicating the number of months) 54 years is entered as 054(0 representing years and 54 indicating the number of years).

With PC-MICAR, the entry operator must indicate exactly where each term is reported on the certificate. Part of the number is either 1 or 2. Lines in Part 1 are indicated by alpha codes. The first line is indicated with an a, and the second with a b. The code number on the line is incremented automatically by the system.

1. With the dictionary on as the user is entering data, the terms are automatically checked against a list of valid words developed from the MICAR dictionary. As soon as the first term has been entered, the system will indicate all spelling errors.

   MASIVE HEAR FAILURE

   MASSIVE is misspelled. The user may select to have the system provide a list of possible terms. If the correct word is on the display, the user highlights the correct word and the system replaces it without additional typing. The second word is HEART, also misspelled. Sometimes the correct selection will not be generated. When this happens, the user may choose to retype a single word.

2. After all words are correctly spelled, the computer will try to match the term to an entry in the MICAR dictionary. In this case, the number 000070 is automatically entered in the column headed with ERN. This is the entity reference number assigned to that term.

3. Duration (the interval between onset of disease and death) must be entered as a coded variable. The code structure generally follows the code structure used to enter age. Special codes are used for terms frequently entered in the duration block: for example, ACUTE - 702; BRIEF - 706. If a date is reported, the user must calculate the duration using the date of death and date reported in the duration block. If dates are reported as a span, for example, 10/1/96 - 10/6/96, the user must calculate the actual duration. In this case, the user would enter 405 for five days.

4. Even if the words are correctly spelled, the full term may not be in the dictionary, for example, MASSIVE HEART FRACTURE. The user can choose either to accept the term as entered or to re-enter the term if an error is noted. If the term is accepted as entered, the ERN is coded as 999999.

5. Many adjectives reported with disease are frequently considered medically insignificant by the ICD. These adjectives do not appear in the MICAR dictionary; however, the user is instructed to enter the words in correct MICAR order. The system is designed to drop a maximum of three words while trying to match a term in the MICAR dictionary. On the first line I entered MASSIVE HEART FAILURE, which is assigned ERN 000070. If I enter ERN 70, the term returned is HEART FAILURE. The system drops the word MASSIVE without the user's knowledge.

6. We know that certain terms are very important to the ICD for a specific group of diseases. For example, PRIMARY almost always affects the ICD assigned to a neoplasm. The system is aware of these limitations and will not drop certain words if the resulting term has been assigned an ICD within a specific range. If I enter PRIMARY LEG CANCER, the system indicates that this term is not in the dictionary. However, if I enter PRIMARY TUBERCULOSIS or PRIMARY TB using an abbreviation, the system assigns ERN 000926. If I enter ERN 926, the term matched is TUBERCULOUS. The word primary was dropped by the system.

   There are cases where this procedure may drop a word that should not be dropped. Scotland has found ACUTE FATTY LIVER DEGENERATION matches ERN 086903. In turn, this ERN generated the term FATTY LIVER DEGENERATION. The ACUTE was dropped by the system. ACUTE should not have been dropped in this case because this term changes the code.

   With the ICD-9 version, we find these errors through comments from both coders using the system and from our quality control procedures. With ICD-10, we are very fortunate to have the index in electronic form. We will use the index to select every condition that has a separate code assigned to the acute form and add that form to the dictionary. These additions should eliminate the majority of errors.

7. Because of the variety and difficulty of reporting, external causes are handled through a system of programmed instructions, called prompts, designed to combine all the relevant information together to form a medical entity.

   For example, if a gunshot wound to the head is reported on the death certificate, the user enters HEAD GUNSHOT WOUND, and then moves to the injury description at the bottom of the screen. "Self-inflicted by 25 caliber hand gun" is reported in the injury description block on the certificate. The user enters a greater than symbol (>) to initiate the prompts. The first screen lists various external causes. For this list, the user selects I for firearms. The next information required is the type of weapon. The user would select 05 for the 25 caliber hand gun. The system also needs to know the circumstances surrounding the external cause. In this case, the coder selects 06 for self-inflicted. The MICAR term generated for this entry is >I0506. This term is assigned ERN 900267. The ERNs assigned to external causes begin

with the number 9. Again, there is no relationship between the ICD and the ERN codes. The leading digit immediately identifies the ERN as an external cause.

SuperMICAR

Even as MICAR was being placed into production in the United States, NCHS began to develop an enhanced data entry version—called “SuperMICAR.” The first thing you will notice is that the interface for SuperMICAR and PC-MICAR are almost identical. We tried to keep all the interfaces alike since some of our coders have switched from PC-MICAR to SuperMICAR; we did not want to introduce possible errors by having a different interface. However, the certificate screen is very different. The SuperMICAR screen matches the cause of death section from our standard death certificates. In Part I, we have lines a, b, c, and d, and there is also a Part II section. At the bottom, the injury description is a separate block. The place of injury is entered in full text.

We have one new item—activity code. This is for ICD-10, but we are experimenting with it using our 1996 data. We have asked our coders who use SuperMICAR to code it for 1996. Eventually, we will automate this information if it is reported frequently on the certificates.

The same demographic variables are required by SuperMICAR: sex, date of death, and age. Sex is entered the same for both systems; however, SuperMICAR retains the alpha code on the screen. If sex is entered using code 1, an M appears in the field. With PC-MICAR, the numeric value is shown on the screen. In SuperMICAR, age is entered as a number of units followed by the type of units in full text. For example, instead of coding 203 to indicate 3 months, the user would enter 3 in the number of units and "months” in the unit’s field. If no units are entered, the system assumes years.

1. There is still the benefit of the spell checker, and the ability to accept, retype, or list options. The correction process is the same for both systems.

2. In SuperMICAR, duration is entered as reported. If the certifier enters ACUTE in the duration field, the user enters ACUTE in the duration field. There is no coding involved. Dates reported in spans are entered as spans.

3. There is one big difference between PC-MICAR and SuperMICAR. SuperMICAR does not assign an entity reference number as the terms are entered. Spelling is checked on an interactive basis, but the assignment of the entity reference number is done in batch mode.

To show the power of SuperMICAR, I could enter the following causes, as examples:

A. PRIMARY HEART FAILURE with duration ACUTE
B. PRIMARY LEG CANCER with duration CHRONIC
C. PRIMARY TB with duration of 10/1/96–10/6/96.

After all records have been entered, the user selects "process all records." The file may be reprocessed as many times as desired. As the file is being processed, a status bar appears on the screen. If it is a large file, it may

take 30 minutes to process. Users can use this time estimate to do another task while the file is processing. Most average size files will run in 5 to 10 minutes.

In example A, PRIMARY HEART FAILURE, SuperMICAR generated the term HEART FAILURE and the duration, reported as ACUTE, is coded 702. Users do not have to do any coding; the word ACUTE is entered. When example B is processed, the duration, CHRONIC, is coded 703, but the term, PRIMARY LEG CANCER, is assigned ERN 999999 (as in PC-MICAR) because the system realized that "primary" could not be dropped. When example C is processed, the term "primary" is dropped and the system calculates the duration as 5 days.

When the gunshot wound was entered in PC-MICAR, the coder had to go through the prompts and answer questions screen by screen. With SuperMICAR, line a shows gunshot head wound, which is acceptable because it is in the dictionary and it also shows the MICAR term >I0506. The current version of SuperMICAR has some of the external causes, but not the more complicated ones.

I have passed out a sample certificate that I am going to use to describe what ACME is doing.

Male, Age 45 years		Code for record
PART I (a)	Congestive heart failure	428.0
	Due to, or as a consequence of	
(b)	Stomach ulcer with hemorrhage	531.9 578.9
	Due to, or as a consequence of	
(c)	Rheumatoid arthritis	714.0
	Due to, or as a consequence of	
(d)		
PART II	Other Significant Conditions	
	contributing to death but not resulting in the underlying cause given in Part I.	
	Myocardial infarction, cancer of breast, and circulatory insufficiency	410 175 459.9

While it is very straightforward to enter cause of death for MICAR, our multiple cause coders have to apply hundreds of rules to get the correct codes. With MICAR, all the rules are built into tables (rather large tables). As an example of what MICAR can do:

1. Cancer of the breast is reported in Part II. MICAR is programmed to reference the sex and assign either 175 (male) or 174.9 (female). In the MICAR dictionary, breast cancer is assigned to the female code, not the male code.

2. There are certain diseases that are commonly reported or coded in the ICD with a site given. We teach our coders to look at other reported diseases to determine the site of that disease. In the example, stomach ulcer with hemorrhage is reported on line b. Instead of entering a code for hemorrhage without a site, stomach hemorrhage is coded by referencing the site of the ulcer.

ACME

After MICAR has processed the record or the codes have been assigned manually, it is ready to be processed by ACME. ACME applies the WHO rules for selection as set forth in the ICD. The first step is to determine what we call a tentative underlying cause (TUC). To do this, ACME must determine if there is a causal sequence between the codes entered in Part I. In our example, ACME determines that 428.0 (code on line a) can be due to 714.0 (the code on line c, which is also the lowest used line in Part I in this example). ACME then determines that both 531.9 and 578.9 (codes on line b) can be due to 714.0. Therefore, the TUC is 714.0, determined through application of the general principle. In this case, there are no modification rules to be applied. The final underlying cause is rheumatoid arthritis (714.0).

TRANSAX

The final system I will discuss is TRANSAX, which is probably the least understood of any of our systems. Before I explain what TRANSAX does, I would like to describe the difference between entity-axis codes and record-axis codes. The input codes to ACME are referred to as entity-axis codes. For entity-axis coding, each condition reported on a death certificate is entered on an entity-by-entity basis with minimal regard to other entries. The location of each code with respect to where the condition is reported on the death certificate is also entered. These entity codes do not always provide a good description of the cause of death from a record standpoint. Record-axis codes provide a better characterization of the cause of death based upon everything reported on a given record.

In our example, the certifier entered stomach ulcer with hemorrhage on line b. The entity-axis codes are 531.9 (Stomach ulcer without mention of hemorrhage) and 578.9 (stomach hemorrhage). TRANSAX uses the ICD linkages to link stomach ulcer without hemorrhage with stomach hemorrhage to come up with a single code, 531.4 (stomach ulcer with hemorrhage).

The entity-axis data may also contain two codes that mean essentially the same thing, but one may be more specific. TRANSAX will eliminate the least specific disease. In our example, circulatory insufficiency is deleted in preference to myocardial infarction. TRANSAX goes through each ICD code on an individual basis and determines what other codes can influence it. The system performs combination and modification linkages first, then it deletes unnecessary codes. When all processing is complete, all means of identifying where a given condition was reported have been lost. Since the record-axis codes cannot tell us where the cause was located on the record, the codes are sorted in ascending code number order.

Here is what we end up with on that record:

Entity-axis codes:

Line 1, code 1:	428.0	congestive heart failure
Line 2, code 1:	531.9	gastric ulcer without mention of hemorrhage
Line 2, code 2:	578.9	gastric hemorrhage
Line 3, code 1:	714.0	rheumatoid arthritis
Part II, code 1:	410	myocardial infarction
Part II, code 2:	175	cancer of the breast for male
Part II, code 3:	459.9	circulatory insufficiency

Record-axis codes:

Code 1:	175	male breast cancer
Code 2:	410	myocardial infarction
Code 3:	428.0	congestive heart failure
Code 4:	531.4	gastric ulcer with hemorrhage
Code 5:	714.0	rheumatoid arthritis

Throughput

Thus far I have discussed what the systems do and how they work, but an important question is how many records can we actually process on an automated basis? With PC-MICAR, we can process 95 percent or more of the records. With SuperMICAR, we can process at least 75 percent of the records. We have not completed the programming for external causes within SuperMICAR. When this is completed, the throughput rate should be almost equal for both systems. In terms of the dictionary, we are trying to set up a system where we do an annual update to the dictionary. Excluding external causes, our acceptance rate for the terms in the dictionary is 99 percent. So, we are finding fewer terms that need to be added, but we want to continue to update the annual dictionary to keep the system current.

MICAR200 is the rules application program. Using this program, breast cancer was converted to 175 after determining that the decedent was male. MICAR200 is able to process 95 to 97 percent. We have not finished programming everything, but eventually most of it will be automated. For example, because of the sensitive nature of maternal deaths, we decided to reject all maternal deaths for review. With ICD-10, there will be three codes for maternal deaths—one for less than 42 days, one for less than a year, and one for more than a year. With ICD-10, maternal deaths will be automated. Records with surgeries or therapeutic misadventures have not been programmed.

With ACME, 98 percent of our records are automatically processed. The remainder has to be coded manually. Therefore, we must keep trained multiple cause coders and underlying cause coders on staff. Because the easy records are coded automatically, the medical coders must be well trained. The rejected records that require manual coding are the most difficult to code.

Accuracy

Based on a reliability study done by the United States with PC-MICAR data entry, the underlying cause assignment has an error rate of 0.33 compared with manual coding. The multiple cause assignment has an error rate of 0.60. Using SuperMICAR, the error rates are a bit higher, but that system does a whole lot more than PC-MICAR. With SuperMICAR, the underlying cause assignment has an error rate of one-half percent and the multiple cause assignment has an error rate of 1 percent. The ACME error rate for underlying cause is at one-half percent. With TRANSAX, the multiple cause codes have a one-half percent error rate.

I want to talk to you about some of the benefits of the systems. Under the ACME system, the nosologists have to apply data entry rules that are in hundreds of pages of our documentation. The MICAR and SuperMicar automated systems definitely improve the accuracy and the consistency of data. Since every medical condition has its own ERN, we can also retain more detail from

the death record than before. The complexity of medical coding is also reduced. Particularly with SuperMICAR, we are getting to the point where all data items from the death records can be captured in electronic fashion during one entry session. In the past, NCHS had a staff of demographic coders who coded the demographic variables and another staff of medical coders who coded cause of death. Therefore, all death certificates were processed by at least two different people. Now, we are building to the potential where the records only have to be processed by one person. In the future, we plan to implement an electronic death certificate system, which will produce coded data as the information is entered by the source provider.

Issues

To review the issues in the background paper, the first one is nosology and the training of nosologists. Even with the automated coding systems, we require specially trained nosologists. We need the nosologists not only to code any records rejected by the automated systems, but also to provide the specifications that make these systems work.

The second is the training of PC support staff that may be unique to the United States, but we went from mainframe computer processing to PC processing and introduced the PC version into our States. A lot of our States were making the same type of transition and with our decentralized system it was very difficult to help them. We introduced an 800 number so that people could get to us, but those on the West Coast and Hawaii are still having difficulties. So, there needs to be a way that we can work with and train people to be what could be considered PC managers.

Third, we also want to discuss the decision tables and mechanisms for updating them and come to some agreements. We want to discuss some of the differences between countries and see if we can come up with a system where we all know what everyone else is doing.

Fourth, bridge coding or comparability studies are very important, particularly if you work with any part of the system. If you implement any part of the automated system, you need to compare the automated results with your manual coding because there will be a difference. Between Great Britain and the United States we found a big difference in Rule 3. With ICD-10 we should come to a compromise, hopefully by loosening our interpretation of Rule 3, while Great Britain tightens their interpretation. However, this issue would not have come up had we not been using these automated systems.

Fifth, editing and querying need to be discussed. We need to determine how to get better data and more effective ways to query the certifier. SuperMICAR actually includes a function to generate query letters. It is not perfect, but it does generate query letters for what we call rare causes. With ICD-10, we will work on making the system even stronger.

External cause is another issue to discuss. We are working on it for SuperMICAR, which is good for the United States, but that leads into the final issue of language. Even in English-speaking countries, the language used for external causes is different. For example, England uses the term “lorry.” In the United States, we don’t even know what that is, and we would not have had it in the dictionary until we started working with other countries.

ICD-10

With ICD-10, there will be some potential changes. Because we have the two data entry packages, PC-MICAR and SuperMICAR, we are eliminating what we used to call MICAR100. It was simply a dictionary match program. The PC programs already do this work. For those countries who use our automated system beginning at the MICAR200 level, there will be no change. You still can use the system. In other words, you can use the rules application program and use whatever mechanism you have defined as a way to match the dictionary and assign entity reference numbers.

We have also made some minor changes to ACME. Most deal with additional information that we are going to collect. First and foremost, the year of death will be four digits. Hopefully, that will get us around the year 2000 problem on the PCs. From ICD-10, we picked up the new activity code, which essentially tells what the decedent was doing when he or she died. It will not be linked with the cause-of-death code, but will be kept as a separate field. Place of injury will be coded similarly—as a separate code and a separate variable that can be linked with the ICD codes. We will also carry the manner of death item. On the U.S. certificate, there is a check box asking whether it is a homicide, suicide, or natural death, etc. In the past, it has influenced how the records were coded, but we could not determine the actual entry reported in the block. We had to have this information for MICAR to assign the correct codes. Therefore, this information will now be included in the final record layout. We are also adding a version control number on all our software.

That is the end of my presentation. We do invite questions and comments. We also will have the computer set up in the back of the room if you would like to see the software. I would particularly like you to see some of the retrieval capabilities of SuperMICAR.

DR. PAVILLON: Concerning the conditions of the training, it seems that the United States has experienced a difference, and I would be very interested in that.

MS. VASQUEZ: Is there anyone who would like to briefly describe what happened in terms of your personnel and training requirements in making the conversion to automation?

MR. ARRUNDALE: In Scotland, we have not really been able to assess the effect it is going to have on the manpower as yet, but we are expecting to have a decrease in the number of coders by about four overall. That is out of about, well, it is a bit difficult because we have got a lot of part timers—a lot of people job sharing—but it is about 4 out of about 14. The coders do not just code cause of death. They code all the demographic data on births and marriages and deaths and divorces as well, but we are hoping to make that kind of reduction.

On training we have not had a great deal of a problem because we have just been converting existing coders to use the automatic system. I think we might have far greater problems when we have to train coders from scratch just to code the hard cases rather than coding all the cases, and I do not know how we are going to go on with that.

MR. JOHANSSON: In Sweden, we implemented automated coding in 1993, and our experience was that at first we needed more coders to build the dictionary, to check the software and so on. Now that the program is running, and we have a fairly complete dictionary, we can do with about half as many coders as before. Of course now we have ICD-10 and cannot reduce any staff because of that. We have no experience training new coders since we introduced automatic coding.

May I go on with a concern of my own while I am standing here? You said that you were thinking of developing the decision tables and basing them on ERNs instead of ICD codes. That could pose a problem to us who work in other languages than English. It is quite difficult to match a Swedish medical term to an English entity term. So, I hope you will keep an ICD code version as well.

MS. GLENN: Thank you. That is a good comment. I had not really thought of that. Keeping it at the ICD level does make it better internationally. Just in terms of building the system, we, too, had to have a lot of nosology help to build the system. It did require extra people, but once it was built we saw the reduction.

DR. SANTO: In Sao Paulo, when we introduced ACME in 1983 we needed to increase the number of coders. Actually, there were five coders that were coding multiple causes for ACME. I would also like to comment on an additional concern that Harry Rosenberg of NCHS expressed—about the loss of nosologists—

our experience with the system is different. In Brazil, each death certificate is processed individually with ACME, and during the processing there is a lot of interaction with the coder. When we have an issue related to the intent of the certifier, a dialog box that interacts with the coder opens up. So, instead of losing nosologists the system is teaching the coder to make the right decision.
Thank you.

MS. GLENN: I like that. That is a good approach. We have lost nosologists at NCHS. At one time, I guess back when we were doing multiple cause coding, we had 20 or 30 coders on staff. To code the MICAR rejects we are down to, I think, 12 coders. We had to do something to make up for the people that we were losing, and the system, particularly SuperMICAR, has enabled us to continue multiple cause coding. We would not have been able to continue multiple cause coding with the number of medical coders we currently have on staff.

Report on the Collaborative International Study on Multiple Cause Analysis

Gerard Pavillon, Ph.D., Head, WHO Collaborating Center for the Classification of Diseases in French and Eric Jougla, Ph.D., Service d'Information sur les Causes Medicales de Deces-SC8, Institut National de la Sante et de la Recherche Medicale–INSERM

Background

In April 1993, at the Heads of WHO Collaborating Centers meeting in Washington, the importance of multiple cause analysis of morbidity and mortality was emphasized, and the organization of an *ad hoc* meeting on this topic was recommended. This meeting was held in London in 1994. Various studies analyzing multiple cause were presented, and, in its final report, the meeting encouraged the publication of routine tabulations and suggested that international comparisons be undertaken.

At the next Center Heads meeting, held in Caracas in October 1994, the French Center proposed to coordinate a collaborative international study on multiple cause, based on methodology presented at the London meeting. This study aimed to improve international comparisons of cause-of-death data using multiple cause analyses, by defining a set of tabulations for routine calculation. These tables would be published in addition to the standard underlying cause data. The study was accepted at the Caracas meeting and the French Center agreed to prepare and circulate a protocol.

Participating countries

The protocol was sent to all Collaborating Centers for the Classification of Diseases in January 1995. It specified the objectives of the study and described the required information. Participating countries were required to send the data by June 1995 so that a preliminary report could be made to the Center Heads' meeting in Canberra in October 1995. Ten countries replied to this mailing, of which seven countries were interested in collaborating on the study. Five countries were able to send the required data by the June 1995 deadline. Data were requested in tabular form for recent years of death. Table 1 shows some characteristics of the information provided by the participating countries.

Table 1. Characteristics of the Data

Country	Data Year	Coding	Observations
Brazil	1994	Automatic	Data refer to the state of S. Paulo (first semester, 1994)
France	1992	Manual	
Latvia	1992	Manual	Data on imprecise causes not available
Russia	1993	Manual	Sample of 1173 deaths. Data on total mentions and violent deaths not available
Sweden	1993	Automatic	

Characteristics of the data provided by the participating countries

Proposition of routine tables

Of the tables described previously, we have selected seven to propose for routine publication. These are intended to provide information on cause of death, in addition to that furnished by the standard tabulations of underlying cause. They describe five aspects of multiple cause: reported causes, coded causes, imprecise causes, ratio of reported causes to underlying cause, and distribution of Nature of Injury (N codes). These tables, presented in the following section, may constitute a consistent source of information from which multiple cause analysis and international comparisons can be performed. The final section presents possible extensions of these tables.

Routine tables

Each table incorporates the data of the five participating countries. Tables 2 and 3 describe the conditions reported on the death certificate. Table 4 is identical to table 3 except coded conditions are described rather than reported conditions. Comparisons between these two tables provide information about the coding stage. Tables 5 and 6 present figures on imprecise causes. They are not strictly multiple cause tables, but they convey information that is not available from routine tabulations of underlying cause and are essential for multiple cause analyses. Table 7 is the most important as it presents, by cause of death, the ratio of total mention of a given cause to mention of it as an underlying cause. Finally, table 8 gives the distribution of N codes for violent deaths.

Table 2 presents the mean number of causes reported by the certifier on Part I or Part II of the death certificate. Part I refers to the diseases leading directly to death and Part II refers to other significant conditions contributing to death. All the conditions are counted as they are reported by the physician (e.g., if diabetic coma is reported, one condition is counted; if coma and diabetes are reported separately, two conditions are counted).

Table 2. Mean Number of Conditions Reported on the Death Certificate, by Part

Part of Certificate	Country				
	Brazil	France	Latvia	Russia	Sweden
Part I	2.4	2.5	2.0	2.5	2.0
Part II	1.3	0.5	0.2	0.3	1.0
Total	3.7	3.0	2.2	2.8	2.9

Table 3 presents the mean number of conditions by age and sex, regardless of which part of the certificate the condition is reported.

Table 3. Mean Number of Conditions Reported, by Age and Sex

Age	Sex	Country				
		Brazil	France	Latvia	Russia	Sweden
0-44	M	3.6	2.6	2.1	2.2	2.9
	F	3.8	2.3	2.0	2.6	2.9
45-64	M	3.7	3.0	2.2	2.7	2.8
	F	3.8	3.2	2.2	2.7	2.8
65 & over	M	3.7	3.2	2.3	3.0	3.0
	F	3.8	3.0	2.3	2.9	2.9
Total	M	3.6	3.1	2.2	2.8	3.0
	F	3.8	3.0	2.2	2.9	2.9
Total		**3.7**	**3.0**	**2.2**	**2.8**	**2.9**

Table 4 shows the mean number of conditions coded on the death record.

Table 4. Mean Number of Conditions Reported, by Age and Sex

Age	Sex	Country				
		Brazil	France	Latvia	Russia	Sweden
0-44	M	2.6	1.9	2.1	2.0	2.9
	F	2.6	1.9	2.0	2.3	2.9
45-64	M	2.7	2.1	2.2	2.4	2.8
	F	2.8	2.0	2.2	2.5	2.8
65 & over	M	2.9	2.1	2.3	2.7	3.0
	F	2.9	2.0	2.3	2.6	2.9
Total	M	2.7	2.1	2.2	2.5	3.0
	F	2.8	2.0	2.2	2.6	2.9
Total		**2.8**	**2.0**	**2.2**	**2.5**	**2.9**

Table 5 gives the percentage of deaths for which the only condition coded on the death record is an imprecise one, that is, is coded 427.5 (cardiac arrest), 799.1 (respiratory failure), or 799.9 (unknown or unspecified cause).

Table 5. Proportion of Death Certificates With a Single Imprecise Cause

ICD-9 Code	Country percent				
	Brazil	France	Latvia(*)	Russia	Sweden
427.5	0.0	0.7	-	0.0	0.1
799.1	0.5	1.2	-	0.0	0.0
799.9	2.9	1.7	-	1.7	0.3
Total imprecise	**3.4**	**3.5**	**-**	**1.7**	**0.3**

(*) Not available

Table 6 gives the percentage of deaths for which only an unknown or unspecified cause is coded (ICD–9, 799.9) by age.

Table 6. Proportion of Death Certificates With a Single Unknown or Unspecified Cause, by Age

Age	Country percent				
	Brazil	France	Latvia (*)	Russia	Sweden
0-44	2.9	6.9	-	4.0	1.3
45-64	3.4	2.1	-	2.6	0.8
65 & over	2.6	1.0	-	0.9	0.1
All Ages	**2.9**	**1.7**	**-**	**1.7**	**0.3**

(*) Not available

Table 7 (page 7-6) lists, for a given cause, the ratio of the number of deaths for which the cause of death is coded to the total number of deaths for which it is coded as the underlying cause. The list of 63 cause-of-death groups is in the appendix (page 7-9). When the ratio is equal to 1, the cause of death is always coded as the underlying cause. As the ratio increases, the cause of death is selected less often as the underlying cause. For instance, in Brazil, the ratio of Infectious diseases (cause-of-death group 1) is 2.9, indicating that the number of Infectious diseases nearly triples when all causes reported on the death certificate are considered.

Table 8 (page 7-8) gives the percent distribution of deaths by Nature of injury and poisoning (N codes) for a short list of 17 categories. All the violent deaths (i.e., all deaths for which the underlying cause is an E code) are taken into account in this table. For each violent death, all the N codes mentioned are counted. The percentage is thus computed from the number of N codes (which can be greater than the total number of violent deaths). For instance, in Sweden, Fracture of the skull (N800-N804) is responsible for 6.8 percent of violent deaths.

Proposed extensions of the routine tables

Additional information beyond that in the routine tables presented in the preceding section has been envisaged. We propose the following modifications:

- Table 6: extend the imprecise conditions 799.1 (respiratory failure) and 427.5(cardiac arrest), as in table 5.

- Table 7: present by sex.

- Table 8: present as a cross-tabulation between the underlying cause (E code) and the nature of injury and poisoning (N code) for violent deaths. The following short list could be used for E causes:

  1. E810–E819, E826–E829
  2. E850–E858, E860–E869
  3. E870–E879, E930–E949
  4. E880–E888
  5. E890–E899, E910
  6. E800–E807, E820–E825, E830–E848, E900–E909, E911–E928.0–8, E929.0–8
  7. E928.9, E929.9
  8. E950–E959
  9. E960–E969
  10. E970–E978
  11. E980–E989
  12. E990–E999

Table 7. Ratio of Reported Causes to Underlying Cause

Group	Country				
	Brazil	France	Latvia	Russia (*)	Sweden
1	**2.9**	**1.8**	**1.8**	**-**	**3.4**
2	2.0	1.7	1.1	-	2.5
3	1.0	1.1	1.0	-	1.3
4	1.0	1.1	-	-	1.7
5	1.4	1.9	1.2	-	2.0
6	**1.1**	**1.1**	**1.0**	**-**	**1.8**
7	1.2	1.1	1.0	-	1.8
8	1.1	1.1	1.0	-	1.3
9	1.1	1.1	1.0	-	1.1
10	1.1	1.1	1.0	-	1.1
11	1.1	1.1	1.0	-	1.2
12	1.1	1.1	1.0	-	1.2
13	1.1	1.0	-	-	1.1
14	1.1	1.0	1.0	-	1.1
15	1.0	1.0	1.0	-	1.1
16	1.4	1.1	1.2	-	1.3
17	1.1	1.1	1.0	-	1.3
18	1.1	1.1	1.1	-	1.1
19	1.1	1.1	1.0	-	1.3
20	1.1	1.1	1.0	-	1.1
21	1.2	1.2	1.1	-	1.4
22	1.1	1.1	-	-	1.3
23	1.1	1.1	1.0	-	1.2
24	1.2	1.2	1.1	-	1.3
25	**2.3**	**3.2**	**3.9**	**-**	**5.2**
26	2.4	3.1	3.8	-	5.2
27	**5.2**	**3.3**	**2.0**	**-**	**6.0**
28	**3.4**	**3.3**	**2.5**	**-**	**3.4**
29	2.8	3.6	1.5	-	4.5
30	-	1.5	3.0	-	1.8
31	**3.4**	**2.2**	**27.1**	**-**	**3.8**
32	1.7	2.3	1.0	-	1.7
33	3.3	2.3	-	-	3.8
34	**1.4**	**1.3**	**1.2**	**-**	**2.3**
35	1.1	1.3	1.2	-	1.5
36	1.4	1.4	1.1	-	1.7
37	**2.7**	**1.8**	**2.4**	**-**	**2.9**
38	2.3	1.7	1.2	-	2.5
39	1.5	1.2	3.0	-	1.3
40	1.6	1.8	3.2	-	2.7

Table 7. Ratio of Reported Causes to Underlying Cause, Con.

Group	Country				
	Brazil	France	Latvia	Russia (*)	Sweden
41	**1.9**	**1.7**	**1.7**	**-**	**2.8**
42	1.5	1.6	1.4	-	2.4
43	2.7	1.4	1.7	-	1.6
44	25.1	1.7	1.7	-	2.4
45	**3.9**	**3.1**	**2.5**	**-**	**4.5**
46	4.1	3.7	6.3	-	5.2
47	2.5	2.1	2.3	-	2.5
48	**-**	**1.2**	**1.0**	**-**	**1.2**
49	**6.6**	**3.3**	**2.7**	**-**	**6.3**
50	**3.5**	**2.5**	**3.7**	**-**	**5.0**
51	**1.4**	**1.5**	**1.1**	**-**	**1.9**
52	1.2	1.3	-	-	2.2
53	**1.3**	**1.3**	**1.0**	**-**	**1.7**
54	**7.9**	**16.0**	**2.8**	**-**	**12.9**
55	1.3	1.0	-	-	2.7
56	1.0	1.0	-	-	2.5
57	**1.2**	**1.0**	**1.1**	**-**	**4.0**
58	1.3	1.0	1.2	-	5.6
59	1.0	1.0	1.0	-	1.0
60	1.8	1.0	1.0	-	2.2
61	1.0	1.0	1.0	-	1.2
62	1.0	1.0	1.0	-	1.2
63	1.0	1.0	1.0	-	1.0
Total	2.1	2.3	1.5	-	2.7

(*) Not available

Table 8. Percent Distribution of Nature of Injury and Poisoning (N codes) for Violent Deaths

Cause N	Country percent				
	Brazil	France	Latvia(*)	Russia(**)	Sweden
Cause N not specified	1.4	3.8	0.0	-	0.1
800-804	2.0	8.4	13.2	-	6.8
805-809	0.9	2.8	2.8	-	3.5
820	0.4	12.4	0.0	-	10.0
810-819, 821-829	0.9	3.2	2.3	-	3.5
850-854	27.5	11.9	16.7	-	11.1
860-869	23.6	20.0	12.7	-	8.3
870-897	0.8	0.7	1.4	-	0.8
905-908	0.0	0.4	0.0	-	2.0
930-939	0.4	7.5	1.4	-	2.4
940-949	1.2	1.5	2.2	-	1.3
950-957	2.4	0.4	0.0	-	0.4
960-979	0.1	3.9	0.0	-	13.4
980-989	0.6	1.7	11.4	-	8.4
990-995	8.8	17.2	0.0	-	14.6
996-999	1.9	3.2	0.0	-	6.8
other causes coded	27.1	1.1	35.9	-	6.7
Total 800-999	**100.0**	**100.0**	**100.0**	**-**	**100.0**

(*) 0.0 means that detailed figures cannot be provided by the coding system
(**) Not available

Appendix. List of 63 groups of diseases

This list is used for tables 5 and 6. For each cause-of-death group, the title and ICD–9 codes are shown.

1.	Infectious and parasitic diseases	1–139
2.	Tuberculosis	10–18
3.	Meningococcal infection	36
4.	AIDS	42–44*
5.	Hepatitis	70
6.	Neoplasms	140–239
7.	Malignant neoplasms	140–208
8.	Malignant neoplasm of lip, oral cavity, pharynx	140–149
9.	Malignant neoplasm of oesophagus	150
10.	Malignant neoplasm of stomach	151
11.	Malignant neoplasm of colon	153
12.	Malignant neoplasm of rectum and anus	154
13.	Malignant neoplasm of liver	155
14.	Malignant neoplasm of rest of digestive organs and periton.	152,156–159
15.	Malignant neoplasm of trachea/bronchus/lung	162
16.	Malignant neoplasm of skin incl. melanoma	172–173
17.	Malignant neoplasm of female breast	174
18.	Malignant neoplasm of cervix uteri	180
19.	Malignant neoplasm of other parts of uterus	179,182
20.	Malignant neoplasm of ovary	183.0
21.	Malignant neoplasm of prostate	185
22.	Malignant neoplasm of bladder	188
23.	Malignant neoplasm of kidney	189.0
24.	Malignant neoplasm of lymph./haematopoietic tissue	200–208
25.	Endocrine/nutritional/metabolic diseases/ immunity disorders	240–279
26.	Diabetes	250
27.	Diseases of the blood and blood-forming organs	280–289
28.	Mental and behavioral disorders	290–319
29.	Alcoholic psychosis, chronic alcohol abuse	291,303
30.	Drug dependence	304
31.	Diseases of the nervous system and the sense organs	320–389
32.	Meningitis	320–322
33.	Parkinson’s disease	332
34.	Diseases of the circulatory system	390–459
35.	Ischaemic heart diseases	410–414
36.	Cerebrovascular diseases	430–448
37.	Diseases of the respiratory system	460–519
38.	Pneumonia	480–486
39.	Influenza	487
40.	Chronic lower respiratory diseases	490–494,496
41.	Diseases of the digestive system	520–579
42.	Ulcer of stomach, duodenum and jejunum	531–534
43.	Chronic liver disease, alcoholic	571.0–571.3
44.	Chronic liver disease without mention of alcohol	571.4–571.9
45.	Diseases of the genitourinary system	580–629
46.	Nephritis, nephrotic syndrome, renal failure	580–589
47.	Hyperplasia of prostate	600
48.	Complications of pregnancy, childbirth and puerperiurn	630–676

49.	Diseases of the skin and subcutaneous tissue	680–709
50.	Diseases of the musculoskeletal system/ connective tissue	710–739
51.	Congenital malformations and chromosomal abnormalities	740–759
52.	Congenital malformations of the circulatory system	745–747
53.	Certain conditions originating in the perinatal period	760–779
54.	Symptoms, signs, abnormal findings, ill-defined causes	780–799
55.	Sudden death, cause unknown	798
56.	Other unknown and unspecified cause	799.9
57.	External causes of injury and poisoning	E800–E999
58.	Accidents	E800–E929
59.	Traffic accidents	E810–E819, E826–E829
60.	Accidental falls	E880–E888
61.	Accidental drowning and submersion	E910
62.	Suicide and intentional self-harm	E950–E959
63.	Assault	E960–E969

* Countries using a different ICD code for AIDS must include AIDS in category 4, ICD–9, 42–44.

Comments

MR. ISRAEL: First, I would like to congratulate both of our colleagues, who have done a lot of very serious thinking and work in the area of multiple cause analysis, and over the years have made some significant contributions. But, I am a little unsure, or perhaps even concerned, about the interpretation of the absolute value of the mean number of conditions reported, because I think it is a function to a very large extent of the WHO instructions and the rules on how to properly complete a death certificate.

So, while there is not a ceiling, and while we also realize that physicians often don't follow the instructions anyway, there is a dampening or cushioning effect on the top that pushes down the number of conditions reported on a death certificate. We do not offer the certifier a blank piece of paper and ask for all the things that he or she thinks might have been significant at the time of death. Instead, we fall back on the concept that Dr. Pavillon and Dr. Jougla have called a "properly completed certificate," and we ask for a sequence and suggest that at least on the lowest line there only be one condition reported for it to be a properly completed certificate. We list only three or four lines in part one, and most countries only provide a single line in part two. Of course, one can write several conditions on that line.

The point I am making is that I am not quite sure of the importance of the absolute value of the mean number of reported conditions. Perhaps there is some value in relative comparisons between countries, given that the countries are following the same constraints that I have just mentioned. I would appreciate your observations on that. Thank you.

DR. JOUGLA: I agree with you that a good indicator is the proportion of causes on the certificate. The number of conditions is not an indicator of good certification, but it does provide important information before undertaking multiple cause analysis between countries. The number of conditions is not a good indicator and in certain cases it may be difficult to solve this type of certification. I worked on a large AIDS project. During one year, we drew a large sample of about 1,200 certificates with AIDS and had to go back to the physicians with questions. The result was that the next year we had certificates with a lot of information, and it was very difficult to certify them.

DR. JOZAN: First of all, may I congratulate you on this very instructive and interesting study. I would like to ask whether you have been in a position to make a distinction between juvenile and senile diabetes, because these diseases are definitely very different, with different outcomes and complications.

Another question is whether you could analyze the validity and reliability of the data on the death certificate by whether the information came from autopsies, from physicians in the hospitals, or from family physicians in the various countries. Thank you.

DR. JOUGLA: For the first question concerning juvenile or senile diabetes, we are most likely to see senile diabetes. I agree with you, it was a problem because those types of certificates are quite difficult to code and classify. When we present the example on diabetes, it tends to make people very pessimistic about the statistical cause of death. But I think it is a particular problem because it concerns all countries.

In France, for instance, we have very few autopsies. In addition, we do not have reliable information on the certificates. We have no information on the people who are certifying the certificate.

MRS. ROBERTS: Are those results based on samples from those countries, or what is it that the numbers refer to?

My second question is, have you tried to identify how much of the work you have done or that could conceivably be incorporated as a standard output of the software that we are discussing here?

DR. PAVILLON: For the first question, we mainly based our results on all death certificates, except in certain cases. In France, for instance, the results are based on representative samples of certificates. For the certification level, the results were based on a representative sample of the death certificates.

The second question seems to be two-fold. We did not use a specific software program to produce the multiple cause analysis. It was just programed in certain databases. It could be possible to think of a software program, but before this step we have to think of the type of tables we would like to have. When the tables are defined, it will be possible to develop the software.

There is a second aspect to your question, what is the relation with multiple cause analysis and automatic coding? I think that multiple cause analysis will help to analyze discrepancies or differences in the decision tables, in the algorithm of automatic coding, and in the multiple cause coding between countries. Multiple cause analysis is interesting at the international level when making comparisons between countries.

DR. ROSENBERG: Some of the software (ACME, TRANSAX, MICAR) does have the capability of counting the number of times for which certain selection and modification rules are invoked, so that you can get some of the counts directly.

The United States now routinely produces multiple cause tables for our data tape documentation. We use these for responding to many public inquiries for the number of times a condition is mentioned. There is a great deal of interest in having this information.

Regarding diabetes, in the United States, we changed the death certificate in 1989. We added a fourth line to Part I on our standard certificate. As a consequence, in 1989 the number of diabetes deaths in the United States increased by about 14 percent. So diabetes is very sensitive to the formatting of the death certificate. Subsequently, the trend continued in a normal fashion, but from 1987 to 1989 the rate went up rapidly. We were able to determine that this is an artifact by looking at multiple cause data, where the condition counts for diabetes did not change from one year to the next during this period. It was merely a shift in the placement of the condition on the death certificate, from Part I to Part II.

I also want to mention that the United States will soon produce the first CD-ROM with multiple cause data. It will have all two million deaths with all conditions mentioned. Instead of costing over $1,000 for a data tape, it will cost between $30 and $60 for one CD-ROM. So, the CD-ROM will introduce new possibilities for analysis. The CD-ROM will include a tabulation program, to make the data more accessible, so that we can do some studies along the lines that Dr. Pavillon and Dr. Jougla are recommending.

MR. JOHANSSON: I just have a comment on the relationship between multiple cause coding and automatic coding. In Sweden we have had manual multiple cause coding since 1961, when we analyzed our data and saw that the ratio between the underlying cause and the number of mentions varied extremely from one year to the next, and we realized that the manual coding was not consistent. That was one reason why we decided to use automatic coding instead of manual coding.

DR. SANTO: Relating to the tables, in Brazil we have produced a software program that tabulates multiple cause data, using output from ACME files. It is a complete matrix from which all the traditional tables can be drawn. It is very easy to use. We can later demonstrate to those who are interested in it.

DR. MORIYAMA: Ever since we started coding causes of death, at least publishing single cause of death per individual, we have been concerned about the loss of information. So it has been suggested we ought to code everything on the death certificate. We are now at a point where we have done that, and I am glad to see that the authors here have gone beyond that to show the anatomy of cause-of-death analysis. I think it was an excellent presentation.

But as Mr. Israel pointed out, I think the greatest concern to me is the difficulty of looking at the data and picking out information. That is, we talk about multiple causes; what do we mean by multiple causes? We talk about count of conditions; what do we mean by that? As you all know, the physicians report almost anything that you can think of on the death certificates. You can have signs, symptoms, and disease reported, the same disease reported in a different way, and manifestations of diseases. There is a lot of duplication and a lot of noise in the information.

So when a country has 2.6 causes per certificate, it somehow leaves me puzzled, because what are we counting? Rather than giving a count of conditions, if we talk about the count of deaths with particular diseases I think we will have a lot better information as far as multiple cause analysis is concerned. Thank you.

DR. PAVILLON: In order to count the different conditions, we decided to count them as they were reported. For instance, if a diabetic coma was reported, we counted it as one condition, diabetes ending with a coma would count as two conditions.

Secondly, what conditions were counted? We counted the conditions that can be coded in ICD–9. If it could not be coded in ICD–9 but can now be coded in ICD–10, we did not count it as a condition.

MR. ROTHWELL: I would like to comment on Dr. Moriyama's point. I think that if we are really going to use multiple cause, it has to be looked at as a scenario of a combination of diseases, a chronic set of diseases. Possibly we need to take a look at the underlying cause with associated diseases and say this is a scenario that we want to publish. That might also improve the information that we are getting on multiple causes, because physicians and certifiers will then know that we are using this information to depict a syndrome, a chronic disease syndrome.

What bothers me about international multiple cause analysis is there can be confusion about the different practices that take place in different countries. I am wondering if you have looked at, just within France, mentions of hypertension or atherosclerosis or diabetes, and looked at it regionally to see if you have great variations in certifiers' responses on the certificates.

DR. JOUGLA: We have not done that type of geographic comparison. That is not the sort of study we do routinely. We are involved especially in querying on case by case, but do not have time to do all the things we would like to. It would be good to have this information, but we do not routinely do geographic comparisons.

MR. ROTHWELL: As we mentioned, in the United States, the States are very different. If you look at multiple cause relationships between States, you get some unusual differences.

DR. JOUGLA: Perhaps you also have the motivation because your system is decentralized. In France, we have a very centralized system.

MR. ISRAEL: I would like to come back to a question that I thought Mrs. Roberts was asking. I would like to add a little bit to that. The relationship of multiple cause data and multiple cause analysis and the topic of this collaborative effort on automation is a very interesting one.

There were studies done on multiple cause-of-death analysis before countries began to automate their coding systems. But on the other hand, when countries began to look at methods of doing automated coding, it became obvious that the underlying cause of death is a summation of a number of conditions written on a death certificate that a human coder assimilates.

It was obvious in the beginning of automation that we had to take the individual elemental units of information on the death certificate and do something with them in order to enable the computer to give us automatically an underlying cause of death.

In doing this automation of underlying cause of death, we had a by-product. We had to have these individual conditions already coded and fed to the computer. So it greatly enhanced the ability to look at multiple cause data because we were beginning to try to automate the underlying cause of death principle. So I think that is part of the relationship between automation and multiple cause analysis.

Internationally, we are still ascribing the underlying cause of death primarily in our statistical tabulations, but we have a great deal of work to do in terms of the methodology of capturing the multiple causes of death, presenting, and analyzing them.

This question of what is a condition is not uniformly resolved around the world in the different countries that are looking at multiple cause data. We don't have uniformity. A number of years ago at a WHO meeting on multiple cause analysis, it was mutually agreed by all present that an international set of guidelines or rules would be difficult to develop, and countries would be encouraged to develop their own systems and to meet periodically and compare results to see if it were not possible to produce some international guidelines on causes to count, how to count them, and how to present them. We are still doing that, and that is partially what we are doing this morning. There is a long history of the relationship between attempts to automate the underlying

cause of death process and the production of multiple causes.

DR. PAVILLON: What you said I think is very important. This is the goal we have to keep in mind while doing multiple cause analysis at the international level, to define all these basic aspects of multiple cause and coding.

DR. PEREZ: I have two questions. First, which rules did you use to code multiple cause of death? And are those rules the same in the countries that you compared? The second one is, how did you measure the strength of association between diabetes and other conditions? Thank you.

DR. PAVILLON: For the first question, we did not use specific rules for coding multiple cause. Each country participating in the study had its own way of producing multiple cause. In France, multiple cause is produced manually. In Sweden, they use TRANSAX, an automated coding system. In Brazil, also, I think, and so on. Each country used its own system. We did the analysis. I hope that we will be able to develop some guidelines for producing multiple cause analysis, but I think that we have to begin with the analysis of the different discrepancies between countries and to end with some guidelines.

DR. JOUGLA: Concerning the measure of the strength of the association, it is computed by the comparison of an observed number of death certificates with the two specific diseases, to an expected number computed on the independence of the two diseases.

DR. PAVILLON: You can have the details in the publication, or in another publication of our own, which takes the details of the calculus. If you want, we can give you the references.

Implementing Automated Coding in England and Wales: How It Affected Mortality Statistics

Cleone Rooney, M.D., Office for National Statistics, England

I would like to talk to you about what happened when we started using automated coding in England and Wales, changing over from our previous manual coding system. I am going give you a quick outline of what our system is, how it works, and how it differs from the system in the United States.

ACCS

Our automated cause coding system (ACCS) is built from modules developed by the National Center for Health Statistics (NCHS)–MICAR, ACME, and TRANSAX– and some additional modules developed in the Office for Population Census and Surveys (OPCS)–TRACER, EXTRACT and COLLECT. (Figure 1 is a diagram of how these fit together.)

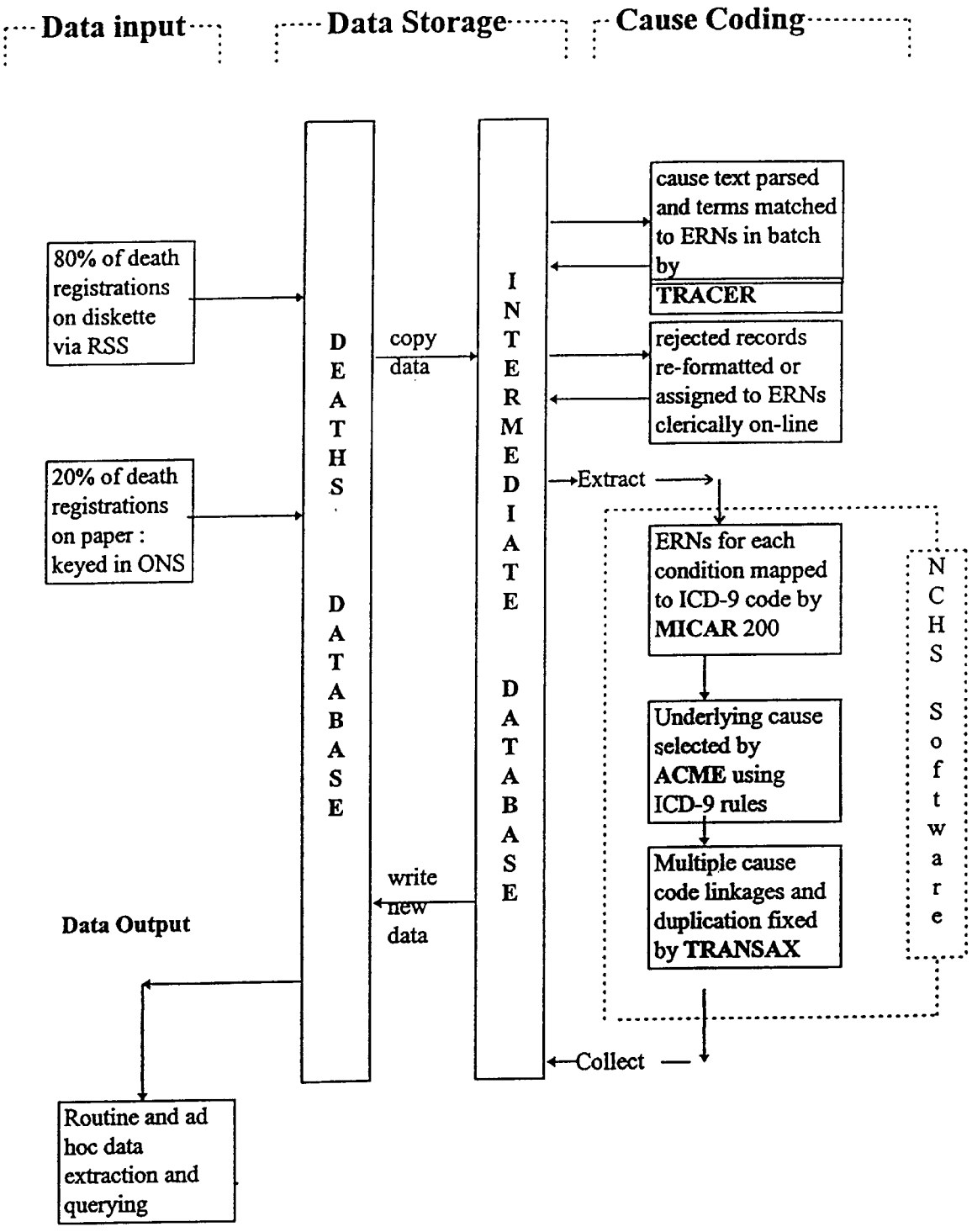
In England and Wales, about 580,000 deaths are registered each year. Most of those are certified by doctors, but about 25 percent are certified by coroners, and a small proportion of those, 20,000 a year, involve a coroner's inquest. Coroners are required to hold a legal inquest into all deaths from accidents, violence, and poisoning–external causes of death.

For most deaths, the doctor completes the paper death certificate. He gives the certificate to the family. They take it to the Registrar of Births, Marriages, and Deaths, who takes all the information about the cause of death exactly from the doctor's certificate: if the doctor has spelled the causes wrongly, the registrar has to spell them wrongly as well. He/she is not allowed to change anything. The registrar gets all the other information, date of birth, occupation, whether the person was married, etc., from the family. The doctor's certificate just provides the cause of death and the date of death.

At that point, the Registrar enters this information directly into the computer with the family there, using what we call registration system software. At the end of every week, the Registrar sends us a floppy disk with all the deaths that have been registered that week. At least 90 percent of all the deaths in England and Wales come to the Office for National Statistics (ONS) on disk, using registration system software. Ten percent still come on paper, and those are entered at the ONS using an almost identical software system. The 10 percent that we enter come largely from offices in very rural areas, with very few events. There may only be 2 deaths a week, and it is just not worthwhile to have a computer or to learn how to use one.

Data is then loaded directly into the electronic deaths database. From there, the information on cause (which is just electronic text and positional codes) and the necessary bits of additional information, like age and sex, are copied across to an intermediate database, and go into the automated coding system. Eighty percent of all deaths go through the entire automation process. They are completely processed electronically, and have not been seen or dealt with by a coder at all. Twenty percent need some intervention.

Figure 1. The Automated Cause Coding System



TRACER is one of the modules developed at ONS . Basically, it takes in the cause of death text, splits the text into individual cause terms based on which line the terms are on or whether there is a comma or a conjunction that splits the terms into more than one cause, matches those to entity reference numbers from the MICAR dictionary, and retains the entity reference number and the position on the certificate.

Most deaths are input, cause coded, and output automatically, with no intervention. With TRACER, the coders can intervene if there are problems, such as if TRACER cannot make a split because there is not a conjunction, if the certifier has written several terms in one long stream, or if there are misspellings that the dictionary does not recognize. (The really common misspellings, such as anaemias and haemorrhages with the As and Es reversed, are in the dictionary.) Unusual misspellings can be dealt with by coders using TRACER, who correct the spelling so that it will match a dictionary entry. If there is no exact match for the term on the death certificate in the dictionary or the thesaurus, the coders can choose the entity reference number which is closest to the phrase, along with the words which are closest to the words the doctor wrote. The coders can actually assign an entity reference number, correct the spelling, or format the record so that it matches.

From there, the records go through another module–EXTRACT. This moves the record from the intermediate database to MICAR, from where they are passed to ACME and TRANSAX.

After TRACER, the records are in a closed system. Records that are rejected anywhere further along cannot be fixed and still be automatically coded. If they are rejected, they have to be coded on line. That results in only ICD codes, not entity codes. However, the majority of records actually go right through without being rejected.

About 20 percent of the records are rejected with MICAR. As expected, these records are for surgical operations, drugs, terms that are not in the dictionary, and a few terms for which there is no exact match. Very few records are rejected with ACME, perhaps five or six records a month. With these rejects, the ACME decision tables usually cannot choose an underlying cause from the variety of ICD codes for the record.

When the underlying and multiple-cause coding are finished, the information is copied back into the intermediate database, which writes a new record for that death to the death database, and all of the outputs can be processed, both the routine and special *ad hoc* querying, from the death database. At the end of that process, the following is stored:

- underlying cause
- unlinked multiple cause
- linked multiple cause
- the position of every cause on the certificate
- ICD codes
- entity codes (if the record was automatically coded)
- original cause text

If the death had to be coded manually, then there are not any entity codes, which means that for those records, there is less detail. The entity codes that Donna Glenn showed are a lot more detailed than some of the ICD

codes, and you can pick up very particular diseases, which are often lost in the other and unspecified bits of the ICD codes.

All of the original text is stored exactly as written by the doctor. That enables a lot of detailed searching because the text can be searched for particular words. For example, mesotheliomas can be sought because in the ICD coding they get coded to the site, which is either a pleural tumor, a peritoneal tumor, or an unspecified site. It is difficult to count them reliably using ICD codes, but they can be found more reliably by searching the text for the word “mesothelioma.”

The verdict is stored on those 20,000 deaths, which are certified after an inquest. These deaths have a legal verdict (suicide, accident, homicide, industrial disease, open verdict, etc.), which corresponds fairly well to the ‘manner of death’ in the United States. We have found it very useful in checking the accuracy of cause-of-death coding by cross tabulating the verdict against the underlying cause.

Information on duration is also stored, if it was on the certificate. Duration is reported on only about a third of the certificates. I think that varies a lot from country to country. Some countries get 2 or 3 percent, while others get 90 percent. Some countries do not even have a space for duration on the certificate.

There have been benefits from automating. Automated coding gives better uniformity. If the deaths are coded automatically, then one does not get different coders doing different things or the same coder doing different things on different days. So it does make coding more uniform. It increases international comparability, and provides multiple cause coding. Cost savings may also be a benefit one day, although I am not sure we have saved any money yet.

We have seen some disadvantages as well. We started using this system for all deaths from the first of January 1993. We did not bridge code the change, which is a shame. Changing from manual to automated coding is a very big change in how death data is produced. Even though we stayed with ICD–9, it really was an enormous change. It certainly was the biggest change since the transition from ICD–8 to 9 and was probably a bigger change. We did not bridge code it, so we spent a lot of time trying to work out exactly what happened to our statistics.

There were some things that we expected to change–particularly how ICD–9, rule 3,(1) was applied. At the same time, we also stopped querying doctors for more information on vague causes, because we were coding these deaths in arrears. We did not start coding 1993 deaths until about December 1993, and we have found in the past that if you send a letter out more than 3 months after the death, you never get an answer, so there did not seem to be much point.

At that point, we also did not think that our medical querying system was giving us very useful information. We were sending queries out on more than a 1,000 causes of death, and hardly any of the replies changed the underlying cause. So we decided to review the system, and try to rationalize it. We have not actually got around to it yet, but are hoping to in the next

couple of years. (I was very interested to hear that NCHS is building in a querying system within the automated system, which I think we could probably adapt for use in England and Wales.)

There were also some big unexpected changes. The worst was that the automated system dealt very poorly with our external-cause deaths. They just suddenly changed, and it took a lot of work to find out why. We are reliant on other people for the development and updating of the coding software. We do not want to change the software ourselves, because the advantage is the comparability, so we do not want our system to be different. But if there are lots of users of these systems, we are putting a lot of work on one country and one fairly small set of people to do the work of maintaining it for the rest of us.

Selection rule 3

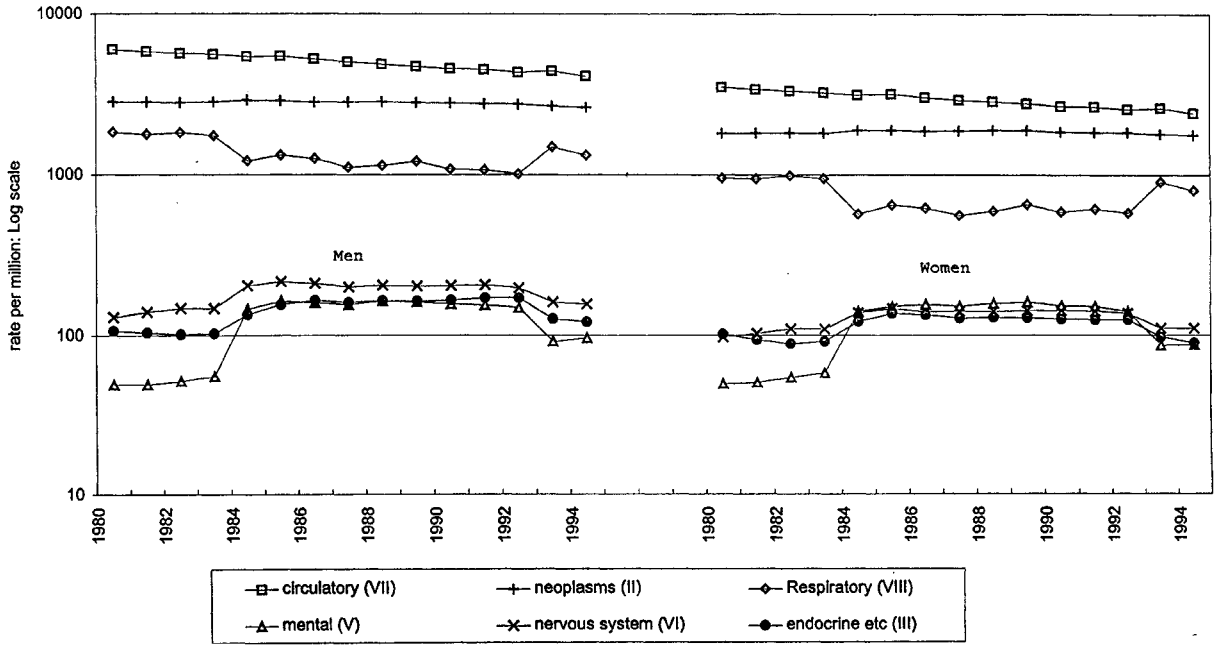
The biggest change that we saw and the biggest effect we expected was due to the different interpretation of rule 3. ICD–9, rule 3, says, "if the condition selected by the general rule and rules 1 and 2 as the underlying cause of death can be considered a direct sequel of a condition mentioned elsewhere on the certificate, select that condition(1)." My office (which used to be called the Office for Population Census and Surveys but is now the Office for National Statistics) changed this rule in 1984 because we had an increasing number of certificates that just said bronchopneumonia in Part I and usually mentioned other important diseases in Part II of the certificate (2). So OPCS broadened the rule to say that if the underlying condition was one of eleven terminal conditions, including pneumonia, deep vein thrombosis, pulmonary embolus, heart failure, liver failure, etc., and if there was any other major condition mentioned anywhere on the certificate, the coder was to select that major condition. You did not have to have any plausible pathological sequence; you just presumed the original underlying cause was a terminal event that could happen to anyone who was already seriously ill.

We did bridge code that change from the strict interpretation of rule 3 to the broader interpretation, so there are detailed tables showing how that changed our statistics in our annual volume of deaths by cause (2). We have a clear idea of what conditions it affected. Obviously, the things that were considered terminal went down, because we were not selecting them any longer.

Most of the conditions that increased were chronic disabling diseases, especially diabetes, dementia, Parkinson's disease, arthritis, rheumatoid arthritis, all sorts of long-term illnesses like that. They are things which our doctors tend to write in Part II, as something that the patients had for 10, 20 or 30 years and have made them unwell, but the doctor does not think of it as the cause that killed them. The doctor thinks of the pneumonia they got in the last week as being the cause of death. Primarily this affects deaths in the elderly, where there are lots of conditions, and it is often quite difficult for the doctor to decide what the real cause was. The conditions that are not affected by this rule change include most cancers and ischemic heart disease (IHD). The effects on causes of deaths at ages under 75 are generally much smaller than in older people. My interpretation of this is that the doctors write these conditions (IHD, cancer, etc.) as being the underlying cause of death. They think of them as causing the death, and the certificate is not completed incorrectly; it is more or less right, so it is easy to pick the right condition.

The graphs of age-standardized death rates by chapter of the ICD from 1980 to 1994 (see figure 2) show which groups of diseases were or were not affected by the change in coding. The top line is circulatory diseases. You can see there is a fairly steady decline throughout that period. There may be a bit of an artifact there, but it is not noticeable. The next line, the next most common cause of death, is cancer. I am not convinced that our rates are really going down yet. (I am sure we could change our coding and make it look as if they were. We can apparently perform magic with respiratory diseases.) The third line is respiratory diseases. It looks as if deaths from respiratory diseases went down about 25 percent in 1984 (this is a log scale), which is quite dramatic, really. The change was really an artifact due to the change in rule 3. It is mostly due to the effect of excluding pneumonia as an underlying cause if there was another major condition on the certificate. We said we were not going to select it as the underlying cause, and the rates went down 25 percent, both sexes. We changed over to the automated coding in 1993, back to the stricter rule in use in the United States, and the death rates went back up.

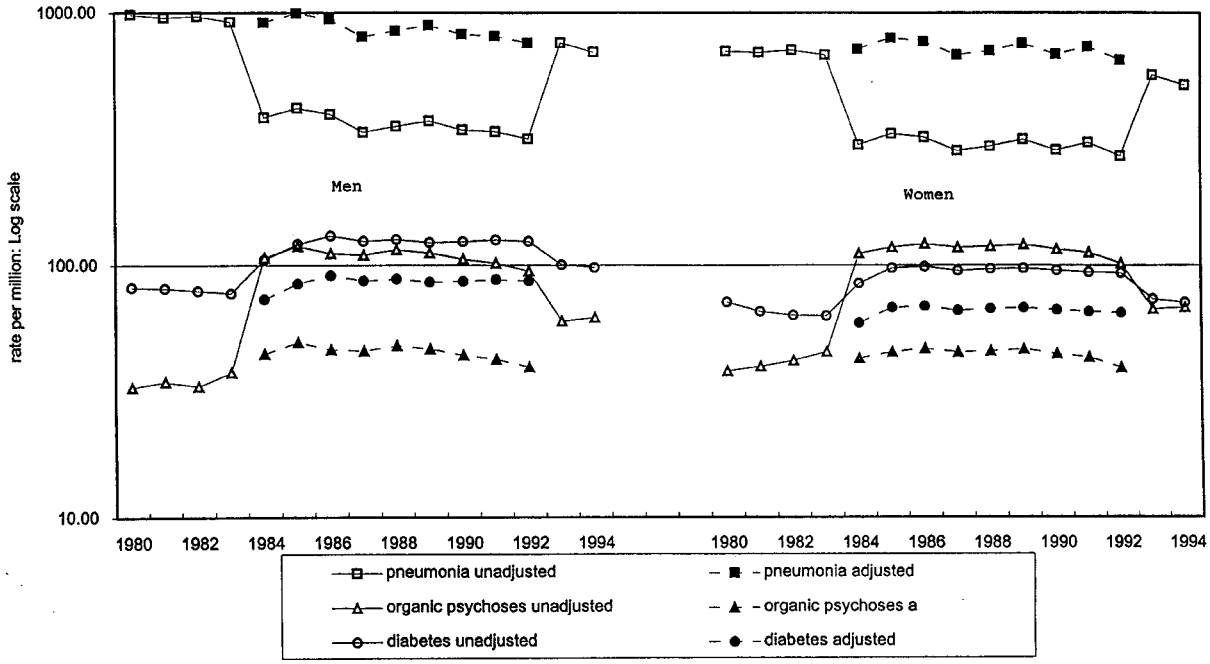
Figure 2. Age-standardized death rates by causes in selected chapters of ICD–9, England and Wales, 1980–1994



The lower lines on the graphs show you what conditions those deaths were being coded to instead. Diseases of the nervous system, musculo-skeletal disorders and endocrine diseases all went up. Most people do not think of mental disorders as being particularly fatal, especially when you accept that depression is hardly ever selected as the underlying cause because of the ICD rules which make you select suicide, even when it is certified as due to depressive illness. Nearly all the deaths coded to this chapter are in fact organic psychoses (ICD–9, 290), which is mostly senile dementia (290.0). The number of deaths in this category went from about 4,000 to about 9,000 in 1984 –a huge difference, and in both sexes. You can see this group came down again

when we began using MICAR and ACME. If you look at some individual diseases, this shows up even more clearly (figure 3). Pneumonia is at the top. Below that are diabetes and dementia. Again, this is a log scale, so these are huge sudden changes in rates. The age-standardized death rate from senile dementia more than doubled in 1984.

Figure 3. Age-standardized death rates by sex, with and without adjustment for rule 3, England and Wales, 1980–1994



I tried looking at what would happen if we took the conversion factors (comparability ratios) from our rule 3 bridge coding in 1984 (2) to see if it could get rid of these artifacts. If you apply the 1984 factors to obtain a corrected pneumonia rate throughout the period to 1992, it pretty much eliminates the whole artifact. If you use these factors to adjust diabetes mortality, there is quite a substantial gap between the 'adjusted' 1992 figure and the 1993 rate. You also find that diabetes went up dramatically in 1984 as well as in 1985 and in 1986. It has increased steadily since then. When it fell in 1993 it settled down between the change in 1984 and the whole 1983-1986 change.

There were other, less documented, changes in mortality coding in OPCS in those years. There was the rule 3 change in 1984, the dramatic one. Then urinary tract infections was added to rule 3 for diabetes in 1985, increasing the rate a bit more. Then, any mention of peripheral vascular disease was linked to diabetes. So reporting a myocardial infarction due to peripheral vascular disease, with diabetes in Part II, was reported as a diabetes death.

If you look at multiple-cause data, which is only available for 1985 (3), 1986 (4), 1993, and 1994 (ONS electronic deaths database), the mentions

of diabetes are absolutely steady; there is no change. This means that the change in underlying cause death rates for diabetes is probably entirely due to changes in how the underlying cause was selected. Looking at organic psychoses (ICD–9, 290–mostly 290.0, which is senile dementia), if I try to adjust the rates, they are fine in 1985 and 1986, but for 1992 there is still a gap of about 40 percent or more between the adjusted figure and the 1993 rate using the automated system. It is absolutely hopeless to try to correct rates in the 1990's using 1984 factors.

I do not think that I understand everything that is going on in the senile dementia data, but I can explain some of it. I know that there has been a big shift from writing 'senile dementia' to writing 'Alzheimer's disease' on the certificates (5). So a lot of these deaths are actually now going into a different chapter, but I still do not really understand why there is such a big gap between 1992 and 1993. Unfortunately, I do not think that there is a way to adjust for this using the available data. Multiple-cause data could be used, but it is more complicated, because these deaths can go to Alzheimer's disease, which is in the diseases of the nervous system chapter. This is a lesson: bridge code any changes in processing the data. Otherwise, you do not have any idea what is really going on in your mortality data.

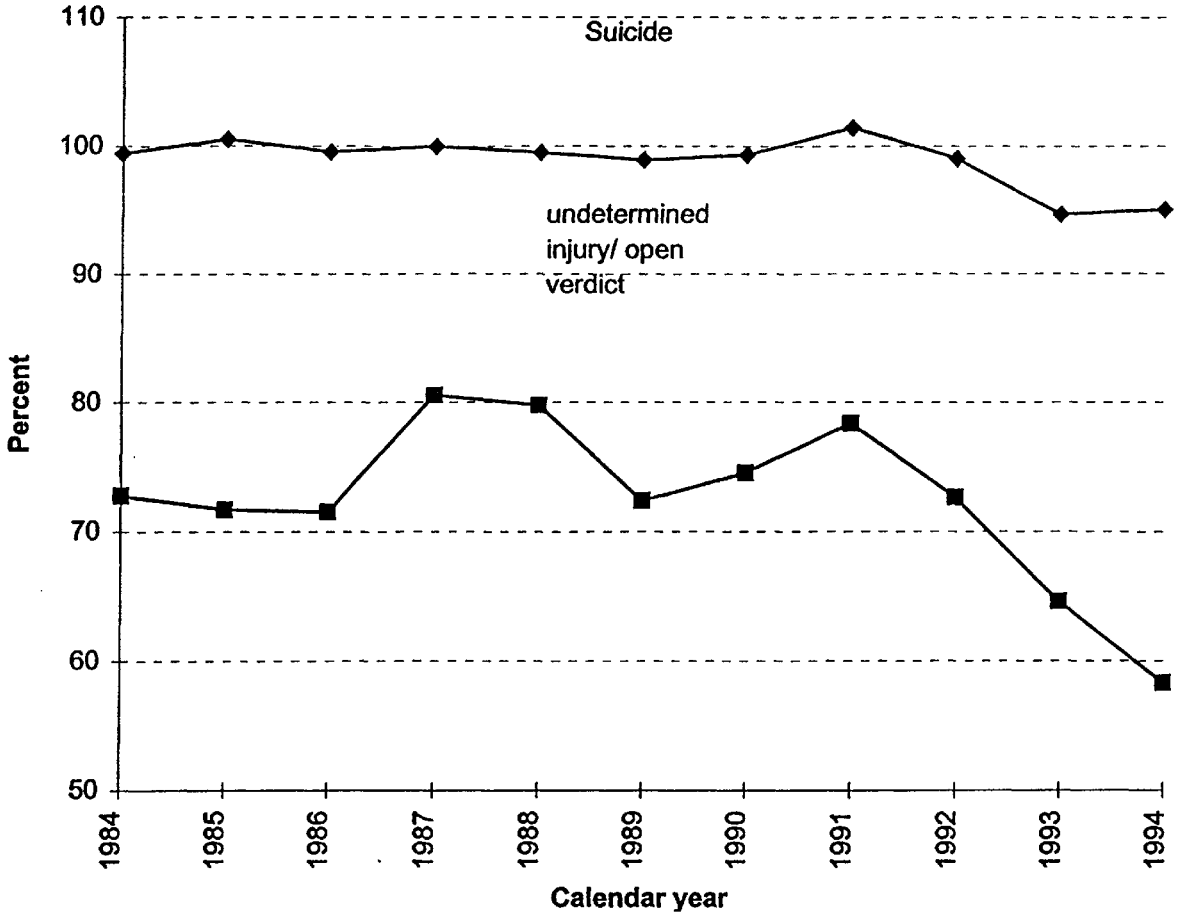
External causes of death

The other big problem we found was in external causes of death. In 1993, we had a sudden, dramatic decline in death rates from external causes as a whole, and particularly from suicides and motor vehicle traffic accidents. We saw a decline of 17 percent in the age-standardized death rate for male suicide. This seemed unlikely to be true. The age-standardized rate had been fairly steady for the past decade, though this concealed decreases in older age groups and increases in those under 45 years old. When we looked at the age-specific data produced by automatic coding, we saw declines in every age group in 1993 and 1994 compared to earlier years. Luckily, we were able to compare our data with an independent source. The Home Office (6) publishes statistics on verdicts from corners' inquests each year. When we compared the numbers of deaths coded to suicide or undetermined injury with the number of coroners' verdicts of suicide or 'open,' we saw a sudden decline in the ONS coded data which was not present in the coroners' data reported to the Home Office (figure 4). Our suicide figures are usually very close to theirs–between 98 and 102 percent (the variation is because of slightly different coverage in time periods and places of death or residence). It is a little bit more difficult when you look at the undetermined deaths. Our deaths coded to undetermined injury are usually between 70 and 80 percent of the Home Office's open verdicts. This is because not all open verdicts are injury deaths. (For example, some may be 'open' as to whether asbestosis was due to the deceased's occupation or not.) Figure 4 shows the position before we corrected the coding of inquest deaths.

For motor vehicle accidents, the picture was similar. The deaths coded by ACCS showed a 20 percent decline between 1992 and 1993. Though these rates had been falling for some years, this was a much larger decline than in previous years. Comparing these data to independent data collected by the Department of Transport on deaths within 30 days of a road accident (7) showed clearly that the ONS data were deficient. We then cross tabulated coroners' verdicts and underlying cause of death for deaths on the ONS mortality database. This showed that there were significant numbers of deaths with a

verdict of suicide, accident, or open which had been coded to a disease category as underlying cause. We have since gone back and re-coded these deaths clerically and published corrected figures for 1993 and 1994 (8).

Figure 4. Uncorrected numbers of deaths coded to Suicide (E950–E959) and injury of undetermined intent (E980–E988, except adjourned inquests) as a percentage of Suicide and Open verdicts reported to the Home Office, England and Wales, 1980–1994



For deaths due to accidental falls, there was no apparent change in the rates for men, while those for women appeared to go up very sharply in 1993 and then down again in 1994. It is very difficult to disentangle this. In England and Wales quite a number of these fall deaths in elderly women are certified as being due to osteoporosis, even when a fall is also mentioned. ONS then codes them to osteoporosis. Most other countries, following the ICD–9 (1) instructions that an external cause such as a fall cannot normally be due to disease, would select the fall as the underlying cause. I think that this is a way of getting out of holding an inquest into these deaths in the elderly. Any death which involves any degree of injury or unnatural cause or accident has to be referred to a coroner by law, and in theory has to have an inquest. In fact, the coroners exercise a bit of discretion here, but the

law says they have to hold an inquest, which is held in public court. We know that doctors are very reluctant to mention falls and fractures on the death certificates of elderly patients (9) probably because it causes delays for the relatives. One result is that England and Wales have about half the apparent mortality rate from falls that Scotland has, because in Scotland these deaths can be certified by a doctor in the normal way.

So what happened when we changed the coding was an odd mixture of artifacts. ACCS coded many deaths certified by doctors as falls (E880–E888) which would previously have been coded to osteoporosis. At the same time, it coded many deaths certified by coroners, with a verdict of accidental death, to diseases because the coroner had not put the injury or the fall in Part I of the certificate. I think we have not yet completely sorted out what is really going on in death rates from falls and fractures.

Dual coding

NCHS nosologists very kindly re-coded some of our deaths, and explained why they selected the underlying cause in each case. We sent them a sample of the certificates that we had difficulty with. These were chosen to show the range of coding problems we encountered. They were not a statistically representative sample of the deaths. We found that there really are differences in the interpretations of what is an acceptable sequence, and what is a default sequence (i.e., one which makes you jump to somewhere else using rule 3 or linkage or one of the other modification rules). There are a few differences in specific codes, but those are fairly rare, and there is a big difference in how we use the verdict or manner of death to select the underlying cause.

The easiest bit, and the bit that we can dispose of very quickly, is specific codes. There are very few of these, because most diseases are indexed. There are a few where we end up in different places. They tend to be diagnostic terms made up of long strings of words. You can end up in different categories depending on the order in which you take the words when using the ICD index, particularly the choice of lead term. For example, multi-infarct dementia, dementia due to having a whole series of minor strokes, the United States codes that to 290.4, which is atherosclerotic dementia. We start with infarct, multiple, and we end up with ‘unspecified other late effects of cerebrovascular disease’ in the cardiovascular chapter. So they are both perfectly sensible, and we end up at 434.9 and they end up at 290.4; we are in different chapters. It is not that common, but it does make some difference.

The most significant difference in actual codes is probably that we were coding AIDS to the disorders of immunity chapter. The code is 279.1, and as you know, the United States uses the block of *042–*044, so that they could get more detail. We knew that change was going to happen. So we can still find our AIDS deaths, and we can explain to people why they are in a different chapter from where they used to be, and why infectious diseases have gone up. That is not a problem.

Using the format of the death certificate recommended in the International Classification of Diseases (ICD)(1), the condition thought to be the underlying cause of death should appear in the lowest completed line of part I:

I. a. disease or condition directly leading to death
b. other disease if any, leading to a
c. other disease if any, leading to b

II. other significant conditions contributing to the death, but not part of the direct sequence

What I am going to do now is show you a few death certificates where ONS gets different codes now from what we used to get, and see what you think they ought to be (see Appendix). First of all, here are some questions about what is an acceptable cause. If you have a death certificate that says cerebrovascular accident or cerebrovascular hemorrhage or cerebrovascular thrombosis, due to myocardial infarction, how many people go for myocardial infarction as the underlying cause? How many people go for cerebrovascular accident as the underlying cause? Almost evenly split there.

In England and Wales, the coders had been told that myocardial infarction cannot cause stroke, and stroke cannot cause a heart attack. In the United States, either one can cause the other, as far as I know, if that is what the certifier has written down. So ACCS codes it as 410, England and Wales code it as 436.

Here is another one where there are differences in the same sort of thing: 1a, Toxic megacolon due to Crohn's colitis, and then 1b, Crohn's disease. England and Wales code the underlying cause as Crohn's disease affecting the colon, 555.1, the ACCS system codes it as 556, which is an unspecified colitis.

Multiple brain contusions due to an epileptic fit, due to epilepsy, with a verdict from the coroner, natural causes. With this one I think there is actually something wrong in the software because I think the ACME decision tables say that injuries can be due to epilepsy, but it is coming out as 928.9 (unspecified accident), and I think that may be our fault in the processing somewhere. England and Wales would definitely code that certificate to 345.9, epilepsy unspecified. I do not actually think that there is a disagreement on whether that is an acceptable sequence or not, but I think we need to work out where the software is making the error.

Now, here is an example where we differ in interpreting what is a default sequence: Pneumonia in Part I, carcinoma of the bronchus in Part II. No durations for either of them. England and Wales select carcinoma of the bronchus, the ACCS system sticks with the pneumonia. I chose this one because we actually had a problem in that a physician epidemiologist complained about the coding of a death like this in a research study. There were only 120 deaths in three years where this made a difference, out of all the thousands of lung cancer deaths. It is not a big problem but there is certainly a real difference in interpretation of the ICD selection rules.

Here is another one which gets a little bit more complicated: acute renal failure due to acute tubular necrosis due to urinary tract infection (UTI) and diabetes mellitus. The certifier cannot decide which of the conditions on the last line causes the diseases above it. How many people use rule 3 to link with diabetes mellitus? ONS would accept the sequence as far as the UTI, but then use rule 3 to select the diabetes mellitus and it is coded as 250.3 (with renal complications). The automated system codes it to the UTI.

Now we move on to a real problem we had with the verdict, which is the manner of death box on the death certificate, and for us is the coroner's legal verdict at the end of an inquest. We have real problems because the cause part of a coroner's certificate gets written by the pathologist after they have done a post-mortem. (The coroners' death certificates have only been in the standard format, (1a,1b,1c, 2) for 3 years, and I do not think we have explained the significance of the sequence properly.) Then, several days or weeks after the pathologist conducts the post-mortem, the coroner holds an inquest, and he or the jury come to a verdict. The verdict sometimes does not agree with what is written in the cause section. It is then very difficult to decide how you are going to select the underlying cause.

For example: Left ventricular failure due to ischemic heart disease (IHD), but then the ischemic heart disease was due to the fractured neck of femur, which was due to a fall in the hospital, and the coroner decided this was an accidental death. I do not think fractures can cause ischemic heart disease, even if they can precipitate an infarct in someone who has IHD. I think it is a rare sort of aetiological sequence there. On the other hand, the coroner says this was an accidental death. That is his verdict, or that is a jury's verdict: this person died as a result of this accident.

The automated system selects ischemic heart disease because that is an acceptable sequence and you believe the certifier here. ONS goes with the verdict. If the verdict says "accidental death," it is selected as the underlying cause.

Here is another one: 1a, Cardiac arrest. The verdict, "took her own life while the balance of her mind was disturbed." We code that to suicide by other or unspecified means. The automated system codes it to cardiac arrest.

PARTICIPANT: If there was a check box that said suicide, suicide would be printed in Part II, the multiple cause, but the underlying cause of death would be cardiac arrest. Because there is no internal cause on how the suicide or the external cause came about, you cannot select suicide. Following the rules, we have to go with cardiac arrest but put suicide in Part II as a multiple cause.

DR. ROONEY Right. ONS would code the underlying cause as suicide by other and unspecified means. I think it is E958.9.

This is the last example: Heart and liver failure, "took an overdose of paracetamol" in Part II. (Paracetamol is roughly equivalent to Tylenol.) Verdict, open.

It is open as to whether it is an accident or a suicide, so it is an injury. We would code that to E980.0, which is undetermined intent, poisoning by aromatic analgesics. ACCS codes it to heart failure, because there is no sequence. If it had been liver and heart failure, then there would be a sequence. Then you could select the poisoning. But if it is heart and liver failure, you are stuck with the heart failure, because of the selection rules.

Conclusions

These are bad certificates. One of the key things to remember is that if the certificate is filled out properly, you hardly ever need the rules. It is when people are filling them out badly that things go wrong and you have to apply rules. If we have differences in the way we apply those rules because of differences in what we understand as an acceptable sequence or a default sequence, it will affect our statistics. It will mean that they are not comparable between countries.

The decision tables in the software are based on what is understood as being a pathological sequence that is acceptable, and what is understood as being so likely, that it should be assumed, and that affects all our statistics. Differences in these underlying assumptions are what caused the big artifacts in England and Wales' mortality data: an apparent doubling or halving of deaths from senile dementia. We are not talking about little changes at the margins.

So, I think we all need to understand what goes into this kind of decision. I would like to think that there would be some way that we can arrive at a consensus on sequences, what is acceptable, what is not. There will be some variation and opinion between countries, but the purpose of using an international classification is to try and get statistics which are comparable over time and between countries.

The purpose certainly for me of using an automated system, which is the same in a lot of countries, is to improve that comparability. Then we can use international comparisons for epidemiology and public health purposes–for identifying probable causes of diseases and what sorts of preventive measures we can take. If we are all coding deaths differently, we cannot possibly reach sound conclusions from data that we have all produced in different ways. So I would like to think that we can come up with some kind of mechanism for arriving at consensus on what should be in our automated software, what decisions the decision tables (10,11) should be making.

Comments

MR. ARRUNDALE: Just to comment on the dementia rates, in Scotland over the last 10 years we have noticed an increase in the age-standardized rates for dementia. I do not have the figures with me, but that could likely happen.

DR. ROONEY: Our rates have gone up, very definitely, but that is fine. My correction factor should still get certificates being written now to the numbers we are getting now, but it does not. Either the certificates are written differently or there is more difference in that coding than I understand. I think it is a combination of both.

DR. ISRAEL: I think that Dr. Rooney's presentation is a very important contribution to the understanding of international comparability of data. We all know that there are differences, but many of them have been more or less either unidentified or swept under the rug, because we are "all using the international advice guidelines and documentation," and we quite glibly make direct comparisons from country to country, knowing that there can be these kinds of problems in the data.

I think what Dr. Rooney has done is show that in an attempt to standardize and make comparable, we are going to find a number of these. In my opinion, which way they are resolved, if they are to be resolved, is less important than they be resolved as an international issue.

I have one other point. On your use of the bridge coding ratios to adjust your data, they were calculated at some point in time. I believe that after 10 years they probably have decayed in their utility and value. Nevertheless, it was what you had, and I thought it was a very useful exercise.

DR. ROONEY: I agree entirely with that. They are not good enough for correcting the data in the last half of that period, but they are all I have. I wish we had bridge coded it.

I would like to say just quickly, thank you again to the people at NCHS who did the dual coding for me, and showed me that the differences are real. In the interpretation of the rules, it isn't just that something is going wrong in the software. It really is that in different countries, people have different opinions on what is an acceptable sequence and what is not. But it was very kind of them to code all our bad certificates for us.

DR. DONOVAN: I think the explanation for the problems in relation to dementia, at least if Australia's experience is any guide in relation to yours, is that the percentage annual increase in dementia deaths is not independent of age. In fact, it is quite strongly age related, and is very much higher at the older ages than at younger ones. That may be why your

correction factor applied at the time you started using it, and not in 1984, if that was the right year, and why that factor was no longer applicable 10 years later.

DR. ROONEY: It is more than that. I have looked at it in 5-year age groups, up to 1995. It is going up at more or less the same ratio in all age groups. The absolute numbers are much higher in the higher ages, but the relative increase is roughly the same for us in different age groups.

The group at the top, the open-ended 95 plus or 85 plus is going up astronomically, because that group is not properly age-standardized. It is open-ended at the top; so you will have more and more 110-year-olds in it over time. But if you look at the actual age-specific rates in the closed age groups in our data, it is going up relatively about the same. It is going up in the 60-year-olds.

DR. JOZAN: But still, this is a very strong argument. I think for the future, taking into account the increase in life expectancy at the older ages, it probably would be advisable to standardize death rates up to age 80 and to look at deaths over the 80 age group using a different approach, even from a nosological point of view, because they are definitely different by nature.

DR. ROONEY: I agree. The deaths over 80 or 85 are very different. You get hundreds of causes written on the certificate. It is often a rather arbitrary choice. Another thing that happened in our data was that the person in charge in 1985 put into the certificate instructions for the doctors that over the age of 75, if you could not decide what the underlying cause was, you could put old age. So we actually have an increase over the years, too. In every age group from 75 onwards, we have a huge increase in deaths that are actually certified as due to old age. The numbers below age 85 are tiny, but above 85 we are getting 5,000 or 6,000 deaths a year now.

DR. JOZAN: For external causes in which falls on the same level are included, they are fairly frequent in Western European countries, the United States, and Canada, where there are old people, old ladies who are in their own home, and fall because of certain reasons. In the Russian Federation or in the former Soviet Union, where the population is much younger, there are much fewer falls on the same level, but there are more accidents, traffic accidents, and homicides. If you compare the age standard, the gap is not as large as in France. So that is a very serious issue.

DR. COLE: It is not really a question to you, but I thought that your talk was excellent, and it served to convey a message that I think is very important. It was suggested on the first day that conversions to automation and ICD-10 should be done at the same time in order to minimize disruption. I think you are compounding and confusing issues you will never fully

understand unless you do the conversions separately and bridge code separately, and try and understand the nature of the problems.

DR. ROONEY: I certainly would prefer to do the changes at different times. This same year, we also changed the coroner's form. We changed from using date of registration for our tabulations to date of occurrence and we changed the standard populations that we used for our age-standardized rates, from the 1981 to the 1991 population, and from world to European.

It drove me absolutely mad. There was nothing stable across this year, so you really could not make sense of it. Now, I could deal with the populations. I can just do that myself, so that is not a real problem. If we had bridge coded it, I am not sure I would have minded how much was due to the coroner's form and how much was due to the coding, as long as I had ratios that I could make sense of the data. I would have preferred to have them separate, but I think you cannot always be an absolute purist on that. Bridge coding is expensive and it is difficult to understand.

I think some people may have to do the two changes at once, as long as you bridge code the changes really thoroughly, so that you understand what your new data means compared to your old data. That is more important than doing them separately, I think. I would prefer to do them separately, but if doing them separately means doing a slipshod quick little comparison or doing a major change and improper bridge coding of a whole year's data or something, I would go for the latter, personally.

References

1. World Health Organization. International Classification of Diseases, Ninth Revision. Geneva, WHO. 1977.

2. OPCS. Mortality statistics England and Wales: Cause 1984, Series DH2 no.11, HMSO (London) 1985.

3. OPCS. Mortality statistics England and Wales: Cause 1985, Series DH2 no.12, HMSO (London) 1986.

4. OPCS. Mortality statistics England and Wales: Cause 1986, Series DH2 no.13, HMSO (London) 1987.

5. Aylin P, C Rooney, F Drever and M Coleman. Increasing mortality from Creutzfeldt-Jakob disease in England and Wales since 1979; ascertainment bias from increase in Post mortems? Population trends 86 (1996) 34–38.

6. Home Office statistical bulletin. Statistics of deaths reported to coroners: England and Wales 1994. London, Government Statistical Service, 1995.

7. Department of Transport. Road Accidents in Great Britain, 1994. London, Government Statistical Service, 1995.

8. ONS. Mortality statistics England and Wales: Cause 1993 (revised) and 1994 Series DH2 no. 21, HMSO (London) 1996.

9. Goldacre, M. Cause specific mortality: understanding uncertain tips of the disease iceberg. JECH 1993: 47: 491–496.

10. National Center for Health Statistics. Instruction manual part 2a; Instructions for classifying the underlying cause of death. Hyattsville, 1992.

11. National Center for Health Statistics. Instruction manual part 2c; ICD-9 ACME decision tables for classifying the underlying cause of death. Hyattsville, 1992.

Appendix.

Examples of death certificates from England and Wales used in the dual coding exercise.

Death certificate number 1

Ia cerebrovascular accident
b acute myocardial infarction
c

II

Death certificate number 2

Ia toxic megacolon due to crohn’s colitis
b crohn’s disease
c

II

Death certificate number 3

Ia multiple brain contusions
b epileptic fit
c epilepsy

II
verdict: natural causes

Death certificate number 4

Ia pneumonia
b
c

II carcinoma of bronchus

Death certificate number 5

Ia acute renal failure
b acute tubular necrosis
c urinary tract infection and diabetes mellitus type I

II

Death certificate number 6

Ia left ventricular failure
b ischaemic heart disease
c fractured neck of femur following fall from bed in hospital

II
verdict: accidental death

Death certificate number 7

Ia cardiac arrest
b
c

II
verdict: took her own life while the balance of her mind was disturbed

Death certificate number 8

Ia heart and liver failure
b
c

II took an overdose of paracetomol
verdict: open

Technical Aspects of Language Conversion: Details of the NCHS MICAR Systems

James Hart, OAO Corporation[1]

Note. Much of the technical information in the NCHS MICAR systems have changed since the presentation of this paper. For updates to the systems, please contact the National Center for Health Statistics.

This presentation focuses on three main areas of the ICD-10 software, currently under development at the National Center for Health Statistics (NCHS). The first covers the technical details of the MICAR systems, how they are created, and the elements that make them work. This is targeted for those who are interested in porting the MICAR software to another platform, such as Unix or OS/2.

The second focus area concentrates on aspects of language conversion. Specifically, what portions of the systems would be affected by a conversion to another language, the feasibility of conversion, and the work required. This includes an overview of how Super-MICAR works, targeted for those who are interested in converting the MICAR software to another language.

The final part of the presentation is a quick overview of the efforts required to modify decision tables in MICAR200 or ACME. This is useful for those who are interested in using the NCHS system, but need to classify terms differently.

MICAR

The software systems are written using Microsoft Visual C/C++ version 1.52, the last version developed by Microsoft for 16-bit applications. We also rely heavily upon Microsoft's Foundation Class Library, which is a library of standard C++ routines for handling the visual elements of Windows-based programs.

Database

Data is stored in an X-base compatible database using the CodeBase 5.0 database engine from Sequiter Software, Inc. This engine does not require a license fee for the users of the software nor a client-server overhead such as SQL or Oracle require. As well as the DOS and Windows versions, there are also special versions for OS/2, Unix, and several other operating systems.

---

[1]Contractor to the National Center for Health Statistics.

System requirements

We are often asked about the minimum computer requirements to run these systems. We have two configurations that we publish, one is a minimum configuration, which is the bare minimum you can have and effectively run the system. We also have a suggested configuration, which we feel allows you to run the systems at optimal performance:

Minimum Configurations for the ICD-10 MICAR Software Systems

Super-MICAR Processing, MICAR100/200 PC-ACME/TRANSAX

Option	Minimum Required	Suggested
CPU	486/50	Pentium 75 or higher
RAM	4 Megabytes	16 Megabytes
HDD	340 Megabytes	850 Megabytes
Monitor	VGA	VGA
Operating System	MS Windows 3.1	MS Windows 3.1
Other	3.5" floppy, mouse	3.5" floppy, mouse

Hash tables

All of the MICAR programs rely heavily on a series of data tables. These can be simple lookup tables or decision tables. The older mainframe versions of the programs used virtual storage files for these tables. The original versions of PC-MICAR Data Entry and PC-ACME/Transax used a variety of data representations, from a dBASE III data base to ASCII tables with specific lookups. As the tables became larger, however, these lookup methods became too slow for data entry and processing. As a result, we have turned to a data structure called a hash table for storing our large dictionaries.

Hash tables are the fastest lookup method available for large data tables that do not change from one time period to the next. The theory is that the key value that you are looking up is used as an index into the file. For example, in the MICAR dictionary, we might want to look up the term “ACUTE MYOCARDIAL INFARCTION” to get its entity reference number (ERN). The hashing routine will take the term and perform mathematical calculations on it, reducing it to a number. That number will be the position of the term in the dictionary. For example, if acute myocardial infarction returns the number 200, then that term will be the 200th entry in the dictionary. We can then go to that point in the dictionary and read the ERN for that term.

The advantage of this method is that the hard drive only has to be accessed once. The most time-consuming aspect of any PC-based program is accessing the hard drive. Early versions of our lookup tables used a sequential search which read every record until it found the one it wanted. If that happened to be the last one in the table, you could be looking for a long time. Later, binary searches replaced the sequential searches, which

significantly reduced the number of times the disk had to be accessed. Keyed database searches further reduced the processing time.

In the 1996 version of the MICAR software, the hash tables were implemented for all of Super-MICAR and PC-MICAR data entry, and for the larger lookup tables in MICAR100/200. We have also begun to use the computer's extended and expanded memory for the hash tables. Loading the tables in extended memory means that the lookup takes place in computer RAM, not on the disk. The lookup time required is a third of what it was previously.

The hash tables are the cornerstone of all of the systems. Any modification to the systems for operating systems, language, or decision tables requires modifications to the hash tables. NCHS can provide the base files from which the tables are built, the programs that we use to build them, and the necessary source codes for hash table building and lookups.

Conversion

Several countries have asked us about conversion of the software, either to another operating system or to another language. One of the reasons that we originally chose the C language (and later C++) for our systems is that these languages are used on almost every type of machine, from personal computers (PCS) to mainframes. Theoretically, this will make it easier to port the programs to another platform. The database, interface and tables need to be considered when porting to another operating system.

Database

As mentioned earlier, we use the Codebase database package because it has several advantages over more traditional client-server databases. First, there are no license fees for multiple users. Developers can buy the package once, create programs that use it, and distribute it freely. Second, because there is no overhead, programs can be stand-alone. There is no requirement to connect to a network or link into a distributed system. One person working alone can use the system without requiring support personnel or the administrative costs of supporting a large system. Third, it has versions for DOS, Windows, OS/2, UNIX, and Macintosh.

Interface

We use the Microsoft Foundation Classes to handle our interface issues, plus a number of Visual Basic controls. To convert to another operating system, you would need to replace our interface routines with those native to your system. Under UNIX, you might want to convert to XWindows/MOTIF or some other interface type. Under OS/2 or Macintosh, you would have to replace our Windows routines with the proper Desktop routines.

Tables

While the hash table concept is viable on any operating platform, there will be some required changes due to the change in the new operating system's architecture. File names may be different. The libraries used for extended and expanded memory management will be useless, and perhaps unnecessary. In either case, you would need to know the underlying principles of our hash routines

and be able to duplicate them on your system. As another option, you could replace the entire hash system with a lookup method or database that is native to the new operating system. This may affect performance, but may be easier to implement.

Language conversion

Only the data entry systems, PC-MICAR and Super-MICAR, process literal text. Therefore, only they need to be converted from one language to another. Since Super-MICAR and PC-MICAR perform the same function, that of collecting data and preparing it for MICAR processing, only one of the two packages needs to be converted.

The first decision is which package to convert. One of the major factors bearing on this decision is the ability of your data entry personnel. PC-MICAR requires trained data entry operators, while Super-MICAR requires little or no training for data entry operators, but requires more detailed conversion by the programming staff.

The underlying backbone of both data entry packages is the MICAR dictionary. The MICAR dictionary is a list of over 200,000 medical conditions that have been reported on the death certificate. Both Super-MICAR and PC-MICAR rely on that dictionary. Both data entry packages format the data entry literals in such a way that the resulting terms can be looked up in the MICAR dictionary to find an ERN, the six-digit code the software uses to identify each discrete cause of death.

In addition to the MICAR dictionary, both PC-MICAR and Super-MICAR include an interactive spelling checker that validates the words as they are entered. The spelling dictionary, called the medical lexicon, is a list of the over 9,000 words that appear in the MICAR dictionary. If the word is misspelled, the software offers a list of suggestions. In the 1996 and prior versions, the software used a traditional Soundex lookup with data contained in a hash table. While this worked well, the suggestion lists were not always the best. For ICD-10, the Soundex lookup will be replaced with a data structure called a TRIE. This is the structure used by the spelling checkers in word processing software. It creates a formatted database that contains within it a representation of all the correctly spelled words, then uses an algorithm to traverse the database, assuming that the word entered contained either an extra letter, a dropped letter, or transposed letters, as these are the most common causes of misspelled words.

PC-MICAR

PC-MICAR requires a trained coder to read the death certificate, determine the causes of death reported on the certificate, and enter those causes in the proper format so that the term entered can be found in the dictionary. However, it also allows for some words to be dropped from the term, such as POSSIBLE or MASSIVE, and for some words to be exchanged for synonymous words, such as ABDOMEN for ABDOMINAL. In addition, it allows for plural words to be replaced with their singular form if the plural form does not exist in the dictionary. After each change, the resulting term is verified in the MICAR dictionary. The best term, the one with the fewest words dropped and exchanged, is kept.

To convert PC-MICAR into another language, the entire MICAR dictionary needs to be converted as well as the lists of droppable words and synonyms. The part of the program that identifies plural words and replaces them with singular forms will also need to be modified. The rules for matching the dictionary, however, does not need to be changed. The lexicon (the list of properly-spelled words) will need to be converted and the TRIE will need to be rebuilt so that the spelling checker will check the spelling in the new language. However, once the MICAR dictionary is converted, gleaning the words for the lexicon is a trivial issue. The software would also have to be modified so that captions, messages, and help files would be in the new language.

Super-MICAR

Since Super-MICAR uses language more during processing, one would suspect that it would be more difficult to convert. However, depending on the language you are converting to, this may not be the case.

Scotland has taken Super-MICAR and converted it from American English to British English. Although it was time-consuming, it was not a huge task. There are not a lot of differences between the two languages. Most of the differences are in spelling. For example, in the United States we spell Esophagus with an ‘E’, while in Britain it is spelled ‘OE.’ Every time the word was entered into Super-MICAR, it would be flagged as a misspelling. If not corrected, it would never match the dictionary.

Super-MICAR contains a dictionary that looks up each word and assigns it a category based upon its definition. This dictionary also can automatically exchange a word for a standard replacement. For example, the abbreviation, “2nd,” is always replaced with its meaning, SECONDARY. This allows Super-MICAR to immediately replace a word that does not appear in the dictionary with an equivalent word that does. To solve the British English problem, Scotland’s programmer only needed to modify the dictionary so that OESOPHAGUS was always replaced with ESOPHAGUS and the word OESOPHAGUS was added to the lexicon. With the word in the lexicon, it was accepted as correctly spelled.

Of course, British English and American English are a lot more alike than American English and, say, French. But it is possible that the same solution could work for French, or any other language. To demonstrate how language conversion would be done with a non-English language, I have an example in French. The English term, ADVANCED LIVER CANCER, translates to CANCER DU FOIE AVANCÉ in French. As each word is entered, it is looked up in the lexicon. If the 9,000 words contained in the lexicon are translated and rebuilt, then once processing begins, each word is looked up in the dictionary as it is converted to the English equivalent.

```
CANCER      ==> CANCER
DU          ==> OF
FOIE        ==> LIVER
AVANCÉ      ==> ADVANCED
```

Each word is assigned a numeric definition as follows:

0: Linking words
1: ACUTE, CHRONIC, SUBACUTE
2: Modifiers
3: Infectious parasites
4: Body site modifiers (Left, right, upper, lower, large, small)
5: Body sites
6: Cancer Modifiers (T Cell, Squamous Cell)
7: Lead terms
8: Neoplasmic lead terms (cancer, carcinoma)
9: Infectious lead terms (pneumonia)

Then the numeric definitions are applied to the terms:

8:	CANCER	CANCER
0:	DU	OF
5:	FOIE	LIVER
2:	AVANCÉ	ADVANCED

Super-MICAR ranks the words by numeric category:

0:	DU	OF
2:	AVANCÉ	ADVANCED
5:	FOIE	LIVER
8:	CANCER	CANCER

Words defined as “0" are dropped since they do not divide or link terms. This leaves ADVANCED LIVER CANCER, which is what would have been matched if it was entered in English.

Language requirements

Super-MICAR also incorporates a number of other helpful tools that aid in conversion. The translation does not need to be a direct one-word-for-one-word translation. Even in English, there are two-word terms that should be kept together-Gunshot wound, for example. When these two terms are entered together, they are linked by Super-MICAR. Once linked, they are treated as a single word for all processing purposes. In cases where it takes two words to express a concept represented by a single English word, the same process is used to combine the words into the English equivalents.

Translation is much smoother when the language contains a defined set of linking words with standard rules. By linkage, I mean a word which causes an action when it appears in a sentence. For example:

CANCER OF THE LIVER AND LUNG

This should be reported as two separate terms, LIVER CANCER and LUNG CANCER. The word AND dividing two body sites with a single lead term indicates that each body site should appear with the lead term. Another example is:

PNEUMONIA WITH LUNG CANCER

In this case, the word WITH divides two lead terms and thus two complete terms. This would be processed as PNEUMONIA as one term, LUNG CANCER as another. A complete translation of Super-MICAR requires:

1. Conversion of the MICAR dictionary
2. Conversion of the Words dictionary
3. Conversion of the Lexicon
4. Conversion of the Drop words, Synonyms, and other associated tables
5. Translation of the External Cause Prompt coder
6. Translation of over 600 word-specific exceptions
7. Complete translation of the code.

There are five conditions that can apply to every word in a phrase processed by Super-MICAR:

1. Can the word be dropped?
2. Does the word have a synonym with which it can be replaced?
3. Is the word the plural form of a singular word?
4. Does the word have a common abbreviation?
5. Is there more than one word in the line with the same category?

Once these parameters have been established and the words ranked, Super-MICAR goes through a series of loops to check every possible combination of drops, synonyms, plurals, abbreviations, and word orders to see if the term is in the dictionary. If it finds more than one combination in the dictionary, it accepts the one that has the greatest number of words with the least number of synonym replacements.

Once all medical processing has been completed, the external cause processor looks for unmatched terms indicating external causes. The related terms are pulled together to build an external cause prompt for each term and to match the dictionary for both the term and any nature of injury codes imbedded in the term.

Modifying tables

The last topic is the modification of the tables in the NCHS system to meet the needs of other countries. Adding a term to the dictionary seems like a simple task, but it can be difficult because the tables can be so complex.

It may only require modifying 2 tables, but it could also require changing up to 30. Someone with a thorough knowledge of the system needs to make the changes. The term needs to be added to the MICAR dictionary used by PC-MICAR and Super-MICAR, and assigned an ERN. Then the ERN needs to be added to the ERN conversion table, CVTERN, used by MICAR200.

In addition to the ERN and the ICD code, this table also contains a series of flags that indicate other tables and circumstances under which this ERN may need to be modified. There are 27 tables that MICAR uses, from age and sex edits to Intent of the Certifier to histological cancers. While no term will affect all 27 of them, the person building the tables must know which tables will be affected. NCHS can provide the program source codes, routines

used for building the tables, and base files from which the tables are created.

Causal relationships

There are some tables that can be easily modified. For example, the causal relationship tables in ACME can be changed quite easily. These tables define causation of disease, specifically, which entities can be caused by which other entities. Some countries disagree with NCHS's definitions of causation, and some relationships are more valid in other countries than in the United States. The causal relationship tables contain an entry for each possible causal relationship. If disease X can be due to condition Y, then that relationship is in the table. To add a relationship, it is only necessary to add the relationship to the data file and rebuild the table. To remove one, simply delete it and rebuild the table. This only affects the underlying cause.

Comments

DR. ROONEY: There are several improvements in Super-MICAR, including this much more sophisticated ability to split terms. There are also some advantages in the Tracer program which essentially tries to do the same thing. (I think Tracer may be a little more flexible.) Is there any way that we can get the best out of both programs, rather than having to decide one way or the other?

MS. GLENN: One project that we have been discussing is the ability to go into the Super-MICAR dictionary on an interactive basis and do more with it than we have done before. Part of our problem is that we are changing from a mainframe system to a PC system, and the dictionary programs are still on the mainframe. We need to get everything into the PC system before we can start talking about those kinds of changes. I really like what can be done in Tracer. There are a lot of benefits. Since our States are using the system, we want to have some consistency, but I would also like to see more flexibility.

DR. SANTO: You said that the choice between PC-MICAR and Super-MICAR would be the need for trained coders to enter input data in MICAR. Are there other differences which should be considered before choosing between PC-MICAR and Super-MICAR?

MR. HART: Well, there is also the question of how difficult it is going to be to convert the software. I do not know if there can be a direct word-to-word conversion of the Super-MICAR dictionary from English to Portuguese. That is something you would have to check into yourself. It may be impossible for you to convert Super-MICAR, but it may be possible for you to take the principles of Super-MICAR and adapt it to Portuguese.

DR. KARDARN: I would like to make a remark about different languages. Sometimes you can translate and reshuffle words, but very often things are said in a quite different way. For example, in German you can combine words to create compound words, which are quite difficult to analyze.

MR. HART: One of the things that can be done in Super-MICAR is to take a single word in one language and convert it to a phrase. An English phrase can be converted to a single word. In terms of combining words, I do not know exactly how that would work.

DR. WELTNER: Have you considered what would happen if you did not translate the words, but just the ICD codes or the ERNs, in the decision tables? Of course, the ICD codes would not give us the detailed information about the disease but might be helpful for translation between languages with very dissimilar grammatical patterns. Perhaps at some point it would be possible to compare mortality data on the basis of a MICAR or Super MICAR program.

MS. GLENN: For clarification, when we say decision tables, we are usually talking about ACME, which uses ICD-9 codes. MICAR uses entity reference numbers in the decision tables. They can be converted to ICD-9 codes but there will be some ambivalent relationships. There will be a set of true ICD-9 or ICD-10 decision tables, but some records will be rejected.

MR. HART: One reason that we used the entity reference numbers instead of the ICD codes was to make the data entry systems independent of the ICD revision. It also gives an extra level of clarity in terms of querying the database.

DR. PEREZ: In Spain, we have four official languages. The physicians write the death certificates in those four languages. In Catalonia, we have two official languages. We wanted to translate the dictionary and use the NCHS automated systems. It was impossible for us to translate both languages. Our alternative was to use neural networks (*see Chapter 17)*. I do not know if that is an option for other countries.

MR. HART: I think that is certainly a possibility. While it would have been possible to convert the MICAR dictionary to both of the languages that you use and have both of those languages in the dictionary, the dictionary would have been much larger. You could have each term in the dictionary and just multiply the size of your dictionary and the size of the job.

DR. MONTELLA: With the language conversion, you reorganize to put the entity in the English order. I think this may be a problem in the sense of the interpretation of this medical entity.

MR. HART: Having the words in a different order would create a different meaning?

DR. MONTELLA: Sometimes the interpretation of the same medical entity is different in a different language. If you reorganize only the words, the medical entity may have a different meaning and perhaps a different code.

MR. HART: I had not noticed any case in which changing the order of the words changes the meaning of the medical entity. Maybe this needs further consideration.

DR. SANTO: What are the differences between MICAR100 and MICAR200?

MR. HART: MICAR100 is going to disappear in the ICD-10 systems because it is simply reprocessing a dictionary match that has already been done. MICAR200 reads the entity reference numbers and converts them to their ICD codes, applying rules and linkages and looking for age or sex modifications. With MICAR200, the data can come from PC-MICAR or Super-MICAR. The number of rejects may be higher with the Super-MICAR data, simply because Super-MICAR can not match the

dictionary as often as a trained coder. Super-MICAR's throughput is five to ten percent lower than PC-MICAR's, although that gap is closing.

DR. ISRAEL: On a slightly different point, can you tell us a little bit about how much work or extra work will be involved in running a set of death certificates through the ninth revision version of one or more of these systems, compared to the tenth revision version, to produce the bridge coding? Will the same original data entry work in the two systems, or do they have to be re-entered?

MR. HART: The same original data entry will work in both systems.

DR. ISRAEL: Is it an oversimplification for me to think that it would be relatively easy to redo a large batch of records in order to get the comparability ratios?

MR. HART: If we use Super-MICAR data, we can reprocess that. With the external cause coder we have now, the external causes can be immediately processed. One of the problems with using PC-MICAR data is that all the external causes would have to be converted, and we would lose some definition unless we used the 1996 or 1997 data. Any data after 1996 already has that in there. The amount of work involved in coding the rejects also needs consideration, and the rejects will be higher with Super-MICAR than with PC MICAR.

DR. ROONEY: You are saying you can do the bridge coding from the original data entry. That is only if you used Super-MICAR and you have stored the text, isn't it? You cannot do it from the entity codes. You have to do it from the text, is that right?

MS. GLENN: We have a problem with what we call “drop words.” We drop the word and we have an entity reference number. To overcome that with the 1996 system, we will match and drop the word, but retain the full text. For example, in a case where the word “acute” makes a difference in ICD-10, but not in ICD-9, the word would not be shown in the ERN. With the national data, we changed this so that the word was dropped with PC-MICAR, but any text that was changed was pulled. We can rerun even our PC-MICAR data, as long as it is 1996 and beyond.

DR. ROONEY: We are using an earlier version of MICAR.

MS. GLENN: You cannot go from just the entity reference numbers.

DR. ROONEY: We actually have text stored on every death certificate anyway, so we can pull it from the electronic text. Would it have to run through the whole system again?

MS. GLENN: Yes, PC-MICAR would also have to go through the whole system again.

DR. ROSENBERG: On the comparability study, we are using the 1996 data file, running it once through ICD-9 and once through ICD-10.

Closing Remarks

Edward J. Sondik, Ph.D., Director, National Center for Health Statistics, Centers for Disease Control, U.S. Department of Health and Human Services

This has been an extraordinarily productive meeting and I want to thank the planning committee, Elizabeth Vasquez for her help in facilitating, and, of course, Harry Rosenberg for leading. This is not the end—it is clearly the start.

I was very pleased to hear the recommendations as well as the formal chartering of the user group. I believe the establishment of the user group will have positive benefits. I was also particularly pleased about the recommendations regarding the home page and, in general, the use of the Internet. I support Charlie Rothwell's point about our willingness to use our offices as much as possible, with your help and guidance.

As for the continuation of the ICEs, I think the ICE on Automating Mortality Statistics should continue. This is incredibly productive. I am struck by the discussion of WHO's role. As Gerard Pavillon pointed out, we ought to examine that and try to firm up the relationships between WHO and the ICEs.

It seems to me that an international organization should in part serve as a clearinghouse for information. It should serve a catalytic role, with motivational power, and the resources to get things moving. I think clearly this effort among all of you is moving and WHO could play an important role by helping to distribute the results of this meeting, particularly to those nations that are not here.

I would hope that in future meetings, we can include those nations who are not as far along in automation, and that we can work to help bring them along. I certainly see a role there for an international organization.

The ICE planning committee will ensure that efforts continue in this area, and the proceedings from this meeting will be published as soon as we can.

I just want to emphasize again that an event like this is not a one-time effort. There is a saying in this country that if something is not broken, you shouldn't fix it. It seems to me that this effort is working so well and is so productive that it should continue. I believe these international collaborative efforts are quite crucial to each of our efforts in our own countries.

Sharing health information across countries can only be done if we all speak the same language. This ICE is an effort toward helping us to speak that language. By building on past research, we can work cooperatively to develop the next steps. So again, I thank you on behalf of NCHS and all those involved for this very productive effort.

Recommendations

Recommendations from the First International Collaborative Effort on Automating Mortality Statistics

During the meeting of the International Collaborative Effort (ICE) on Automating Mortality Statistics there were 20 small discussion sessions that focused on developing recommendations on issues related to international automation, the central theme of the conference. Detailed descriptions of these topics are included in "International Collaborative Effort on Automating Mortality Statistics: Background and Issues" (see Chapter 1). Recommendations from the sessions were reviewed and refined by the ICE Planning Committee which met in June 1997. [Note: The status of follow-up activities on specific recommendations is shown in italics]

Topic one: Nosology and/or the training of nosologists in an automated environment

Recommendation 1

The National Center for Health Statistics (NCHS) should provide a standard definition of nosologist.

At the 1997 World Health Organization (WHO) Center Heads meeting, NCHS proposed the following definition: "Mortality medical coders are able to apply the International Classification of Diseases (ICD) rules for coding underlying and/or multiple causes of death. Nosologists specializing in mortality medical coding are those who have achieved high levels of expertise in the practice of medical coding; in the interpretation and application of the ICD classification; in the training, apprenticeship, and qualification of new mortality medical coders; and in the implementation of special projects on causes of death."

Recommendation 2

Every country should take steps to strengthen the nosological skill and expertise of its medical coders. Countries should develop courses and seminars that give nosologists more background, information, and knowledge about automated systems. They can also use an automated system as a learning tool. WHO and the ICE on Automating Mortality Statistics can assist by developing accreditation and curriculum standards and by developing and teaching an international, recurrent ICD-10 coding course. WHO should note in their official statistics which countries use certified nosologists.

Recommendation 3

To strengthen the status of nosologists, programs that employ nosologists should increase opportunities for career advancement and publicly recognize nosologists who excel in teaching and training other nosologists. WHO and the ICE on Automating Mortality Statistics could assist by creating an international society of nosologists recognized by WHO.

Recommendation 4

With the use of automated systems, fewer, but more highly skilled, nosologists are needed. Countries should increase awareness of the need for nosological expertise.

Recommendation 5

As the number of nosologists declines, countries should consider cross-training employees (for example, statisticians could also be trained as nosologists).

Topic two: Decision Tables (and program logic) and mechanisms for updating them

Recommendation 1

NCHS, with the help of other countries using automation, will develop consensus decision tables and algorithms to improve comparability between countries.

NCHS has sent out ICD-9 decision tables to a few countries to test with test decks (by comparing manual underlying cause-of-death coding with ACME). Printed versions are also available on the NCHS website (http://www.cdc.gov/nchswww/).

Recommendation 2

Countries should test decision tables and algorithms for specificity and clarity.

Recommendation 3

NCHS will create a test deck for country and system comparisons on ICD codes and multiple causes.

Recommendation 4

Countries should use bridge coding, or comparability studies, to assess changes from ICD-9 to ICD-10, from manual coding to automatic coding, and on any subsequent changes made to ICD-10. Bridge coding will need to be done on underlying cause and done from literal text, or the diagnostic expressions from the certificate, rather than from Entity Reference Numbers (ERNs).

Recommendation 5

Because external cause comparisons between countries are difficult due to variations in sources of information, legal systems, and the amount of information, the ICE on Automating Mortality Statistics, in collaboration with the ICE on Injury Statistics (Injury ICE), should work to develop uniform

rules. The use of automated coding may also help to improve consistency between countries.

Recommendation 6

To address systemic errors in coding software, countries should be encouraged to participate in the Beta test of ICD-10 software and provide feedback to NCHS on errors and inconsistencies. NCHS will make the commitment to acknowledge and explain reported differences in the United States software and to make appropriate adjustments to the software. The software also needs version control, and the system will carry a version number.

Recommendation 7

Developers of automated systems should make the decision tables used in their systems broadly available to all countries.

Recommendation 8

An advisory committee, composed of members of WHO collaborating centers, should be established to help in interpretation of decision tables, but not in changing decision rules. This committee would report to WHO, which is viewed as ultimately responsible for ICD decisions.

Recommendation 9

The ICE on Automating Mortality Statistics will develop a process to establish consistency or recognize the differences in interpretation of the rules, and to disseminate information on international differences.

The establishment of a mortality reference group (MRG) was proposed and adopted at the 1997 WHO Center Heads meeting. The international MRG is the ultimate repository of nosological skills and knowledge. The group consists of nosologists, epidemiologists, statistical analysts, and systems design persons with knowledge of mortality medical coding. The ICE sees this group as embedded in the WHO and Center Heads mechanisms, and requested that WHO delegate to the MRG final decision making authority on technical questions relating to mortality medical coding.

Topic three: Data quality and editing

Recommendation 1

Automated coding is a step toward improving data quality, consistency, and comparability, but it is not the only step. To improve data quality, the quality of certification must also improve. Countries can help physicians complete more accurate death certificates by:

a. using a querying system for corrections and as an educational tool;

b. sending letters to physicians explaining how to certify deaths in specific cases;

c. exploring ways to make the system more accessible for physician input (for example, an electronic death certificate);

d. training doctors with a PC-based, interactive system, which includes test cases [in the United States the National Association of Medical Examiners (NAME) maintains a cause-of-death tutorial on the Internet (http://www.thename.org/main.htm)];

e. conducting quality control of medical certification through peer review.

The United States has developed a prototype tutorial recommended for use by the States who are implementing the electronic death certificate. The prototype can be made broadly available.

Recommendation 2

WHO should recommend that questions on death certification, including the concepts of sequence and underlying cause, be incorporated into medical board examinations, implying that it will be added to the curriculum and that receiving continuing medical education (CME**)** credit may be a possibility.

Recommendation 3

Countries should pursue ways to handle constraints (on budgets, time, resources, and legal issues) which restrict their ability to query physicians.

Recommendation 4

Coding supervisors should encourage the use of literal text entry to reduce the likelihood of coders changing input to avoid error messages.

Recommendation 5

Countries should use final edits of age, sex, and cause of death to verify consistency and validity among variables on the death certificate.

Recommendation 6

NCHS will put edit procedures on the Internet.

The NCHS instruction manuals are on the NCHS website (http://www.cdc.gov/nchswww/).

Recommendation 7

When countries revise their death certificates, ways to improve data quality through format, content, and instructions of the death certificate should be considered.

Recommendation 8

The ICE on Automating Mortality Statistics should establish a group to review the algorithms that interpret the ICD rules. The group will focus on clarification and specification.

Topic four: Training of automation (PC) support and mechanisms for technical support and for training users

Recommendation 1

NCHS will provide training on automation support to the best of its ability.

Recommendation 2

Countries who have already received automation support training should work to strengthen networks with each other.

Recommendation 3

Each country is encouraged to try to understand their own automated systems and to be able to change their systems when needed.

Recommendation 4

Countries should recognize that two types of support skills are needed for: (1) daily operations and general computer skills, and (2) systems and platform support, including training. Systems and platform support personnel will need to be highly skilled in computer applications.

Recommendation 5

Each country’s site should have a daily operations support person.

Recommendation 6

Each country will be responsible for general computer training.

Recommendation 7

Technical support of automation should be hierarchical:
Local
State
Country
Region

Recommendation 8

Developers of automated systems should include useful and understandable messages in their systems.

Recommendation 9

Developers of automated systems should document changes in their system and inform users of these changes.

Recommendation 10

The ICE on Automating Mortality Statistics should explore ways of reimbursing NCHS for training and support time.

Recommendation 11

Countries should provide medical coders and nosologists with computer training along with their nosological training.

Canada provided assistance to the United States in conversion training.

Topic five: Language issues

Recommendation 1

WHO and the Collaborating Centers should provide technical support and act as clearinghouses. They may also provide help with networking between countries and guidance on translation.

Recommendation 2

Although countries should be responsible for their own changes, language groups should share translations.

Recommendation 3

The ICE on Automating Mortality Statistics should assist countries in sharing their experiences of system revisions. The ICE should also work to increase networking between countries undergoing system revisions.

Recommendation 4

Existing medical dictionaries should be used in updating and translating dictionaries.

Recommendation 5

Translation should aim at translating ERNs, while ensuring that it is always possible to enter ACME through ICD codes.

Recommendation 6

Translators should translate dictionary terms according to frequency so that the most frequently used terms are translated first.

Recommendation 7

Countries should involve nosologists when updating their dictionaries.

Recommendation 8

Before translating the MICAR dictionary, countries need to understand how MICAR works and how to create their own dictionary from scratch. Countries should start with a sample of their death certificates, not with the MICAR dictionary.

For some countries, the translation of MICAR was a stumbling block in creating their own dictionary. This is why each country needs to understand how MICAR works and how to create dictionaries from scratch. One country took a sample of their most frequently coded causes and began their own dictionary, which will include about 6,000 terms when they are finished. Another country also used the most frequently occurring terms, which represent about 82-85 percent of their cases, but had some difficulty resolving the remaining cases.

Topic six: Implementation issues

Recommendation 1

The ICE on Automating Mortality Statistics recognizes the importance of WHO in coordinating and providing leadership in automation relative to the classification of cause of death. WHO should continue in this leadership role.

Recommendation 2

NCHS, as well as other countries who develop automated systems, should emphasize the transfer of expertise and methodology, not just products.

Recommendation 3

NCHS will establish a web presence for automated systems.

Recommendation 4

The ICE on Automating Mortality Statistics will establish a user group.

The recommendation for the creation of an automated systems users group was presented at the 1997 WHO Center Heads meeting. The users group, established by NCHS, offers technical assistance and systems support, input to software development and general information sharing with news and updates of ICD-9 and ICD-10. This group works closely with the Mortality Reference Group to ensure that decisions are reflected in general software design.

Recommendation 5

The ICE on Automating Mortality Statistics should establish an e-mail network for the general sharing of news, ideas, and questions. This e-mail network would be open to all countries using automation as well as those who are considering moving toward automation.

Recommendation 6

The ICE on Automating Mortality Statistics encourages the establishment of language based e-mail groups.

Contributed Papers

Automatic Coding of Cause of Death: The Use of the U.S. Systems in Scotland, Preliminary Results from a Bridge-Coding Exercise

Jack Arrundale, M.Sc., Susan Cole, M.D., Lesley Fraser, Jan Hannah, and Helen Lamb, General Register Office for Scotland

Introduction

From January 1, 1996, deaths in Scotland have been coded using the U.S. automatic coding software (USACS) (Super-MICAR, MICAR, ACME, and TRANSAX) developed by the National Center for Health Statistics (NCHS) in the USA. This paper describes the system within which this software has been used and presents preliminary results from an exercise to recode, using the software, a sample of manually coded deaths from 1995.

Background

In Scotland approximately 60,000 deaths per year are registered by about 350 registrars on receipt of a medical certificate of cause of death signed by a qualified medical practitioner. This certificate (Appendix 1) is similar to that recommended by WHO (Appendix 2). The registrar sends a copy of the information on this certificate, together with other demographic information relating to the deceased, to the General Register Office for Scotland (GRO(S)). At present 85 percent of such information is sent on diskette, the remaining 15 percent is keyed by GRO(S) into the same format as the diskettes.

Coders then process and vet the demographic data on computer using a system written by GRO(S). When the demographic data is 'clean,' an extract, including the cause of death text, is processed with the cause of death software. This software consists of a number of separate modules: Super-MICAR, MICAR, ACME and TRANSAX.

Super-MICAR reads the cause-of-death text and allocates an entity reference number (ERN) to each individual cause and records its position on the certificate. It is necessary to correct spelling errors and differences between British and American usage and spelling at this first stage. The next module, MICAR, converts these ERNs to ICD-9 codes. At this stage, coder intervention is needed if Super-MICAR has not been able to allocate an ERN.

The output from MICAR is then submitted to ACME which, using the ICD codes, their position on the certificate, and other information such as duration of each condition, selects the underlying cause of death. Again the coder has to deal with rejections where the software is uncertain whether the sequence on the certificate is valid or where further information, for example, in the case of violent or drug deaths, is needed for precise coding. These results are passed to TRANSAX, which sorts out any linkages between causes and eliminates repeated causes to produce the final output of up to 20 causes for each death. The output from both ACME (Entity Axis) and TRANSAX (record AXIS) are then passed back to the main GRO(S) system. However, only 10 causes from each axis are stored for analytical use.

Prior to 1996 the causes on the death certificate were coded manually by a team of trained coders. They selected the underlying cause and from 1974 coded up to three other causes mentioned on the death certificate. The coding was checked by having each death recoded by another coder. Any queries were passed to a medical advisor for a final ruling.

Bridge Coding

Any change in coding practice can have a significant effect on the final results and can cause discontinuities in time series for particular conditions. Such changes therefore have a serious effect on the work of epidemiologists and medical researchers who use the results to compare changes over time or the causes of death in different countries. It was decided to quantify these changes by coding some deaths using both manual and automatic coding.

In 1995 about 75 percent of deaths were registered in offices that used the Scottish Registration Software for the entire year. The written text of the cause of death was available for these deaths in computer-readable form. Deaths for 1995 had already been coded manually as part of normal processing, so 75 percent of deaths (about 45,000) were used as a sample and were recoded using USACS. To date, 11,499 have been recoded and this paper presents the results from this preliminary sample. The remaining records will be recoded in the near future.

Results and Discussion

When the GRO(S) coders first used USACS they studied the results that emerged from the system and made a note of deaths that were coded differently from the way they would have been coded manually. While such differences were noted, no changes were made. A list of some of the more important changes is given in Appendix 2, along with an indication of whether we accept the difference in the interpretation of the WHO rules or whether we wish to discuss them with NCHS and other users. This exercise was not a statistically rigorous one nor was it exhaustive, but it gave a preliminary impression of the changes in the final statistics that may result from the move to automatic coding. The appendix is divided into three parts:

Part 1. individual conditions or textual strings that have attracted a different code;

Part 2. sequences where a different underlying cause has been selected; and

Part 3. new codes that have been introduced by NCHS.

Process Measures

Not all the records are processed successfully by USACS without manual intervention. In order to minimize this intervention, the records in the Scottish system are browsed prior to their submission to USACS. This is part of the normal processing of the death registrations and at this stage multiple

conditions are separated by semicolons and any obvious spelling errors are corrected. Spelling errors are also checked and cleared in Super-MICAR. Approximately 16 percent of records were rejected by Super-MICAR and required manual intervention; 14 percent of records required manual intervention at the ACME stage, and 1.5 percent of those required the manual selection of underlying cause.

Tables

Table 1 gives the number of records that were originally coded to each ICD chapter and the number which USACS coded to the same chapter, ICD block of codes (see Appendix 3), 3-digit code, and 4-digit code. Table 2 gives the percentage distribution of the total number originally coded to each ICD chapter. Table 3 presents ICD block, 3-digit code, and 4-digit code as percentages that were coded within the same chapter, as well as the total number coded to each chapter by USACS.

Table 1. Deaths coded to the same ICD chapter, number coded to the same ICD block, 3-digit code, and 4-digit code, by both manual and automated coding, and the number coded to each chapter by automated coding

Manual coding		Automated coding				
ICD chapter	Number of deaths	Number of deaths in same				Total coded to chapter
		chapter	ICD block	3-digit code	4-digit code	
I	66	57	53	53	52	97
II	2,888	2,846	2,775	2,759	2,665	2,865
III	149	112	112	112	108	130
IV	31	26	26	22	21	32
V	338	312	303	281	278	323
VI	166	157	156	154	150	177
VII	5,041	4,904	4,748	4,583	4,548	5,021
VIII	1,539	1,467	1,417	1,404	1,400	1,548
IX	402	375	368	344	330	406
X	192	162	160	153	152	190
XI	1	0	0	0	0	0
XII	13	12	12	12	12	16
XIII	69	55	53	51	49	61
XIV	46	38	38	33	33	59
XV	37	31	31	29	29	34
XVI	64	60	60	50	49	66
SUPP	457	434	368	325	307	474
Total	11,499	11,048	10,680	10,365	10,183	11,499

From these tables we can see that 96 percent of deaths overall were coded to the same ICD chapter. As expected, the main causes of death were coded to Chapter VII, Diseases of the circulatory system. Of the 5,041 deaths originally coded to this chapter, 97 percent were coded to the same chapter by USACS. Of the 2,888 deaths coded to Chapter II, Neoplasms, by GRO(S) coders, 2,846 or 99 percent were in the same chapter and 96 percent were in the same ICD Block.

Table 2. Percentage of deaths coded to the same ICD chapter, ICD-block, 3-digit code, and 4-digit code by both manual and automated coding

Manual coding		Automated coding			
ICD chapter	Number of deaths	Percentage in same			
		chapter	ICD block	3-digit code	4-digit code
I	66	86	80	80	79
II	2,888	99	96	96	92
III	149	75	75	75	72
IV	31	84	84	71	68
V	338	92	90	83	82
VI	166	95	94	93	90
VII	5,041	97	94	91	90
VIII	1,539	95	92	91	91
IX	402	93	92	86	82
X	192	84	83	80	79
XI	1	0	0	0	0
XII	13	92	92	92	92
XIII	69	80	77	74	71
XIV	46	83	83	72	72
XV	37	84	84	78	78
XVI	64	94	94	78	77
SUPP	457	95	81	81	81
Total	11,499	96	93	90	89

Some of the other causes were not treated as consistently, however. For example, only 86 percent of the deaths coded to Chapter I by GRO(S) were coded to the same chapter by USACS. Overall, only 79 percent attracted the same 4-digit ICD code. These and the other divergences will require more detailed investigation.

An interesting example is the one death originally coded to Chapter XI, "Complications of Pregnancy, Childbirth, and the Puerperium." This death, for which the death certificate stated:

1(a) Pulmonary Thromboembolus - Left Vein Thrombosis,

was coded to 673.2, "Obstetric blood-clot embolism," by GRO(S) coders as the doctor had indicated on the death certificate (see Appendix 1) that the deceased had died within 6 weeks of pregnancy. This information is not used by USACS, which coded the death to 415.1, "Pulmonary Embolism." This code excludes "when complicating pregnancy."

Table 4 gives the numbers coded to each chapter by each system and the gross and net movements between chapters. These movements are important in assessing the changes introduced. The net movement between chapters is, in most cases, comparatively small so that final figures for, say, Chapter VII look much the same. The gross movements are much larger, however, and it is these that will require further study. Table 5 is a matrix of movements between chapters and is the first stage in the analysis of changes between chapters, ICD blocks, and 3-digit and 4-digit codes. A final analysis will be published and made available to users when a larger sample of the 45,000 records have been recoded.

Table 3. Percentage of deaths coded to the same ICD chapter that are also coded to the same ICD block, 3-digit code, and 4-digit code, by both manual and automated coding

Manual coding		Automated coding			
ICD chapter	Number of deaths	Number in same chapter	Percentage in same		
			ICD block	3-digit code	4-digit code
I	66	57	93	93	91
II	2,888	2,846	98	97	94
III	149	112	100	100	96
IV	31	26	100	85	81
V	338	312	97	90	89
VI	166	157	99	98	96
VII	5,041	4,904	97	93	93
VIII	1,539	1,467	97	96	95
IX	402	375	98	92	88
X	192	162	99	94	94
XI	1	0	0	0	0
XII	13	12	100	100	100
XIII	69	55	96	93	89
XIV	46	38	100	87	87
XV	37	31	100	94	94
XVI	64	60	100	83	82
SUPP	457	434	85	75	71
Total	11,499	11,048	97	94	92

Table 4. Deaths coded to ICD chapters by both systems showing gross and net movements between chapters

ICD chapter	Manual coding UCD	Automated coding UCD	Coded to same chapter by both	Coded to chapter by manual coding only	Coded to chapter by automated coding only	Net movement from manual to automated	% movement from manual to automated
I	66	97	57	9	40	31	47
II	2,888	2,865	2,846	42	19	-23	-1
III	149	130	112	37	18	-19	-13
IV	31	32	26	5	6	1	3
V	338	323	312	26	11	-15	-4
VI	166	177	157	9	20	11	7
VII	5,041	5,021	4,904	137	117	-20	0
VIII	1,539	1,548	1,467	72	81	9	1
IX	402	406	375	27	31	4	1
X	192	190	162	30	28	-2	-1
XI	1	0	0	1	0	-1	-100
XII	13	16	12	1	4	3	23
XIII	69	61	55	14	6	-8	-12
XIV	46	59	38	8	21	13	28
XV	37	34	31	6	3	-3	-8
XVI	64	66	60	4	6	2	3
SUPP	457	474	434	23	40	17	4
Total	11,499	11,499	11,048	451	451	0	0

Table 5. Deaths coded to chapters by manual and automated coding

Manual coding	Automated coding																	
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	SUPP	Total
I	57						1	1	1						3	2	1	66
II	4	2,846				1	9	6	5	2				11		2	2	2,888
III	16	1	112				15	2		2							1	149
IV		1		26	1		1	1	1									31
V			1		312	3	15	5		1							1	338
VI	1					157	3	4									1	166
VII	6	8	12	4	2	7	4,904	43	13	19		1	1	6			15	5,041
VIII	2	3	1		2	1	41	1,467	3			1	3	1		1	13	1,539
IX	3	1		1	4		10	3	375	3		1				1		402
X	3	4	1		1	1	4	8	3	162		1	1	2			1	192
XI							1											1
XII	1											12						13
XIII	3		2			1	3	3	1				55				1	69
XIV			1				1	2	3				1	38				46
XV				1		2	1			1				1	31			37
XVI																60	4	64
SUPP	1	1			1	4	12	3	1								434	457
Total	97	2,865	130	32	323	177	5,021	1,548	406	190	0	16	61	59	34	66	474	11,499

Appendix 1.

MEDICAL CERTIFICATE OF CAUSE OF DEATH FORM 11

This certificate is intended for the use of the Registrar of Births, Deaths and Marriages, and all persons are warned against accepting or using this certificate for any other purpose. See back of this form for notes about registration of a death.

To the Registrar of Births, Deaths and Marriages

Registrar to enter
District no . . . . . .
Year . . . . . .
Entry no . . . . . .

Name of deceased . . . . . .

Date of death	Day	Month	Year

Time of death . . . . . hours
(Enter approximate time if exact time not known)

Place of death . . . . . .

I hereby certify that to the best of my knowledge and belief, the cause of death and duration of disease were as stated below.

CAUSE OF DEATH *(PLEASE PRINT CLEARLY)*		Not to be entered in register: Approximate interval between onset and death		
I	I	years	months	days
Disease or condition directly leading to death*	a . . . . . *due to (or as a consequence of)*			
Antecedent causes Morbid conditions, if any, giving rise to the above cause, the underlying condition to be stated last	b . . . . . *due to (or as a consequence of)*			
	c			
II	II			
Other significant conditions contributing to the death, but not related to the disease or condition causing it	. . . . .			

* *This does not mean the mode of dying such as heart failure, asthenia, etc; it means the disease, injury or complication which caused death.*

Please ring the appropriate letter and appropriate figures:—

Certified cause takes account of post-mortem information . . . . . A
Information from post-mortem may be available later . . . . . B
Post-mortem not proposed C

Seen after death by me 1
Seen after death by another medical practitioner but not by me 2
Not seen after death by a medical practitioner 3

The deceased woman died during pregnancy or within six weeks thereafter 1
The deceased woman died between six weeks and twelve months after pregnancy 2

Please tick box if appropriate

I may be in a position later to give, if asked by the Registrar General, additional information as to the cause of this death for the purpose of more precise statistical classification ☐

Procurator Fiscal has been informed ☐

Signature . . . . . .

Date . . . . . 19 . . . . .

Name in BLOCK CAPITALS . . . . . .

Registered medical qualifications . . . . . .

Address . . . . . .

. . . . . .

For a death in hospital
Name of consultant responsible for deceased as a patient . . . . . .

Appendix 2.

Differences in coding between GRO(S) manual coding and automatic coding (USACS)

Part 1. Individual conditions

Condition	USACS	GRO(S)	Comment
1. Acute fatty degeneration of liver	571.8	570 (acute)	Micar dictionary ignores acute. *Query why not acute 570?*
2. Advanced neoplasia (primary site unknown)	239.9 unspecified	199.1 malignant	Neoplasia is unspecified *Query.*
3. Aortic and mitral valve disease.	424.0, 424.1 non-rheumatic	396 rheumatic	New ruling. Instruction manual 2b, p. 122, due to Ischaemic heart disease. Treat as non-rheumatic. ICD code 396 when disease. 424.0, 424.1 are disorders. *Query.*
4. Cerebrovascular accident over 1 year.	438 old	436	New. GRO(S) would only code 438 if there was a residual effect (eg hemi-plegia) and progression (another CVA) with duration originally 1 year+. *Accept.*
5. Chronic cardiac failure	428.9	428 chronic	Chronic not indexed. GRO(S) instruction to treat chronic as congestive. *Accept.*
6. Deep vein thrombosis	453.8	451.1	Normally coded to 453.9 by system. GRO(S) code to site in legs 451.1. *Query.*
7. Dementia	294.9 organic	298.9 as indexed	Auto appears to contradict index. Dementia NOS as indexed 298.9. 281 deaths coded to 298 in 1993. *Query.*
8. Drug abuse	305.9 non-dependent	304.9 dependent	304.9 correct. When with alcohol, alcohol selected as underlying cause. *Query.*
9. Ischaemic cardiomyopathy	414.8 Ischaemic heart disease	425.4 cardio-myopathy	Not a common term. Unlikely to affect numbers. *Accept.*

-

Condition	USACS	GRO(S)	Comment
10. Lower lobe pneumonia	486 pneumonia	481 Lobar	Lower lobe pneumonia a term in Micar dictionary. *Query 486.*
11. Lymphangitis carcinomatosis	199.0, 457.2 carcinoma and lymphangitis	196.9 carcinomatosis of lymph glands	Not indexed. Coded separately. *Accept.*
12. Malignant pleural effusion with a primary cancer	511.9 non-cancerous	197.2 secondary pleural cancer	GRO(S) consider 197.2 correct. *Query.*
13. Metastatic carcinoma when secondary	199.1 secondary neoplasm	199 spread	*Accept.*
14. Multi-infarct dementia	290.4 acteriosclerotic	294.1 dementia in conditions elsewhere	New. 294.1 is now an invalid code. Dementias previously coded there now coded to physical condition and 294.9. There were over 100 deaths coded to 294 in 1993. *Accept.*
15. Myelodysplasia	742.5 congenital	238.7 neoplasm uncertain behaviour	Indexed as 742.5. GRO(S) were advised by specialist to code 238.7. *Query.*
16. Myelodysplastic syndrome	289.8 other diseases of blood	238.7 as above	As above. *Query.*
17. Obstructive airways disease	496 chronic	519.8 other diseases of respiratory system	*Accept.*
18. Old pulmonary and renal tuberculosis	137.0, 016.0 old TB active	137.0, 137.2 both old	GRO(S) would accept both sites as old or inactive. USACS does not. *Query.*

Condition	USACS	GRO(S)	Comment
19. Perforated viscus	799.8 indexed	569.8 intestine	799.8 ill-defined but as indexed. GRO(S) coded to intestine as medical enquiry replies gave that site over all. *Accept.*
20. Pneumonias due to immobility	514 hypostatic pneumonia	486 pneumonia	New. GRO(S) preferred other codes on certificate to 514 unless ill-defined. *Accept.*
21. Previous myocardial infarction. No duration.	410, under 8 weeks	414.8, over 8 weeks	Previous implies over 8 weeks *Query.*
22. Previous myocardial infarction. 27 years duration.	410 as above	414.8 as above	Duration of 27 years. *Query.*
23. Septicaemic shock	785.5 as indexed	38.9 septicaemia	*Accept.*
24. Subarachnoid haemorrhage	431	430	Indexed as 430 but USACS appears to put any brain haemorhage to 431. Query.

Part 2. Differences in selection of underlying cause.

Cerebrovascular Conditions

1. 1a	Bronchopneumonia	485	No durations
1b	Immobility		
1c	Cerebrovascular disease	437.9	
II	Cerebrovascular accidents	436	

USACS 436 GRO(S) 437.9

Cerebrovascular disease (437.9) is ignored when mentioned with cerebrovascular accident (436) and 436 is selected as the underlying cause. The US developers are aware of this problem.

2. 1a	Bronchopneumonia	485	No durations
1b	Cerebrovascular disease	437.9	
1c	Left hemiplegia	342.9	
II	Dementia	294.9	

USACS 342.9 GRO(S) 437.9

Cerebrovascular disease (437.9) ignored and hemiplegia (342.9) selected as underlying cause.

3. 1a	Cerebrovascular accident	436	10 days
1b	Hypostatic pneumonia	514	10 days
1c	Ischaemic heart disease	414.9	60 years
II	Multi-infarct dementia	290.4	30 years
	Previous cerebrovascular accident	436	

USACS 414.9 GRO(S) 436

GRO(S) does not accept that ischaemic heart diseases cause cerebrovascular diseases except for embolism.

Chest infection

4. 1a	Chest infection	519.8	5 days
1b	Dementia	294.9 (organic)	10 years

USACS 519.8 GRO(S) 298.9 (in organic dementia)

5. 1a	Chest infection	519.8	No durations
1b	Immobility		
II	Cerebrovascular accident	436	
	Epilepsy	345.9	

USACS 519.8 GRO(S) 436

Chest infection 519.8 has been selected as underlying cause despite being due to immobility with a condition 436 in Part II of the certificate.

6. 1a Acute myocardial infarction 410 No durations
1b Bronchopneumonia 485

USACS 485 GRO(S) 410

Bronchopneumonia 485 selected as causing acute myocardial infarction 410. This is a possibly acceptable sequence. If it were to appear frequently the number of acute heart deaths could decrease but the frequency is doubtful.

Lymphoma

7. 1a Metastatic small cell carcinoma (primary source not identified) 199.1 No durations
II Lymphoma 202.8

USACS 202.8 GRO(S) 199.1

GRO(S) have coded to primary source not identified (199.1) but CARCINOMA now considered secondary to conditions classifiable to 200-203.

8. 1a Carcinomatosis 199.0 No durations
1b Melanoma right eye 198.4
1c Lymphoma 205.8

USACS 202.8 GRO(S) 190.9
(melanoma as primary)

Deep vein thrombosis

9. 1a Bilateral pulmonary thromboembosis 415.1 No durations
1b Deep vein thrombosis 415.1
1c Immobility due to fracture neck of femur 887
II Old cerebral infarction 438

USACS and GRO(S) both selected E887 as the underlying cause. Deep vein thrombosis

USACS 415.1 GRO(S) 451.1
(assumed deep vein thrombosis of leg)

10. 1a Renal failure 586 10 months
Pulmonary embolism 415.1
II Previous cerebrovascular accident 436
Previous deep vein thrombosis 434.0

USACS 486 GRO(S) 586

Deep vein thrombosis

USACS 434.0 GRO(S) 451.1
USACS associating DVT with cerebral site.

11. 1a Bronchopneumonia	485	7 days
II Left deep thrombosis	453.9	
Obesity	278.0	

USACS 485 GRO(S) 485

Deep vein thrombosis

USACS 453.9 GRO(S) 451.1
No site of DVT given.

Ischaemic heart disease

12. 1a Myocardial infarction	410	No durations
1b Pneumococcal pneumonia	481	
1c Acute renal failure	584.9	

USACS 584.9 GRO(S) 410

No sequence, should be 410?

13. 1a Myocardial infarction	414.8	10 months
1b Congestive cardiac failure; cardiomyopathy, arterial thrombosis	428.0, 425.4, 410	
II Thromboembolic Disease	444.9	

USACS 425.4 GRO(S) 410

14. 1a Cardiovascular failure	428.9	No durations
1b Myocardial infarction	410	
1c Renal failure, pneumonia	586	

USACS 586 GRO(S) 410

15. 1a Myocardial infarction	410	Same durations
1b Pulmonary embolism	415.1	
II Bronchopneumonia Thyrotoxicosis	485, 242.9	

USACS 415.1 GRO(S) 410

VARIOUS SEQUENCES

16. 1a Myocardial infarction	410	No durations
1b Myocardial ischaemia	414.8	
1c Anaemia	285.9	

USACS 285.9 GRO(S) 410

17. 1a Cardiorespiratory arrest	427.5	10 days
1b Cerebrovascular accident	43	10 days
1c Aspiration	91	10 days
II Ischaemic heart disease	414.9	

USACS 912 GRO(S) 436

18. 1a Stroke disease	438	6 years
1b Severe peripheral vascular disease	443.9	6 years

USACS 443.9 GRO(S) 436
(although 1 year +, no residual effect)

This could be an acceptable sequence with PERIPHERAL VASCULAR DISEASE causing MULTIPLE EMBOLI. But with the same durations GRO(S) think it more likely that both are due to GENERALIZED ARTERIOSCLEROSIS. Therefore 436 would have been what GRO(S) would have used.

19. 1a Bronchopneumonia	485	No durations
1b Dementia	294.9 (organic)	
1c Ischaemic Heart Disease	414.9	

USACS 414.9 GRO(S)to Dementia 298.9 (non-organic)

No sequence between b and c. GRO(S) would use DEMENTIA N.O.S. 298.9 (as indexed).

20. 1a Bronchopneumonia, renal failure	485, 586	No durations
1b Chronic obstructive airways disease	496	
1c Auricular fibrillation	427.3	

USACS 427.3 GRO(S) 496

Both conditions in 1(a) are Terminal - (b) and © are not sequences.

In examples 16-20 either the durations are the same or there are none. In every case the underlying cause has been the condition entered on the lowest line in Part I of the certificate whether there is an acceptable sequence of events or not.

Part 3. Additional codes created for automatic coding

USACS CODES		ICD-9 CODES
012.9	Tuberculosis N.O.S.	011.9
428.2	Arteriosclerotic myocarditis	429.0
428.3	Arteriosclerotic myocardial Degeneration	429.1
428.4	Arteriosclerotic Cardiovascular Disease	429.2
430.0	Any term in ICD-9 to 430	430
	Except	
430.1	Ruptured Cerebral Aneurysm	430
	And	
430.2	Ruptured congenital cerebral aneurysm	430
442.4	Congenital Aneurysm (Peripheral)	747.6
442.5	Congenital Aneurysm Brain (Arteriovenous)	747.8
487.9	Influenza NOS	487.1
518.9	Disease Lung (Chronic) NOS	519.8, 519.9
535.7	Haemorrhage, Duodenum	537.8
537.7	Disease, Stomach NOS	537.9
569.7	Perforation or Rupture Stomach	531.9, 537.8
570.0	Acute or Subacute necrosis of liver	570
	Any term indexed to 570	
	Except	
570.1	Acute Hepatic Failure	570
572.9	Hepatic Failure (Chronic)	572.8
582.6	Chronic Nephritis NOS	582.9
582.7	Chronic Nephropathy NOS	582.9
	Chronic Renal Disease NOS	582.9
799.7	Cause Unknown	799.9

Appendix 3.

ICD "BLOCKS"

I. INFECTIOUS AND PARASITIC DISEASES

Intestinal infectious diseases (001-009)
Tuberculosis (010-018)
Zoonotic bacterial diseases (020-027)
Other bacterial diseases (030-041)
Poliomyelitis and other non-arthropod-borne viral diseases of central nervous system (045-049)

Viral diseases accompanied by exanthem (050-057)
Arthropod-borne viral diseases (060-066)
Other diseases due to viruses and Chlamydiae (070-079)
Rickettsioses and other arthropod-borne diseases (080-088)

Syphilis and other venereal diseases (090-099)
Other spirochaetal diseases (100-104)
Mycoses (110-118)
Helminthiases (120-129)
Other infectious and parasitic diseases (130-136)
Late effects of infectious and parasitic diseases (137-139)

II. NEOPLASMS

Malignant neoplasm of lip, oral cavity and pharynx (140-149)
Malignant neoplasm of digestive organs and peritoneum (150-159)
Malignant neoplasm of respiratory and intrathoracic organs (160-165)
Malignant neoplasm of bone, connective tissue, skin and breast (170- 175)
Malignant neoplasm of genitourinary organs (179-189)

Malignant neoplasm of other and unspecified sites (190-199)
Malignant neoplasm of lymphatic and haematopoietic tissue (200-208)
Benign neoplasms (210-229)
Carcinoma in situ (230-234)
Neoplasms of uncertain behaviour (235-238)
Neoplasms of unspecified nature (239)

III. ENDOCRINE, NUTRITIONAL AND METABOLIC DISEASES AND IMMUNITY DISORDERS

Disorders of thyroid gland (240-246)
Diseases of other endocrine glands (250-259)
Nutritional deficiencies (260-269)
Other metabolic disorders and immunity disorders (270-279)

IV. DISEASES OF BLOOD AND BLOOD-FORMING ORGANS (280-289)

V. MENTAL DISORDERS

Organic psychotic conditions (290-294)
Other psychoses (295-299)
Neurotic disorders, personality disorders and other Mental retardation (317-319)

VI. DISEASES OF THE NERVOUS SYSTEM AND SENSE ORGANS

Inflammatory diseases of the central nervous system (320-326)
Hereditary and degenerative diseases of the central nervous system (330-337)
Other disorders of the central nervous system (340-349)
Disorders of the peripheral nervous system (350-359)
Disorders of the eye and adnexa (360-379)
Diseases of the ear and mastoid process (380-389)

VII. DISEASES OF THE CIRCULATORY SYSTEM

Acute rheumatic fever (390-392)
Chronic rheumatic heart disease (393-398)
Hypertensive disease (401-405)
Ischaemic heart disease (410-414)
Diseases of pulmonary circulation (415-417)

Other forms of heart disease (420-429)
Cerebrovascular disease (430-438)
Diseases of arteries, arterioles and capillaries (440-448)

VIII. DISEASES OF THE RESPIRATORY SYSTEM

Acute respiratory infections (460-466)
Other diseases of upper respiratory tract (470-478)
Pneumonia and influenza (480-487)
Chronic obstructive pulmonary disease and allied conditions (490-496)
Pneumoconioses and other lung diseases due to external agents (500- 508)
Other diseases of respiratory system (510-519)

IX. DISEASES OF THE DIGESTIVE SYSTEM

Diseases of oral cavity, salivary glands and jaws (520-529)
Diseases of oesophagus, stomach and duodenum (530-537)
Appendicitis (540-543)
Hernia of abdominal cavity (550-553)
Noninfective enteritis and colitis (555-558)

Other diseases of intestines and peritoneum (560-569)
Other diseases of digestive system (570-579)

X. DISEASES OF THE GENITOURINARY SYSTEM

Nephritis, nephrotic syndrome and nephrosis (580-589)
Other diseases of urinary system (590-599)
Diseases of male genital organs (600-608)

Disorders of breast (610-611)
Inflammatory disease of female pelvic organs (614-616)
Other disorders of female genital tract (617-629)

XI. COMPLICATIONS OF PREGNANCY, CHILDBIRTH AND THE PUERPERIUM

Pregnancy with abortive outcome (630-639)
Complications mainly related to pregnancy (640-648)
Normal delivery, and other indications for care in pregnancy, labour and delivery (650-659)
Complications occurring mainly in the course of labour and delivery (660-669)
Complications of the puerperium (670-676)

XII. DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE

Infections of skin and subcutaneous tissue (680-686)
Other inflammatory conditions of skin and subcutaneous tissue (690-698)
Other diseases of skin and subcutaneous tissue (700-709)
Arthropathies and related disorders (710-719)
Dorsopathies (720-724)

Rheumatism, excluding the back (725-729)
Osteopathies, chondropathies and acquired musculoskeletal deformities (730-739)

XIV. CONGENITAL ANOMALIES (740-759)

XV. CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD (760-779)

XVI. SYMPTOMS SIGNS AND ILL-DEFINED CONDITIONS

Symptoms (780-789)
Nonspecific abnormal findings (790-796)
Ill-defined and unknown causes of morbidity and mortality (797-799)

SUPPLEMENTARY CLASSIFICATION OF EXTERNAL CAUSES OF INJURY AND POISONING

Railway accidents (E800-E807)
Motor vehicle traffic accidents (E810-E819)
Motor vehicle non traffic accidents (E820-E825)
Other road vehicle accidents (E826-E829)
Water transport accidents (E830-E838)

Air and space transport accidents (E840-E845)
Vehicle accidents not elsewhere classifiable (E846-E848)
Accidental poisoning by drugs, medicaments and biologicals (E850-E858)
Accidental poisoning by other solid and liquid substances, gases and vapours (E860-869)

Misadventures to patients during surgical and medical care (E870-876)
Surgical and medical procedures as the cause of abnormal reaction of patient or later complication, without mention of misadventure at the time of procedure (E878-E879)
Accidental falls (E880-E888)
Accidents caused by fire and flames (E890-E899)
Accidents due to natural and environmental factors (E900-E909)

Accidents caused by submersion, suffocation and foreign bodies (E910-E915)
Other accidents (E916-E928)
Late effects of accidental injury (E929)
Drugs, medicaments and biological substances causing adverse effects in therapeutic use (E930-E949)
Suicide and self-inflicted injury (E950-E959)

Homicide and injury purposely inflicted by other persons (E960-E969)
Legal intervention (E970-E978)
Injury undetermined whether accidentally or purposely inflicted (E980-E989)
Injury resulting from operations of war (E990-E999)

Impact of Automated Coding in Australia

John Donovan, M.D., B.S., Australian Institute of Health and Welfare

Australia has adopted automatic coding for deaths registered from 1 January 1997. This assessment of its likely impact on statistics is based on a review of records from trials of automated cause of death coding conducted in 1995. The assessment updates earlier ones distributed at the October 1996 (Tokyo) meeting of Heads of WHO Collaborating Centers for Classification of Disease and at the November 1996 ICE on Automating Mortality Statistics.

In the trials, 13,907 medical certificates of cause of death were coded both automatically and manually. The automatic coding was done first, using the systems MICAR, SuperMICAR, and ACME as supplied by the U.S. National Center for Health Statistics (NCHS). At the time of this study, information on cause of death was lacking for 320 deaths, which were almost all under coroner investigation.

The manual coding was done by two groups working separately on different medical certificates, and their work has been combined for this report. There were 13 missing manually allocated codes; 8 of which lacked information at the time of automatic coding, so there were 13,574 deaths where codes from both methods were available for comparison. Except where stated, these 13,574 deaths provide the source data for the discussion that follows. The age distribution of the deaths with complete data is shown in table 1.

Table 1. Numbers of deaths available for study and percentages with different codes for underlying cause of death, by age group

Age	Codes different	Valid pairs	Percentage different codes
Under 1	71	137	52%
1 to 74	1,073	5,706	19%
75 and over	1,384	7,731	18%
Total	2,528	13,574	19%

The manual coding used existing Australian Bureau of Statistics (ABS) interpretations of the International Classification of Diseases manual. It also used long-standing ABS rules for when queries were to be made of certifying doctors. The manual codes were adjusted to reflect the responses to these queries. The codes allotted automatically were not adjusted in this way, and that difference in procedures contributed greatly to the 2,528 instances where different underlying causes resulted. The numbers of queries and a summary of responses are shown in table 2. Queries were also made in respect to the 320 deaths referred to earlier, where information was not available at time of coding; these are not shown in the table.

Table 2. Frequency of queries of certifying doctors and response categories

Query action	Number of cases
None	12,123
Queried, no response or response giving no new information	100
Queried, response changes fourth digit only of code for underlying cause	176
Queried, response changes first three digits of code for underlying cause	959
Death certificate would be queried except states that further information not available (code used for editing file)	223

The effect of queries is taken into account in the discussion on individual causes of death that follows. It should also be noted that ABS practice is not to make queries when there is no potential for changing the code at the third digit level. Thus the wording “bowel cancer,” 159.0, will be queried as to primary site, usually with amendment to a four-digit code in categories 153, 154, or 152. However, “cancer of colon,” 153.9, is not queried as the potential replacement code is another category in 153. Australian practice with queries also varies with age of the deceased, and this has influenced the data shown above. Queries are not conducted for deaths to individuals age 75 and over, except with respect to certain causes; those relevant to this report include neoplasms, gastroenteritis, and trauma.

Because findings in this paper for the automatic coding may, particularly at ages under 75, have been influenced by the inability to allow for queries of certifiers, the National Center for Classification in Health (Brisbane) and the Australian Bureau of Statistics plan to repeat this study, coding by both methods all deaths registered in Australia in January and July 1997.

In accordance with ABS practice of right-adjusting codes, three digit codes that lack further subdivision in ICD-9 have ".9" added. Thus, chronic renal failure, 585 in ICD-9, is shown as 585.9. Except where stated, all other codes used in this paper follow the ICD-9 Manual. Some longer titles are paraphrased.

Diarrhoea, etc.

This cause of death is queried (infectious or not) at all ages. Even though Australia is surely, in the terms of the ICD–9 manual, a country where diarrhoea "can be presumed to be of noninfectious origin" (page 317, Vol 1), ABS enquiries about diarrhoea deaths in the elderly almost always yield the answer that it is infectious. An example is:

Male, 90

I(a)	Dehydration
I(b)	Diarrhoea
I(c)	
II	Profound chronic dementia with immobility to emaciation (sic)

There were six such occurrences in the sample, all coded automatically to 558.9, but manually to 005.2, 008.8 (2), 009.0 and 009.1 (2). There were three other deaths coded to intestinal infections both manually and automatically, and one coded to this area manually but automatically to a mycobacterial infection.

For the present, ABS is continuing its policy of querying diarrhoea, gastroenteritis, and related causes of death, but if querying is ceased there will be a transfer of most of these deaths from infectious causes to noninfectious.

Septicaemia

Manual coding yielded 59 deaths to Septicaemia, automatic coding 76 deaths, and in 51 of these both methods yielded the same four-digit codes. Nine of the 25 "gains" from conversion to automatic coding and two of the eight "losses" were from and to gastrointestinal disorders. However, the changes varied greatly with age. It appears that introduction of automatic coding will cause a substantial increase (ratio 32:10) in the number of deaths ascribed to septicaemia at ages up to 74, but a much smaller increase (ratio 52:46) at higher ages (table 3). This age effect will occur because death certificates

where manual coding gave an underlying cause of septicaemia have been queried and hence liable to change at ages up to 74, but not at ages 75 and over.

Table 3. Frequency of allocation of septicaemia as underlying cause by age at death and by method of coding

Coding method / Age	Manual	Automatic	Both
Under 1	–	1	–
1 to 74	13	23	10
75 and over	46	52	41
Total	59	76	51

HIV and AIDS

Australia has used the codes 279.1 for AIDS itself and E875.0 for transfusion associated HIV, but the automatic system uses the codes ranging from 042 to 044, which were adapted from those in ICD-9-CM. Until 1994, ABS also interpreted the notes on “highly improbable” ((a) and (b) on pages 721–2, Vol 1) as applying to AIDS. From 1995, this was abandoned. Thus, until 1995 a certificate such as the following was attributed to cryptococcal meningitis,
but from 1996 to AIDS:

Male, 42

I(a) Cryptococcal meningitis
I(b) Human immunodeficiency virus infection
I(c)
II Cerebral lymphoma

Manual coding to 279.1 yielded 44 deaths. Automatic coding to 042-044 yielded 52 deaths, and 38 were allotted to these codes by both methods. The 14 gains with the transfer to automatic coding included eight from opportunistic infections, one from transfusion associated HIV, and five from malignancies. All of these malignancies except one lung cancer and one Burkitt’s tumour were frequent complications of AIDS.

One case, with the underlying cause coded manually to AIDS, but coded automatically to pulmonary infiltration, 518.3, is interesting. It was not queried. The intent of the certifier is clear, even if the wording could be improved. A minor improvement in the automatic system would be to accept pulmonary infiltration as a consequence of AIDS:

Male, 39

I(a) Pulmonary infiltration
I(b) Acquired immune deficiency syndrome
I(c)
II Kaposi's sarcoma

Earlier in the AIDS epidemic, Australian death certificates commonly showed complications such as opportunistic infections but did not mention HIV or AIDS. This is no longer the case. There were only four certificates that might have come into this category, one death from disseminated fungal infection and three from Pneumocystis carinii pneumonia.

Neoplasms

Manual coding yielded 3,613 deaths, automatic coding 3,523 deaths, and 3,500 of these were in the range 140.0 to 239.9 in both cases. Within the 3,500 there were many coding differences, which reflect responses to queries of certificates using terms such as "disseminated cancer" or "cerebral tumour" where more information was available for manual than for automatic coding. There were 73 deaths where the automatically coded cause was intestinal cancer, 159.0, all but 10 of which were manually assigned to more specific sites following query. Similarly, there were 20 instances of cerebral tumour, 239.6, all but 4 of which were able to be assigned later as benign or malignant. However, there was a much smaller proportionate reduction from 267 to 192 in the number of cancers without specification of site and coded to 199. Apart from these differences resulting from queries, there were few differences between manual and automatic codes relating to deaths where both methods classified the underlying cause as a neoplasm.

Some common wording, "Liver secondaries," with which the automatic system had difficulty is illustrated by this poorly written certificate, coded automatically to 155.2, malignant neoplasm of liver not specified as primary or secondary:

Male, 70

I(a) Adenocarcinoma of rectosigmoid
I(b) Liver secondaries
I(c)
II

There were 107 instances where the manually allocated cause of death was a neoplasm but the automatic allocation was not. These included 20 cases of myelodysplasia or myelodysplastic syndrome and related disorders that were classified manually as neoplasms of uncertain nature, 238.7, but which the automatic system classified elsewhere depending on the wording. "Myelodysplasia" was classified by the automatic system as a malformation in accordance with the ICD-9 index; this problem is adjusted in ICD-10. The other major losses with the change were to pneumonia (22), to trauma (18), and to gastrointestinal disorders (12). The losses to pneumonia reflect

different interpretations of Rule 3. An example is given later under Pneumonia. The losses to trauma were disparate, but two examples were:

Male, 72

I(a)	Pneumonia
I(b)	Infected thoracic wound/metastases
I(c)	Adenocarcinoma unknown primary
II	Chronic bronchitis Peptic ulcer disease

Female, 75

I(a)	Ruptured uterus
I(b)	Sepsis
I(c)	
II	Carcinoma of uterus

My assessment of these certificates is that the certifier intended the malignancy to be the underlying cause of death in each case and that the manual coding was correct, with application of Rule 3. The decision tables in the automatic system might be reviewed. The losses to gastrointestinal disorders are mainly from gastrointestinal cancers, but two are from breast cancer and one from lung cancer to unrelated gastrointestinal disorders. There were 19 miscellaneous instances where the automatically allotted underlying cause was a neoplasm but the manually selected cause was not. One was an interesting example of the system capturing the intention of the certifier in a death certificate with a long sequence of events:

Female, 84

I(a)	Cardiac arrest (seconds)
I(b)	Acute on chronic heart failure (2 days)
I(c)	Acute small bowel obstruction (3 days)
II	Small bowel adhesions due to radiotherapy Cervical cancer (5 years)

The automatic coding attributed this death to cervical cancer, manual coding to heart failure.

Diabetes

In certificates where information was available from both coding systems, manual coding yielded 300 deaths to diabetes, automatic coding 275 deaths, with 255 of these coded to the same four-digit codes and eight coded to different four-digit codes. The changes in frequency did not appear to relate to age at death, and hence to different use of queries before and from age 75 (table 4).

Table 4. Frequency of allocation of diabetes mellitus as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
1 to 74	132	123	116
75 and over	168	152	147
Total	300	275	263

After the change to automatic coding, diabetes mortality will decrease by about 8% to 275/300 of its present level. The main losses with the change were to cardiovascular diseases (11) and pneumonia (7); the gains were mainly from cardiovascular diseases (10). These changes are very different from those observed in the U.S., and reflect different applications of Rule 2 and Rule 3. An example is:

Male, 68

I(a) Renal failure
I(b) Hypertension
I(c) Non-insulin dependent diabetes
II Ischaemic heart disease

This certificate was assigned manually to 414.9 and automatically to 250.0, but I would not accept hypertension as a direct complication of diabetes. The certifier's intention is not clear to me, and I would have wished to query this certificate.

Cachexia, marasmus, malnutrition, etc. (261.9 and 263.9)

Present Australian practice is that these terms are not accepted as underlying causes of death at ages 0 to 74 without query. They were used as underlying cause of death 13 times by both coding systems, 10 of these instances being in common. In the three instances where these underlying causes were allotted automatically but not manually, the manually allotted causes seem to me to have been those intended by the certifiers (one case each of alcoholic liver disease, anorexia nervosa, and depression). In the three cases where the change of codes was from a manually selected underlying cause of malnutrition, the automatic coding was to an equally non-specific underlying cause (one case each of pneumonia, aspiration pneumonia, and anaemia, all in persons over 80).

Perhaps a more important consideration is that terms such as cachexia or marasmus appear on medical certificates in many wasting diseases. This certificate is an interesting example:

Male, 22

I(a)	Cardiac failure
I(b)	Anaemia
I(c)	Malnutrition
II	Machado-Joseph disease [hereditary spino-cerebellar degeneration]

The words in brackets were on the certificate. Both the manual coding and the automatic system attributed this to malnutrition, 263.9. A query was made whether the disabling disease in Part II caused the malnutrition but there was no response. To me it was the obvious underlying cause.

In ICD-10 coding malnutrition will be the "assumed direct consequences of another condition" (new Rule 3, Vol. 2, page 39) for malignancy. However, it can follow other debilitating conditions such as those mentioned four paragraphs above, and it is not clear to me whether cachexia, etc., are trivial conditions for the purposes of Modification Rule B, but if that is the interpretation, the queries might be dispensed with as ICD-10 is adopted. The Australian Collaborating Center for Classification of Diseases is recommending to other Centers that the Modification Rule be clarified.

Dementia

ABS has several longstanding local rulings relating to coding of dementia. "Alzheimer's disease," so worded, is coded to 331.0 following the ICD-9 manual. The code 290.1 for "Alzheimer's dementia," so worded, is not used after age 65, and "dementia" and "senile dementia" are coded to 290.0, 290.1, or 298.9 according to age. Multi-infarct dementia is coded to 434.9 rather than 290.4 as in the automatic system; manual coding of other arteriosclerotic dementias has not been consistent, with some use of 290.4 and some use of codes in the range 436 to 440. The coding of the most frequently used terms is shown in table 5.

Table 5. Comparison of Australian and U.S. codes for frequently used terms relating to dementia

Wording	Australia	United States
Alzheimer's dementia 0-64	290.1	290.1
Alzheimer's dementia 65+	290.0	290.1
Senile dementia 0-64	290.1	290.0
Senile dementia 65+	290.0	290.0
Dementia 0-64	298.9	298.9
Dementia 65+	290.0	298.9
Multi-infarct dementia	434.9	290.4
Alzheimer's disease	331.0	331.0

As well as differences in coding, there are differences in application of Rule 3 to items in Part II. The single most common cause of difference between the present Australian and the automatic systems in underlying cause selections is in certificates such as this:

Male, 74

I(a) Pneumonia
I(b)
I(c)
II Alzheimer's disease

To us, this has been coded to Alzheimer's disease, after query at ages 0 to 74 or accepted at 75 and over, but in the automatic system it is coded as pneumonia. In the sample, there were 123 manually coded underlying causes of Alzheimer's disease (331.0), 104 automatically coded, and 103 by both methods; 17 of the 19 losses with the change to automatic coding were to pneumonia.

The incidence of dementia as an underlying cause depends heavily on what ICD categories are regarded as dementia, but under our present rules the greatest single number are coded to 290.0. To illustrate how greatly automatic coding will affect our statistics relating to dementia, there were 250 manual allocations of 290.0 as the underlying cause; the automatic allocations of these included senile dementia, 290.0, (45), pre-senile dementia, 290.1, (45, worded as "Alzheimer's dementia" at age 65 and over), dementia unspecified, 298.9, (63) and pneumonia of various types (64). Table 6 shows the frequency of all underlying cause codes related to dementia, by age group. It can be used to predict changes in reported mortality from these categories.

Table 6. Frequency of allocation of dementia codes (290.0, 290.1, 298.9, 331.0) as underlying cause by age at death and by method of coding

Age	1–64		65–74		75 and over	
Coding	Manual	Auto	Manual	Auto	Manual	Auto
290.0	–	–	18	2	232	44
290.1	1	1		11	–	38
290.2	–	–	–	–	1	1
290.4	–	–	–	4	16	33
298.9	–	1	–	4	–	61
331.0	2	2	13	10	108	92
434.9	8	6	22	16	87	65

Other diseases of nervous system

Parkinson's disease, 332.0, was the underlying cause of death in 70 manually coded cases, but only in 57 automatically coded cases; 55 of these certificates were coded to Parkinson's disease by both methods. The losses, to pneumonia, were similar below and from age 75, and thus do not appear to be affected by the use of queries to age 74. A reduction in mortality attributed to Parkinson's disease can thus be expected in 1997 data. There were 25 deaths attributed to epilepsy NOS by manual coding and 22 by automatic coding, with 20 deaths in common. Again, the losses were to pneumonia.

There were 34 deaths attributed to motor neuron disease, 335.2, by manual coding and 29 by automatic, with 28 in common. The losses were to other paralytic syndromes in 344.8 (1 case) and 344.9 (3). The last three are cases of Steele-Richardson syndrome, which is not indexed in ICD-9, is automatically coded to 344.9, and has been coded differently by ABS at different times. This anomaly will be rectified with the introduction of ICD-10; the condition is indexed in that revision.

Rheumatic heart disease

Manual coding yielded 37 deaths, automatic coding 30, 27 being to the same four-digit codes. There were two instances of different four-digit codes within the range 390 to 398. The losses (5) were mainly to other cardiac disorders, none occurring more than once, and only one valvular (this certificate of a death at age 82 was not queried).

Hypertension

Manual coding yielded 125 and automatic coding 123 deaths with hypertension as the underlying cause. There were 99 instances of the same four-digit codes within the range 400 to 405 and seven of different third or four-digit codes. The greater use of queries of certifying doctors at ages below 75 does not seem to have influenced these findings (table 7).

Table 7. Frequency of allocation of hypertension as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
Under 1	–	1	–
1 to 74	31	31	26
75 and over	94	91	80
Total	125	123	106

The losses with the change to automatic coding were mainly to other heart disease (6), stroke (3), and pneumonia (3), but 5 of the 19 gains were from renal disease. This reflects different applications of Rules 2 and 3, but these are not so great as to alter the prominence of hypertension as an underlying cause of death.

The importance and usefulness of mortality statistics of hypertension will increase with the introduction of automatic coding, but there are even greater advantages with the introduction of multiple cause tabulations that automatic coding permits. In both ICD-9 and ICD-10, hypertension is not used as an underlying cause of death if many other conditions, including ischaemic heart disease, are also mentioned on the medical certificate; the conditions are listed in ICD manuals.

Ischaemic heart disease

For code 410, as a whole, there were 2,030 underlying causes allotted manually, 1,998 allotted automatically, and 1,918 in common. Fourth digit differences are considered later in this section. With so many deaths it was possible to demonstrate that querying practices may have influenced the findings. At ages 1 to 74 there were 40 manual but only 21 automatic selections of underlying cause not replicated by the other method (table 8). At ages 75 and over these numbers were nearly equal at 56 and 57, respectively. No single diagnosis could be identified as accounting for the relative deficit below age 75 (or perhaps for a surplus at 75 and over) of death certificates with automatic codes of 410.9, but with different manual codes.

Table 8. Frequency of allocation of acute ischaemic heart disease as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
1 to 74	731	712	691
75 and over	1,283	1,284	1,227
Total	2,014	1,996	1,918

The main losses from manual coding were to diabetes (4), valve disorders (4), myocarditis (5), pneumonia (13), and chronic lung disease (5). The single most important source of gain with automatic coding was from heart valve disorders, codes 424.0 to 424.9, (18 instances, 13 from 424.1); next was dementia, 290.0, (5 cases).

For code 414 there were 1,267 manual allocations of underlying cause and 1,224 automatic allocations, 1,137 being in common at the third digit level. As with code 410.9, it was possible to demonstrate that querying practices may have influenced the findings. At ages 1 to 74 there were similar numbers, 37 manual and 34 automatic, of selections of underlying cause not replicated by the other method (table 9). At ages 75 and over these numbers were quite different at 93 and 53, respectively. Among the deaths manually selected to 414, there were 4 deaths under age 75 and 33 deaths at age 75 and over that were automatically assigned to pneumonia. Similarly, among the deaths manually selected to 414, 1 death under age 75 and 8 deaths at age 75 and over were automatically assigned to pulmonary embolism. Within category 414 there was much greater manual use of code 414.0, as table 10 shows. The explanation for this is that Australia uses 414.0 rather than 414.9 when atherosclerosis is mentioned on the certificate.

Table 9. Frequency of allocation of chronic ischaemia heart disease as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
1 to 74	412	409	375
75 and over	855	815	762
Total	1,267	1,224	1,137

Table 10. Frequency of allocation of four-digit categories of chronic ischaemic heart disease as underlying cause by method of coding

Automatic / Manual	414.0	414.1	414.8	414.9
414.0	224		5	38
414.1		1		
414.8			106	10
414.9			2	751

For code 414, the major losses with the automatic coding were to diabetes (4), pulmonary embolism (8), cardiomyopathy (5), dysrhythmia (4), heart failure (5), and pneumonia (37). The major gains were from diabetes (5), valve disorders (10), heart failure (5), and chronic respiratory disease (10).

Pulmonary embolism

The historical importance of this condition is that when an increase in mortality of young women was noticed in the 1960's and the death certificates were examined, it was found that there were many mentions of embolism, but that the deaths had been coded to other underlying causes. The problem appears to be still with us, at least for pulmonary embolism. This was the underlying cause of death in 19 manually coded certificates and in 63 automatically coded. All 19 manual codes were repeated with automatic coding. Table 11 shows that the difference between the numbers of deaths attributed to pulmonary embolism by the two coding systems does not appear to differ with age group. All but one of the certificates at ages up to 74 were queried, but none of those at ages 75 and over. The conclusions are that for pulmonary embolism, the differences in statistics that result from the two systems are due to interpretation of the Rules, and that the querying of certificates up to age 74 is of little value.

Table 11. Frequency of allocation of pulmonary embolism as underlying cause by age at death and by method of coding

Coding method / Age	Manual	Automatic	Both
1 to 74	7	28	7
75 and over	12	35	12
Total	19	63	19

Causes of death on manually coded certificates that were automatically coded to pulmonary embolism included a range of malignant disorders (5), ischaemic heart disease, 410 to 414, (11), aortic valve disease (2), thrombosis of deep leg veins (8), asthma (2), and chronic obstructive lung disease (2). Valvular disease and thrombosis are included in the list of conditions for which Rule 3 should be applied in ICD-10, but the others are not; I do not think the list needs to be changed. Multiple cause tables from 1997 should be monitored in case the larger sample displays other associations where the automatic coding does not lead to pulmonary embolism becoming the underlying cause.

Heart valve disorders (424)

The classification of non-rheumatic mitral valve disorders, 424.0, will be slightly affected by the change to automatic coding. This was the underlying cause from manual coding 16 times, and from automatic coding 19 times, including 15 times from both methods. There were 3 instances where manual codes of 424.0 were coded automatically to 410.9.

The classification of non-rheumatic aortic valve disorders, 424.1, will be more substantially affected. This code was the underlying cause from manual coding 69 times and from automatic coding 48 times, including 39 times from both methods. There were 20 instances where manual coding was to aortic valve disease and automatic coding to ischaemic heart disease, but only 6 where the change was in the opposite direction. The next certificate was coded manually to aortic stenosis and automatically to ischaemic heart disease:

Female, 82

I(a)	Severe acute pulmonary oedema due to aortic stenosis
I(b)	Ischaemic heart disease due to coronary atheroma
I(c)	
II	Previous acute myocardial infarction

I prefer the manual interpretation as a reflection of the intention of the certifier, as I understand that. The next certificate was coded manually to 424.1 and automatically to 414.8. Again I prefer the manual interpretation:

Male, 83

I(a)	Cardiac arrest
I(b)	Myocardial ischaemia
I(c)	Aortic stenosis
II	Emphysema

In this next certificate, manually coded to ischaemic heart disease, I prefer the automatic assignment of aortic valve disease as the underlying cause:

Female, 82

I(a)	Acute pulmonary oedema
I(b)	Left ventricular failure
I(c)	Aortic stenosis and Ischaemic heart disease
II	

I think we need to reconsider the decision tables relating to aortic valve disease.

Atrial fibrillation

There were 51 deaths attributed manually to Atrial fibrillation and 46 automatically, with 38 attributed to both methods. Its importance from the viewpoint of classification is that cerebral and other arterial embolism are frequent complications of it, and in ICD-10, Rule 3 will be applied to these complications. Cerebrovascular disease was frequently mentioned on these death certificates, but cerebral and mesenteric embolism only rarely. This certificate suggests cerebral embolism even though the condition was not mentioned:

Female, 83

I(a)	Pneumonia
I(b)	Congestive cardiac failure
I(c)	Atrial fibrillation
II	Left arm monoplegia

The great majority of the transfers to and from atrial fibrillation were to or from other cardiovascular diseases.

Heart failure

Heart failure is of decreasing importance as a cause of death in Australia with the relative declines more rapid at ages up to 75, where the certificate yielding heart failure as the underlying cause of death is queried, than at ages over 75, where it is not. Thus, the great majority of deaths (90% in manual codes in the study sample) attributed to cardiac failure are at ages over 75.

Manual coding yielded 329 deaths, automatic coding 301 deaths, and 249 of these were identical four-digit codes. Within category 428, there were 14 more with changes at the fourth-digit within category 428, many of which were due to biventricular failure coded manually as 428.1, whereas the automatic coding choice of 428.9 is preferable and compatible with ICD-10 where the condition is indexed. There were 29 transfers to pneumonia with automatic coding, and chronic obstructive lung disease was the single most important source (12) of gains. The decision tables may benefit from review, as this certificate coded manually to 496.9 and automatically to 428.9 seems to me to have been correctly coded manually:

Male, 71

I(a)	Bronchopneumonia
I(b)	Chronic obstructive airways disease
I(c)	Congestive cardiac failure
II	Carcinoma of prostate

Cerebrovascular disease

In the sample, there were 1,377 deaths with manual coding and 1,294 with automatic coding. There were 1,170 with the same four-digit codes and 76 more with different codes within the range 430 to 438. The proportions of death certificates ascribed to cerebrovascular disease under one system, but not the other, was the same before and past age 75 years (table 12). There were 38 transfers to pneumonia with automatic coding, but there was no clear main source of gains.

Cerebral haemorrhage was the underlying cause of death in 169 manually coded and 159 automatically coded certificates, 148 being in common. The automatic system coded as cerebral haemorrhage two perinatal deaths that had been coded manually to respiratory distress syndrome and to intraventricular haemorrhage.

The changes noted with pulmonary embolism do not apply to cerebral embolism. This was the cause of death in eight certificates coded manually and in eight certificates coded automatically; five of these were in common. Only one of these cases mentioned atrial fibrillation. In two cases the code 434.1 seems to be in error. In one there was incorrect keying of a manual code of 424.1. In the other the assignment was by the automatic system, from the wording “arterial occlusion”:

Male, 72

I(a)	Cerebrovascular accident
I(b)	Arterial occlusion
I(c)	Atheroma
II	Parkinson’s disease Postural hypotension Vomiting

This may result from the wording in the ICD-9 index, page 380, which reads,
"Occlusion, artery—see also Embolism artery." The ambiguity is corrected in the index to ICD-10.

Table 12. Frequency of allocation of stroke as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
Under 1	–	3	–
1 to 74	301	288	275
75 and over	1,076	1,003	971
Total	1,377	1,294	1,246

There is substantial movement between other codes and 438 (late effects of cerebrovascular disease), as shown in table 13. This shows that Australia has been using different interpretations from the United States. However, the differences might be better pursued in relation to ICD-10 than by detailed examination of differences under ICD-9.

Other cardiovascular disorders

Aortic aneurysm, 441, was the underlying cause on 167 manually coded certificates and on 168 automatically coded, 160 of these were in common at the third digit level, 151 at the fourth-digit level. Most of the changes were to and from other cardiovascular diseases. Manual coding classified 76 deaths to cardiomyopathy, 425, mainly to unspecified primary cardiomyopathy, 425.4, (62). There were 85 deaths classified by automatic coding. There were 62 deaths coded to the same four-digit category and a further six to different four-digit categories within 425. With the change to automated coding, all but one of the losses were to ischaemic heart disease. Only three of the gains were from ischaemic heart disease.

Deaths are not commonly classified to conduction disorders, but there were 11 by manual coding and 14 by automatic. There were seven deaths classified to the same four-digit codes, but four where the wording “complete heart block” was classified manually to complete atrio-ventricular block, 426.0, (correct) and automatically to unspecified heart block, 426.9. This is due to an index deficiency in ICD-9, corrected in ICD-10.

Atherosclerosis, 440.9, was the underlying cause of death in 71 manual allocations and 65 automatic, with 59 in common. The gains and losses were both from and to other cardiovascular diseases. There were 80 deaths manually coded to 443.9, peripheral vascular disease NOS, 86 coded automatically, and 70 in common.

Manual coding resulted in 14 deaths attributed to leg vein thrombosis and thrombophlebitis, 451.1, but only four deaths were automatically coded to this cause, and three of these matched the manual coding. Eight of the losses were to pulmonary embolism: under ICD-10 such deaths will again be attributed to thrombosis, by Rule 3.

Table 13. Frequency of allocation of late effects of cerebrovascular disease as underlying cause by method of coding

Automatic / Manual	434.9	436	438	Elsewhere 430–438	Elsewhere
434.9	79	2	9	1	26
436	3	648	15	8	37
438	2	14	48	1	25
Elsewhere 430–438	1	3	4		
Elsewhere	2	13	7		

Pneumonia

This is another area where major change will occur, first with the adoption of automatic coding and again in 1999 with the change to ICD-10.

Manual coding yielded 210 deaths to Pneumonia, automatic coding 560 deaths, and 175 of these were in common at the four-digit level. Another 20 certificates were automatically coded to 485.9 or 486.9, manually coded to pneumococcal (lobar, 13) or other (7) forms of pneumonia. The increase in death rates with the change to automatic coding will be about 50% at ages up to 74 but about 200% at ages 75 and over (table 14).

Table 14. Frequency of allocation of pneumonia as underlying cause by age at death and by method of coding

Coding method / Age	Manual	Automatic	Both
Under 1	–	1	-
1 to 74	70	106	64
75 and over	140	453	131
Total	210	560	195

The automatic system has difficulty coding (lobar) pneumonia of a specified lobe, as in this example:

Male, 77

I(a) Cardiac failure
I(b) Right lower lobe pneumonia
I(c)
II Renal failure

The manual code in this example was 481.9 and the automatic code was 486.9. Major sources of gains in pneumonia deaths with the transfer to automatic coding were from neoplasms (22), senile dementia (64), Alzheimer's disease, as worded (17), ischaemic heart disease (52), cardiac failure (29), stroke (38), and chronic obstructive lung disease (22). The transfers among, to and from pneumonia categories are summarized in table 15.

Table 15. Frequency of allocation of forms of pneumonia as underlying cause by method of coding

Automatic Manual	480–483	485	486	Elsewhere
480–483	32	4	14	6
485		53	2	7
486			90	9
Elsewhere	20	147	199	

Under the broader Rule 3 in ICD-10, pneumonia can be regarded as due to any other condition:

Male, 46, Aboriginal

I(a) Respiratory failure
I(b) Pneumonia - organism not identified
I(c)
II Squamous cell carcinoma of tonsil

This certificate was coded both manually and automatically to 486.9, but my own assessment is that despite a response to a query confirming the certificate as written, the pneumonia was highly likely to have been due to the cancer. Under ICD-10 this certificate would be classified to the cancer without query. It needs to be realized that allowing pneumonia to be due to anything is not necessarily helpful. Consider this certificate:

Female, 91

I(a) Bronchopneumonia
I(b)
I(c)
II Congestive cardiac failure, Age and Frailty

It is not uncommon for the very old to die after a long period of deterioration, without the benefit of recent full assessment. There may well have been no clear cause of either the bronchopneumonia or the congestive cardiac failure in this case.

Chronic obstructive lung disease, etc. (490 to 496)

Manual coding yielded 770 deaths, automatic coding 698 deaths, and 665 of these were to the same four-digit code. There were only a further eight instances of different codes within the range. The effect of the changes did not appear to be influenced by age at death (table 16).

Table 16. Frequency of allocation of chronic obstructive airways disease as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
1 to 74	345	315	307
75 and over	425	383	366
Total	770	698	673

The significant losses were to ischaemic heart disease (14), cardiac failure (12), other cardiac disorders (8), pneumonia (22), other respiratory disorders (17). Eight of the gains were from ischaemic heart disease.

Other respiratory disorders

Manual coding attributed 25 deaths to diffuse pulmonary fibrosis (Hamman-Rich syndrome, 516.3) and automatic coding 20, with 19 in common. Five of the losses with the automatic coding were to other diseases in the range 510–519. The two coding systems yielded very different numbers of deaths attributed to residual categories in the respiratory system, as shown in table 17. This is because the two systems handle “chronic airways disease” differently; it was coded manually to 519.8, automatically to 519.9.

Table 17. Frequency of allocation of residual categories of respiratory disease as underlying cause by method of coding

Coding method Category	Manual	Automatic	Both
519.8 Respiratory disease not elsewhere classified	40	28	20
519.9 Respiratory disease unspecified	6	22	3

The sample included 17 certificates of deaths at ages to 74 mentioning "chronic airways disease." These were all queried. In all of the 14 cases where there was a response, the death was reclassified as due to chronic obstructive airways disease. I therefore think we should either be querying the phrase "chronic airways disease" at all ages, or preferably regarding it as "chronic obstructive airways disease." The latter would require amendment of the ICD-10 index, and the Australian Center has included this in its 1997 recommendations to WHO and to other Collaborating Centers for Classification of Disease, as a recommended amendment to ICD-10.

Digestive system diseases

Among certificates coded by both methods, there were 460 deaths with manual coding, 461 with automatic coding, and 412 coded by both methods to Digestive system diseases. Major losses were to pneumonia (11) and to septicaemia (9), and gains were from neoplasms (12), but both gains and losses were widely distributed.

Peptic ulcer, 531 to 534, was the underlying cause of death in 72 certificates coded manually and in 63 coded automatically, and by both methods in 57 cases. Differences were more common at ages under 75 because of queries (table 18), but the only notable ones were two cases attributed manually to peptic ulcer but automatically to septicaemia. Mortality attributed to peptic ulcer will decrease with the change to automatic coding.

Table 18. Frequency of allocation of peptic ulcer as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
1 to 74	15	12	9
75 and over	57	51	48
Total	72	63	57

There were 39 deaths from intestinal ischaemia by manual coding and 45 by automatic, 38 of these being in common. Mortality attributed to this will increase with the change to automatic coding. Intestinal obstruction is also an important cause of death. It was selected as the underlying cause 52 times in manual coding, 46 times in automatic coding, and 37 times on both occasions.

There were 76 deaths from alcoholic liver disease, 571.0 to 571.3, with manual coding and 66 with automatic, 65 of the deaths being in common. Five of the losses were to other liver disease, but only two of these followed queries; it is not clear why the automatic system should not have allowed for the phrases "alcohol abuse" and "liver disease (alcoholic)." There will be a reduction in reported mortality from alcoholic liver disease with the move to automatic coding.

Renal disorders (580–589)

With manual coding there were 179 deaths due to Renal disorders, 580-589, and with automatic coding there were 192 deaths, 142 of these were allocated to this group with both coding methods. The losses were to hypertension (5), other heart diseases 420–429 (7), and pneumonia (13). The major gain was from ischaemic heart disease (17). The two coding systems have very different ways of coding chronic renal disease. Some examples are:

- Analgesic nephropathy still causes some deaths in Australia, although the number is decreasing. It is indexed to 583.8 but all four instances in the sample were coded manually to 582.8; with automatic coding three cases were coded to 583.8 and one to 582.8.

- Of 12 deaths attributed manually to unspecified chronic glomerulonephritis, 582.9, six were so attributed automatically and another four were coded to renal disorders.

- Of 132 codes allocated automatically to any form of renal failure (584–586) for deaths at age 75 years and over, 25 were allocated elsewhere on manual coding.

- Ten of 68 deaths manually attributed to chronic renal failure, 585.9, were allotted by the automatic system to acute renal failure, 584.9. These are cases of “acute on chronic renal failure” or of acute renal failure with chronic renal failure also mentioned. An example of the latter is:

  Female, 72

I(a)	Acute renal failure
I(b)	
I(c)	
II	Chronic renal failure, fractured neck of femur

- There is no provision for “acute on chronic renal failure” in ICD-10 and this is being pursued separately by the Australian Center. Clinical advice is that acute on chronic renal failure should be regarded as acute renal failure.

- There were 17 deaths manually coded to urinary tract infection without specification of site, 599.0, but 36 coded automatically, with only 12 of these being so coded by both methods. At ages under 75, where all certificates were queried, there were 7 deaths by manual coding, 7 on automatic coding, and 5 in common; both the losses on transfer to automatic coding were to septicaemia. At ages 75 and over no certificates were queried, there were 10 deaths from manual coding, 29 from automatic, and 7 in common; the gains with the adoption of automatic coding were from diverse sources including diabetes (2), dementia (2), Parkinson’s disease (2), heart failure (2), cerebrovascular disease (5), and renal failure (2).

The changes to renal disease mortality with the adoption of automatic coding will thus be complex.

The mortality attributed to renal failure is also greatly influenced by queries. Australian Bureau of Statistics procedures require that, at ages under 75, a certificate that gives any form of renal failure (584–586) as the underlying cause be queried. There were 35 certificates in the sample relating to deaths under age 75 and coded automatically to 584–586. Three of these were coded manually to ischaemic heart disease and hence not queried. Of the 32 queries, there were no replies to six, and one of these deaths was coded manually to respiratory disease. Of the 26 replies, seven resulted in confirmation of the underlying cause and 19 in changes, only four of these to a renal cause in the code range 580–583.

Diseases of skin, etc.

The numbers of deaths to diseases of the skin were small, but there were 30 manually allotted underlying causes, and 26 automatic, of which 20 were in common. The two losses to septicaemia deserve mention.

Musculoskeletal disorders

Mortality attributed to Musculoskeletal disorders will be greatly reduced by the change to automatic coding. With manual coding there were 73 deaths, with automatic coding there were 50 deaths, and 47 of these were in common. The main losses were to heart failure (3), pneumonia (10), and trauma (5). At ages up to 74 there were 28 deaths with manual coding, 19 with automatic, and 17 in common. At ages 75 and over there were respectively 45, 31, and 30 deaths. There was thus no apparent relationship with age and hence with queries of certifiers.

Malformations

Manual coding yielded 59 deaths to Malformations, automatic coding 67 deaths, and 53 of these were in common. Six of the automatic codes, five of them at ages 75 and over, were 742.5 for “myelodysplasia” or related terms; these result from a problem with the ICD-9 index already discussed under Neoplasms. The age distributions in table 19 exclude these deaths:

Table 19. Frequency of allocation of malformations as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
Under 1	37	36	34
1 to 74	21	23	18
75 and over	1	2	1
Total	59	61	53

In infants, the gains and losses were to and from perinatal disorders; at higher ages they were to and from many parts of the classification.

Perinatal disorders

There were 80 manually coded deaths to perinatal disorders, 66 automatic, and 62 coded by both methods to the range 760–779. However, only 18 deaths were coded to the same four-digit code by both methods. Losses to other parts of the classification included to communicable diseases (3), stroke (3) and malformations (4). Three of the four gains were from malformations. One particular coding problem has been with the term “Pulmonary hypoplasia” in premature infants. As used in Australia, this term refers to incomplete but otherwise normal development of the lungs, and we code it to 770.7. However the automatic system, following the ICD-9 index, regards it as a malformation and codes it to 748.5. The distinction is clarified in the index to ICD-10.

Symptoms, etc.

In the entire sample of 13,907 deaths, the causes of most of the deaths assigned to this category under either or both methods of coding were still under investigation at the time. For example, there were only 28 infant deaths as coded manually and 35 as coded automatically, 12 were cases of sudden infant death syndrome according to both coding methods. This syndrome was also the manual diagnosis in 15 of the remaining 23 automatic code assignments. If the comparison is confined to deaths where both codes were available, then there were 12 infant deaths, all attributed to sudden infant death syndrome by both methods. At ages 1 to 74 there were 5 deaths with manual coding, 8 with automatic, and 3 in common; most of these death certificates were queried with the certifying doctors. At ages 75 and over there were 12 deaths with manual coding, 9 with automatic, and 6 in common.

Trauma

Findings in respect of these deaths were also affected by the substantial proportion where information was still awaited at the time of automatic coding, and this may have biased the findings that follow. Where deaths were coded by both methods, there were 667 where the underlying cause was manually coded as external, 698 where it was coded automatically, and 627 where an external cause was coded by both methods. However, there were only 259 instances where the same four-digit code was selected as the underlying cause. The findings by age group are summarized in table 20.

Table 20. Frequency of allocation of trauma as underlying cause by age at death and by method of coding

Coding method Age	Manual	Automatic	Both
Under 1	1	1	1
1 to 74	505	521	197
75 and over	161	176	61
Total	667	698	259

Acknowledgments

The coding was supervised by Sue Walker (National Center for Classification in Health) and by Coleen Hill (ABS). Data were prepared by Peter Burke (ABS), Rod Hall (Australian Institute of Health and Welfare, AIHW) and Mike McGrath (AIHW).

Clarification of the Mortality Coding Instructions

Lars Age Johansson, Statistics Sweden, Nordic WHO Center for the Classification of Diseases

Background

Sweden will introduce the ICD-10 on 1 January 1997. At Statistics Sweden, we started our training sessions on mortality coding in May 1996. During the training, however, we experienced some difficulties with the instructions on underlying cause coding in Volume 2. There seem to be contradictions and other inconsistencies. In some places the coding instructions apparently retain earlier coding practice although the corresponding parts in Volume 1 have been changed, and in some important cases Volume 2 gives no guidelines at all.

At the recent meeting for the Heads of the WHO Collaborating Centers for the Classification of Diseases, the Nordic Center presented a paper on "Need for clarification of the ICD-10 mortality coding rules" (WHO/HST/ICD/C/96.18), based on the experiences from the Swedish training sessions. The result of the ensuing discussion was that the Nordic Center, together with the WHO Secretariat, was asked to coordinate an international Working Group on clarification of the mortality coding rules. The aim is to present suggestions for such clarifications at the 1997 Center Head meeting.

International comparability of mortality statistics requires uniform and consistent application of the ICD coding rules. A universally accepted standard for automated coding is certainly a way to achieve that, and the ICE meeting on mortality statistics thus seems to be an appropriate forum for a discussion on how to proceed with the clarification of the coding rules and guidelines. The intention behind this paper is to provide a background for that discussion.

Points for discussion

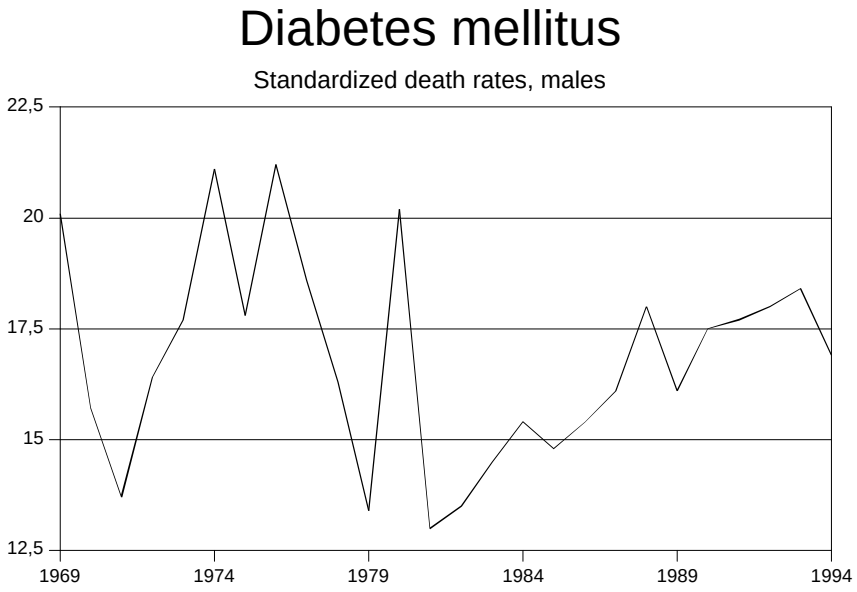
1. *Should coding instructions be complete?*

When mortality coding instructions are discussed, two rather different opinions can easily be distinguished: some people think that there should be explicit coding instructions for every conceivable case and that as little as possible should be left to the judgement of the individual coder, while others think that the coding instructions should only give general guidelines and that problematic cases should be solved on a common sense basis.

While computerized mortality coding obviously requires precise instructions on all cases that the computer software is supposed to handle, statistical reasons can be advanced for the "instructions for everything" approach as well: If coding is not performed according to an established standard, but depends on the judgement of the individual nosologist or medical advisor, then it is bound to be subjective and the influence of individual coders will show in the statistics. The resulting inconsistencies will make both international and national comparisons problematic.

This is amply illustrated by the Swedish mortality statistics for the 1970's and early 1980's. In the 1970's, a "common sense" approach to underlying cause coding was adopted by both medical advisors and coders. The ICD rules were seen as general but not binding recommendations. If a condition mentioned on the certificate "obviously" had caused the death, Statistics Sweden would "correct" the certificate and select that condition as the underlying cause, regardless of the ICD selection rules.

However, it became increasingly apparent that neither medical advisors nor coders were consistent in their selection of the underlying cause, and that the inconsistencies introduced by personal idiosyncrasies were serious enough to make statistical analysis of trends practically impossible. In 1981, it was decided to abandon the "common sense" coding method, a detailed manual of mortality coding was compiled and the coders were told to apply the instructions rigidly. The official Swedish statistics of deaths due to diabetes mellitus illustrate the effect of this change in coding practice.



While it is quite difficult to assess the trend in the 1970's, it becomes far easier after 1981, when rule-based coding had been introduced. Similar artefacts for the 1970's can, regrettably, be found for many other conditions.

Those who think that "common sense" should be used in applying the coding rules certainly have a point. Meticulous application of the coding rules will sometimes result in a fairly improbable underlying cause of death. However, if that is the price one has to pay to achieve consistent statistics, it does not seem too high. The alternative might be, as with the Swedish statistics on diabetes, that the statistics are useless altogether.

The Swedish experience—and it is presumably not unique to Sweden—suggests, then, that coding instructions must be followed slavishly, otherwise the statistics will not be consistent. This leads to a further conclusion: if the coding instructions are to ensure consistent and comparable statistics, then they must be as complete as possible. Gaps in the instructions will open the field for personal improvisations, and statistical comparability might be lost. It is important, therefore, to try to make the ICD rules and guidelines as complete and unambiguous as possible.

2. *Collect background information*

Presumably, difficulties with the interpretation of the ICD-10 coding rules and their associated commentaries and guidelines are experienced in more places than Sweden. A first task for the Working Group is, therefore, to collect questions and other observations on the coding instructions. The WHO Collaborating Centers are of crucial importance in collecting that information. The e-mail network for discussion of ICD-10 mortality coding could also be used for that purpose.

It should be stressed that the aim is to solve *practical* coding problems that might cause difficulties significant enough to impair the international statistical comparability. The discussions should, therefore, be based on death certificates actually encountered by the coders.

3. *Processing the information*

What is the appropriate level of ambition?

The "clarification" of the ICD instructions should not include attempts to change the existing rules and guidelines, unless it is necessary to resolve contradictions. Rather, the aim is to make the existing coding instructions easier to apply consistently, e.g.:

- Correct evident mistakes (e.g., the discrepancy between the cited code numbers and the explanatory text in the Volume 2 *Notes* on T36-T50, p. 61).

- Resolve inconsistencies (e.g., the *Note* on F01-F09, p. 52, says that these codes should not be used for underlying cause coding if the underlying physical condition is known, while the *Note* on I70.9 says that F01 and not arteriosclerosis should be coded as the underlying cause of death if arteriosclerosis has been reported as the cause of dementia).

- Clarify such instructions that could be interpreted in highly differing ways, e.g., the instruction on Rule 3 dealing with pneumonia and bronchopneumonia—quite important for the international comparability!

- Add instructions on important cases that are not included in Volume 2 (e.g., that suicide may not be due to any other condition).

- Provide more coding examples.

- Rewrite complicated coding instructions to make them easier to understand, e.g., as a decision-tree.

- Try to develop an exhaustive mortality coding manual.

It is obviously impossible to cover all these points in the time left until the 1997 Center Head (CH) meeting, and some order of priority must be established. The following order seems reasonable, and the Working Group will attempt to cover the first three points by the 1997 Center Head meeting.

1. Correct obvious mistakes and resolving inconsistencies
2. Clarify ambiguous instructions
3. Add essential instructions
4. Rewrite complicated instructions
5. Develop an exhaustive coding manual
6. Provide more coding examples

How to reach agreement on the clarifications?

According to the decision of the recent CH meeting, the Nordic WHO Center will function as coordinator of the undertaking. That would include:

- keeping a list of the members of the Working Group, and of the work the individual participants are prepared to undertake

- collecting information from ICD users on difficulties with the mortality coding rules and guidelines

- circulating that information to the WHO Secretariat and the Working Group

- working out suggestions for clarification based on the comments from the Secretariat and the Working Group

- reporting on the Working Group's activities at the Center Head meeting

Timetable

Corrections to Volume 2 will be considered an update of the ICD, and any suggestions for such updates must be submitted to the WHO secretariat by April 1997. Obviously, it will not be possible to produce a satisfactory text in such a short time.

However, the fact that the ICD-10 is now being implemented in many countries makes it necessary to cover at least the most crucial clarifications as soon as possible. The Nordic Center is compiling a document for the 1997 Center Head meeting with preliminary suggestions for clarifications, and is also drawing up a document of final suggestions for recommendations that the 1997 Center Head meeting may arrive at. A formal decision can be made at the 1998 Center Head meeting.

4. *Disseminating the result*

Once the proposed clarifications have been submitted to the Center Head meeting, and a decision has been reached, it is important that the result—the clarified rules and guidelines—is distributed to the WHO member states as soon as possible. For obvious reasons, a reprint of Volume 2 is out of the question, but other forms of publication could be used such as Internet, newsletter, or errata sheet.

MIKADO: A PC Software for Coding of Multiple Causes of Death

Lars Age Johansson, Statistics Sweden—Health and Social Welfare Statistics

Background

MIKADO (an acronym for "**M**ult**I**pel**K**odning **A**v **D**öds**O**rsaker"—Multiple Coding of Causes of Death) is a PC software program for automated coding of multiple causes of death, developed at Statistics Sweden, Stockholm.

Automated cause-of-death coding is, of course, afflicted with the same problems as manual coding. In general, it is time-consuming, expensive, and liable to systematic errors. Additionally, cause-of-death coding has its own specific problems. By international agreement, causes of death are coded according to the International Classification of Diseases (ICD), which is both extremely large and abounds in apparently arbitrary exceptions to its alleged general principles.

The coding of a certificate may be influenced by medical facts which are not explicitly stated, only implied. This makes the coding dependent on the medical knowledge of the coder. Since the coders' medical knowledge and familiarity with the ICD inevitably vary, considerable efforts are required to maintain acceptable coding stability. In our experience, it takes at least 2 years to train a new coder—if he/she has a basic knowledge of medical terminology and pathology.

On each death certificate, several conditions may be reported. When coding a certificate, the coder first assigns an ICD code to each one of the conditions reported (multiple cause coding), and then goes on to select a principle cause of death (named "the underlying cause of death" in ICD terminology) according to selection rules specified by the ICD. Most statistical tabulations and analyses are based on the underlying cause of death.

Since the late 1960's a software program has been available which, starting from multiple cause coded certificates, selects an underlying cause of death according to the ICD selection rules. This software, ACME, is developed and maintained by the U.S. National Center for Health Statistics and is the *de facto* international standard in its field.

ACME was introduced at Statistics Sweden in 1987. In 1989, we decided to try to automate the multiple cause coding as well, hoping that the coding would be faster, higher quality, and less dependent on the individuals who perform it.

Multiple cause-of-death coding—MIKADO

The aim of our project was to develop a module that translates the medical terms reported on the certificate into multiple cause codes. The following criteria were specified for the module:

- it must accept the language actually found on the certificates
- if several conditions are reported in the same field, the module must be able to code them separately
- it must allow supplementary and implicit information to influence the coding
- the computer-assisted (interactive) coding must be as similar to manual coding as possible
- the output must be in ACME compatible format

Of the coding systems available in 1991, none met all these criteria. By the end of 1991, we therefore decided to develop a multiple cause coding module of our own. A prototype, called AKK, was available in January 1993. After some modifications to AKK (including renaming it to AMK), we started a full-scale test in July 1993. A year later, the present version (named MIKADO) was introduced.

We decided to work according to the "prototyping" model, i.e., we did not start our project with an attempt to write a complete specification of the coding software. Instead, a primitive prototype was developed very early in the project and functions and refinements successively added to it. The main part of the work was done by two persons working part-time on the project (50 percent for the first 2 years, 25 percent for the last year of the project). One was an experienced database programmer, the other a senior coder with previous knowledge of software development. Once the full-scale test was mounted, all coders took part in the evaluation of the software.

Considerations on matching strategy

Our first plans were to use near-exact matching, which seemed to be the obvious way to avoid inconveniently large dictionaries. Our first results were disappointing. We tried first a "most discriminating compound" method and then a strategy based on a computed measure of similarity. However, both would yield a large number of theoretically possible, but unfortunately incorrect dictionary matches. To achieve a reliable match we would have to use a very high threshold value, and we soon realized that exact matching would give the same matches—and much faster.

The explanation of this result, which seems to be at odds with experiences from many other automated coding applications, lies probably in the structure of medical language. Medical terms are often compounds of a comparatively restricted set of basic elements. These elements denote, e.g., anatomical site or type of tissue (cervico-,

neuro-, myo-, cardio-), or the nature of a disease process (-itis, -oma, -osis, -pathy). Many elements are quite similar, especially in Swedish spelling, that tends to truncate suffixes ("myocardosis" will be "myokardos") and remove letters that are silent in Swedish pronunciation (e.g., "h" in "cirrhosis" or "p" in "symptom"). Sometimes a single letter makes the difference between two quite separate entities, e.g., "arter-" and "artr-" ("artery" and "joint," respectively), or the Swedish words "hjärt-" and "hjärn-" (heart, brain). Moreover, medical terms are often quite long ("kardioartierionefrocerebroskleros"), and essential information on the nature of the disease is often given by the very last syllable ("myocardit", "-it" denotes "inflammation"). This means that word truncation and weighting methods that give higher weight to the early parts of the word will return many incorrect matches.

Therefore, we decided to base the automatic coding proper (the part of the coding that will not be reviewed manually) on exact matching only. In the interactive coding, however, the coder has access to near-exact matching.

Of course, the performance of a system that uses exact matching only will be very dependent on the efficacy of the phrase standardization (parsing). Much effort has been spent on the MIKADO parsing procedures, which are described in Appendix 1.

Approximately 2 percent of the items to be coded are coached in "ordinary," nonmedical language. In such cases (mainly descriptions of accidents and violence) exact matching is quite clearly not suitable, and the rate of automatically coded responses is low. Our experiences suggest, then, that exact matching is preferrable when scientific terminology is concerned, since such terminology consists of a comparatively small number of basic elements and even small variations can be of crucial importance. Exact matching is not, however, appropriate for coding of responses in ordinary, nonscientific language.

System Overview

MIKADO runs on IBM-compatible PCs with a 486 processor or higher. It uses the Paradox Data Base Manager, version DOS 4.5, and has been developed in the Paradox Application Language (PAL). For the time being, it is designed as a stand-alone application, not for PC network.

About 100,000 cause-of-death certificates are sent to Statistics Sweden each year. The certificates are microfilmed and keyed to an ASCII file. A few standard abbreviations are used, otherwise all information on the certificates is entered manually exactly as it appears.

The ASCII files are converted to Paradox format, and then divided into work lots of about 450 records. The work lots are processed by MIKADO in a batch process, and problem terms or records will be flagged for manual review.

The work lots, including the coding suggested by MIKADO, are then examined interactively by a coder. Any editing is done using a "working copy" of the input text, while the original version of the text is stored separately. To facilitate the work of the coder, we have tried to make the interactive coding as similar to manual coding as possible. Thus, the screen layout imitates the certificate form, the coder always has access to the entire text of the certificate, and the necessary operations can be performed in any order the coder prefers. Before the coder is allowed to return a work lot, MIKADO checks (among other things) that all conditions entered by the certifier have been coded.

Typically, the reviewing may include operations such as correcting misspellings, and supplying codes for expressions not found in the dictionary. The coder can browse the dictionary in alphabetical or code order, and there are several search facilities available. Problematic records can also be referred to a senior coder.

Expressions not previously included in the dictionary will be copied to a provisional dictionary update file. The provisional dictionary update file will be reviewed by a senior coder and only then included in the dictionary. A "cloning" feature is available, by which it is possible to copy the codes and modification variables (see below) of an expression already included in the dictionary to a new expression. Each time the dictionary is updated, a check is run that ascertains that expressions with the same standardized text have been coded in the same way.

Text standardization and phrase separation

To keep the dictionary reasonably compact, and to increase the number of matches, the phrases are standardized prior to coding. The standardization procedure used by MIKADO includes steps such as removal of strings that do not influence the coding, replacement of some strings with synonyms, separation of phrases, alphabetical reorder of words in a phrase, etc.

A special feature of MIKADO is that some strings will be coded separately when they are removed, e.g., expressions indicating surgery or other forms of treatment, or the duration of a condition. These supplementary codes may be used later to modify the code of the medical condition itself.

For a more detailed description of the standardization procedure, see Appendix 1.

Dictionary of diagnostic expressions

There are two versions of MIKADO's dictionary of diagnostic expressions. One contains the expressions in their original, nonstandardized form, whereas in the other, the expressions have been standardized according to the specifications in the current standardization tables. Thus, an up-to-date version of the standardized dictionary can be prepared whenever the standardization specifications are changed.

Basic code and modified code

Code modification is a salient feature of ICD coding. This means that a medical term may have several different codes, depending on other information on the certificate. Even very common terms, like "heart attack" and "pneumonia," are subject to code modification. Therefore, an important part of MIKADO is the ability to handle such modifications automatically.

For every expression, the dictionary gives a *basic code*, that is, the ICD code to use if there is no other information on the certificate that modifies the coding. In many cases there is also a *modified code*, that is, the ICD code to use if there is indeed information present that influences the coding.

If an expression can have different ICD codes, the criteria for which code to use are specified by the *modification variables*. There are nine of these:

1. The duration of the condition
2. Conditions reported elsewhere on the certificate
3. Recent surgery
4. Complications to surgery
5. Recent injury
6. In cases of external violence, possible intent (e.g., suicide, homicide, accident)
7. The age of the deceased
8. The sex of the deceased
9. Specific expressions (text strings) used elsewhere on the certificate

For modifications depending on the basic codes of other reported conditions or on other specific expressions, MIKADO also recognizes eight different *relations*: the modifying condition/expression immediately precedes the expression to be coded, immediately follows it, immediately precedes or follows the expression to be coded and the entities are separated by a word that expresses a causal relationship, is reported on the same line, on a line above, on a line below, or anywhere on the certificate.

If an expression can have only one ICD code, there will also be only one record in the dictionary which contains only a basic code. If an expression can be coded in several ways, there will be one record in the dictionary for each cause to modify the coding. Each record will have both a basic code and a modified code, and a specification of under what circumstances the modified code is used rather than the basic one. For example, there is only one record in the dictionary for "alcohol-induced cirrhosis of liver," since no other information on the certificate can modify the coding of that expression. On the other hand, there are about 40 records for "cerebral hemorrhage", reflecting the possibility to code the hemorrhage as spontaneous, old, traumatic, congenital etc.

Code priorities

If an expression can be coded in different ways, and consequently there are several dictionary records containing that expression, MIKADO checks for each case whether the conditions specified by the modification variables are met by the circumstances in the present case. If more than one of the dictionary records meet the criteria, the records are ranked according to a set of priority rules.

If there is more than one dictionary record with the same rank, and the records give different modified codes, the coder has to determine interactively which dictionary record to use.

Results and experiences

We have now coded about 400,000 certificates using the AMK and the MIKADO. It has brought indisputable advantages: the coding is more accurate, much faster, and there is less need for continuous quality checks. In 1992, the coding error (underlying cause, most detailed level) was estimated at 7.2 percent. In 1993, after the introduction of AMK, the estimated error was 3.1 percent. Of these, about 0.7 percent were attributable to the automated coding proper, 1.5 percent to the interactive coding, and only 0.3 percent to keying mistakes.

The batch processing of a work lot (450 certificates) takes about 15 minutes. Before standardization of the phrases, a dictionary match is found for about 40 percent of the terms. After standardization, the success rate is now about 90 percent, compared with about 70 percent when AMK was first put into operation. For about 65 percent of the certificates, MIKADO codes every term on the certificate, and no manual review is necessary. With manual coding, it would take an experienced coder the best part of a day to code a work lot, with MIKADO, it takes less than half that time.

It is important to remember, however, that this does *not* mean that the coding is now 90 percent (or even 60 percent) cheaper than before. MIKADO takes care of the uncomplicated certificates and leaves the difficult ones to the coders, who sometimes get the impression that the coding is now slower and more difficult than before. The new technology has also generated several new tasks, such as running the batch jobs and reviewing dictionary updates. Most of this work is done by the coders themselves, and not by the computer staff.

The expenses of data entry are, of course, substantially higher, and to some extent use up what is gained at the coding stage. Full phrases are both longer and more difficult to type than digit codes, especially since the typists do not always understand the expressions they are copying. Besides, many a doctor's handwriting is quite as bad as is generally reputed. We have tried to scan the certificates, but the character recognition has not been very successful. Even though about 70 percent of the characters were correctly interpreted, it took more time to correct the remainder than to key the certificates from scratch.

The introduction of automated coding has made it possible to work off the backlog we have had since 1987, even though the coding staff has been reduced from eight coders to five. Due to the backlog, however, no financial savings has been made. Presumably, there will be no substantial savings until today's keying of the certificates at Statistics Sweden can be replaced by some form of electronic death certificate.

The acid test still remains, however. In 1997, Sweden will implement the Tenth Revision of the ICD. This is a major operation that requires, among other things, retraining of the coders and independent recoding of each certificate until acceptable uniformity of coding has been achieved. When the Ninth Revision was introduced in 1987, the cause-of-death statistics were almost 2 years delayed. It is our hope that MIKADO will make the transition to the Tenth Revision smoother. If so, that will fully justify the resources invested in it.

A great problem with sophisticated automated coding systems is that coding expertise is lost. Since the coders learn that MIKADO is usually right, they tend more and more to accept the coding suggested by the software, and gradually lose both their ability to code without computer assistance and to evaluate the performance of MIKADO. To counteract this, and to maintain the coding abilities that are needed to update the software, the coders are required to regularly code training sets of certificates manually.

Planned modifications to the multiple cause coding module

The present version of MIKADO was developed in Paradox for DOS, a software that is no longer supported by the manufacturer. Obviously, we will have to switch to another platform.

The Tenth Revision of the ICD will be introduced in Sweden on January 1, 1997. We plan to have an ICD-10 version of MIKADO available by May 1.

For validity checks, we still use a mainframe system developed in 1986. These checks will be transferred to MIKADO.

The ampersands, required in some circumstances by ACME to identify the starting point of a medical sequence, must be supplied manually in the present version of MIKADO. Depending on the requirements of the ICD-10 version of ACME, we are considering including a feature that will supply these ampersands automatically, or at least support the coders when assigning them.

For the ICD-10 version of MIKADO, a new method of handling dictionary synonyms will be used:

! For each ICD-10 category, the LEXBAS file will contain the full set of possible modifications for one diagnostic term only. This diagnostic term, with its set of modifications, is referred to as the "matrix" for expressions coded to that ICD category. For instance, while "pneumonia" (ICD-10 code J18.9) is coded in the same way as the Swedish expressions "pneumoni" and "lung

inflammation", only one of these (e.g., "pneumoni") will have records corresponding to the applicable coding modifications of "pneumonia", e.g., postprocedural pneumonia.

- The synonyms will contain a pointer to the matrix, in this case to "pneumoni". In this particular case, "pneumonia" and "lung inflammation" will have a pointer to "pneumoni".

- When a LEXIKON is produced, MIKADO will, for each expression pointing to a matrix, automatically generate the same set of modifications for the synonyms as for the matrix. So, if "lung inflammation" and "pneumonia" have a pointer to "pneumoni", these expressions will be coded, and if necessary modified, in the same way as "pneumoni".

Appendix 1.

(1) The dictionary is searched for the text string to be coded.

If the string is not found in the dictionary:

(2) Trim blanks—any blanks first and last in the string are deleted, double blanks in the phrases are replaced by single ones.

(3) Exceptions—flagging of strings to NOT standardize in the usual way. Using this feature, e.g., "left" and "right" can be retained in connection with heart failure, where it influences the coding, but deleted in other cases, where it does not.

(4) Hyphens are removed or replaced by other characters (see Appendix 2).

(5) Prefixes and suffixes are removed and replaced.

(6) Deletions—words and strings that do not affect the coding are removed, e.g., "the patient had...", "probable."

(7) Replacements—spellings and expressions are standardized.

(8) Periods—remaining periods are removed or replaced (see Appendix 2).

(9) Standardization of phrase separators—strings indicating the beginning or end of a diagnostic expression are replaced by one of three standard separators (";" for enumeration, "*>>*" for a "giving rise to"-type relationship, "*<<*" for a "caused by"-type relationship).

(10) Surgery—expressions indicating surgery or medical treatment are coded separately and then deleted.

(11) The exception sign "#" is removed.

(12) If an expression has been deleted in its entirety, it is replaced by a "not known" string.

(13) The dictionary is searched for the standardized string.

If still not found:

(14) Durations—expressions indicating the onset of or the duration of a condition are removed and, if possible, coded separately. If automated coding of the duration is not possible, the expression is marked for manual duration coding.

(15) The dictionary is again searched for the standardized string.

If still not found:

(16) All remaining blanks are removed from the standardized string, and a corresponding field in the dictionary (containing the standardized diagnostic expressions with all blanks removed) is searched for a match. If no match is found, the blanks are restored.

(17) The words of the phrase are sorted in alphabetical order, and the search is repeated, this time in an alpha-sorted field.

If still not found:

(18) Phrase separation—the string is searched for any standard separator (";", "*>>*", or "*<<*"). If a separator is found, each substring will be standardized as described above (1–17) and a dictionary search performed.

If still not found, or if no phrase separators are found, then mark the expression for interactive coding.

Appendix 2.

Replacement of hyphens

(In this and the following description, "B" stands for blank, "#" for digit, and "@" for letter.)

hyphens first and last in the string are deleted

#B-B#	>>	#-#
B-B	>>	B;B
@B-@	>>	@B;B@
B-##	>>	B##
@B-@	>>	@B@
@-B@	>>	@-B@
#-B@	>>	#B@
#-#	>>	#-#
I-I	>>	I-I
@-@	>>	@@
@-#	>>	@-#
#-@	>>	#-@
@---@	>>	@@

Replacement of periods

periods first or last in the string are deleted

B.B	>>	B;B
#.#	>>	#·#
B#.@	>>	B#B@
##.@	>>	##B;B@
@.#	>>	@B#
B###.	>>	B;B
.###@.	>>	B;B
B###@.	>>	B;B
B###B@.	>>	B;B
.@.@.	>>	B@B@B
.@.@@.	>>	B@B@@B
.@@.@.	>>	B@@B@B
.@.	>>	B@B
B@.	>>	B@B
.@@.	>>	B@@B
B@@.	>>	B@@B
.@@@.	>>	B@@@B
B@@@.	>>	B@@@B
.@@@@.	>>	B@@@@B
B@@@@.	>>	B@@@@B

any period still remaining is replaced by ";"

Automatic Codification of Underlying Causes of Death in Catalonia, Spain

Gloria Pérez, M.D., Ph.D.; Nuria Montellà, M.D.; Jaume Domènech, M.D., Ph.D., Information and Studies Service, Department of Health and Social Security, Generalitat de Catalunya, Barcelona, Spain

Introduction

The Autonomous Government of Catalonia signed a treaty with the Spanish National Institute of Statistics in 1983 that allowed us to code every death occurring in our territory and to analyze our mortality statistics. The underlying cause of death was selected manually using the International Classification of Diseases, ninth revision (ICD-9)(1). This coding system was applied to every death occurring in Catalonia by a team of four mortality coders.

In 1992 the main objective of the Catalonian Mortality Register was to use an automatic system for coding underlying cause of death and to incorporate the coding of multiple causes of death (2,3). Before deciding to develop a new system, we examined several coding packages and systems, including the MICAR, ACME, and TRANSAX systems from the U.S. National Center for Health Statistics (NCHS). At the end of 1993, a member of NCHS visited our center. We discussed the results of a preliminary study that compared manual and ACME automatic coding of underlying cause of death in a random sample of death certificates (DC).

The objective of this study was to establish the degree of agreement for underlying cause of death by manually and automatically coding every 1994 death occurring in Catalonia. Both underlying and multiple causes of death were coded. The second objective was to determine differences in underlying cause-of-death mortality rates between the manually and automatically coded records.

Methods

Currently, the Catalonian Mortality Register coding team selects the underlying cause of death on every DC. In order to manually code multiple causes of death, a nosologist and physician trained the coding team on the NCHS interpretation of the ICD-9 rules (4). When a high level of agreement was reached between the coders, the training was concluded.

Each month two sets of procedures were followed: 1) the classical procedures for obtaining official data on underlying cause of death, and 2) the new procedures, using ACME and TRANSAX, incorporated in a parallel system, that allowed us to read all codes on every DC after manual coding, automatically select the underlying cause of death (5) and then link the manually coded and automatically coded underlying causes of death. The ACME and TRANSAX modules were installed on a mainframe computer. For every death, we retained the manually and

automatically coded underlying cause of death codes, every code stated on the DC, and the multiple codes before and after TRANSAX.

Agreement between the manually and automatically coded underlying cause of death for three and four digits was calculated. The manually coded underlying cause of death was considered the standard to determine the sensitivity (S) and the value predictive of positive (VPP). The accuracy of each indicator was established using Percy's et al. criteria (6). Death rates were calculated for the manually and automatically coded groups of causes. The population used in the rate denominators was a projection derived from the 1991 census (7). Differences between rates were established with a confidence level of 5%.

Results

In 1994, 52,180 deaths occurred in Catalonia. The number of causes coded was 155,649, with a mean of 2.9 causes per DC. The agreement between the manual and automatic coding of underlying cause of death using ICD-9 three digit codes was 90.6%, and using four digit codes was 88.7%. The analysis of 17 groups of causes (table 1) showed differences in the codification of *signs and symptoms not elsewhere classified* group.

Table 1. Frequency of 17 Groups of Cause of Death. Catalonia, Spain, 1994

Cause of death	ICD-9 codes	Manual		Automatic	
		Number	Percent	Number	Percent
Infectious diseases	001-139	1,431	2.7	1,575	3.0
Neoplasms	140-239	14,335	27.5	14,201	27.2
Endocrine diseases	240-279	1,451	2.8	1,493	2.9
Blood Disorders	280-289	266	0.5	273	0.5
Mental Disorders	290-319	1,812	3.5	1,736	3.3
Nervous system diseases	320-389	1,151	2.2	1,195	2.3
Circulatory system diseases	393-459	19,582	37.5	20,211	38.7
Respiratory system diseases	460-519	4,368	8.4	4,424	8.5
Digestive system diseases	520-579	2,821	5.4	2,624	5.0
Genitourinary system diseases	580-629	975	1.9	921	1.8
Pregnancy and childbirth	630-676	1	0.0	1	0.0
Skin diseases	680-709	75	0.1	82	0.2
Musculoskeletal diseases	710-738	425	0.8	414	0.8
Congenital	740-759	166	0.3	161	0.3
Perinatal period	760-779	72	0.1	69	0.1
Not elsewhere classified	780-799	725	1.4	249	0.5
External causes	E800-E999	2,524	4.8	2,551	4.9
Total		52,180	100.0	52,180	100.0

The results show a high degree of accuracy (sensitivity above 80%) except for *endocrine diseases, blood disorders, genitourinary system diseases, and signs and symptoms not elsewhere classified*. The VPPs were below 80% for *endocrine diseases, blood disorders*, and *skin diseases*. The death rates were significantly different for *infectious diseases, digestive system diseases, skin diseases, and signs and symptoms not elsewhere classified* (table 2).

Table 2. Death Rates per 100,000 Population for Manual and Automatic Codification for Specific Groups of Cause of Death: Catalonia, Spain, 1994

Cause of death	ICD-9 codes	Manual	Automatic
Intestive infections	001-004,006,008,009	0.6	0.7
Respiratory tuberculosi	010-012	1.1	1.0
Remainder tuberculosi	013-018,137	0.7	0.6
Remainder bacterial infections	005,007,020-027,030-041, 080-083,087,090-098	2.9	3.9
Viral infections	045-057,060-066,070-079,138	0.9	1.8
Remainder infections	084-086,088,099,100-104, 110-118,120-136,139	0.4	0.8
Leukemia	204-208	6.9	6.6
Remainder endocrine diseases	240-246,251-279	4.1	4.4
Chronic rheumatic heart diseases	393-398	4.5	3.9
Hypertensive diseases	401-405	12.2	10.3
Other heart and lung diseases	390-392,415-417,420-428	75.9	96.9
Cerebrovascular diseases	430-438	101.0	96.1
Diseases of veins	441-448,451-459	11.5	10.7
Respiratory infections	460-466,487	1.9	1.7
Pneumonia	480-486	12.4	15.2
Chronic respiratory diseases	490-496	40.3	37.5
Gastritis	531-535	2.9	2.2
Cirrhosis and other liver diseases	520-530,536,537,540-543,550-553, 555-558,560-570,572-579	21.3	18.3
Urinary diseases	580-599	15.4	14.5
Female genital organs	610,611,614-629	0.1	0.2
Skin diseases	698,700-709	1.2	1.3
Not elsewhere classified	780-799	12.0	4.1
Industry accidents	E916-E921, E923-E927	0.5	0.5

In 88.7% of the cases, there was agreement between the manual and automatic codification of underlying cause of death. For the remainder, nearly half (4.6%) of the disagreement was due to differences in the use of ICD-9 rule one, and 2.5% was due to differences in the use of ICD-9 rule three (table 3). Spanish special instructions, which are specific interpretations for Spanish Mortality Registers accounted for 1.5% of the difference (table 3).

Table 3. Level of Agreement and Differences in in the ICD-9 Rules for Selection of Manual and Automatic Underlying Causes of Death: Catalonia, Spain, 1994

Level of agreement and differences	Number	Percent
Agreement	46,284	88.7
Rule 1	2,400	4.6
Rule 3	1,304	2.5
Rules 4 and 5	626	1.2
Spanish special instructions	783	1.5
Other queries	783	1.5
Total	52,180	100.0

Discussion

The present study shows a high level of agreement between the manual and automatic codification of underlying cause of death. This allows us to use the ACME and TRANSAX system for coding cause of death in Catalonia. The ACME system is more advantageous than the manual system for the maintenance of homogeneous criteria for codification of underlying cause of death over time and for distributing multiple causes of death.

The limitations of the automatic system are due to differences in the interpretation of ICD-9 rules for underlying cause-of-death codification between the ACME system and our coding team. This could modify trends in mortality rates of i*nfectious diseases, digestive system diseases*, and skin diseases. The group of *signs and symptoms not elsewhere classified* needs a special mention: when *cardiac arrest* and *signs and symptoms not elsewhere classified* were stated on the same DC, our team selected the second cause of death, but ACME selected *cardiac arrest*. Our physicians state cardiac arrest on a high number of Dcs.

Acknowledgment: The present work was funded by a grant 94/1195 of the Fondo de Investigación Sanitaria.

References

1. Organización Panamericana de la Salud. Manual de Clasificación Estadística Internacional de Enfermedades, Traumatismos y Causas de Defunción. Novena Revisión. Washington: OPS, 1975.

2. Israel R, Rosenberg H, Curtin L. Analytical potential for multiple cause-of-death data. Am J Epidemiol 124 (2): 161–79. 1986.

3. Israel R. Automation of mortality data coding and processing in the United States of America. Wld Health Statist Quart 43:259–62. 1990.

4. National Center for Health Statistics. Instructions for classifying multiple causes of death. NCHS instruction manual; part 2b. Hyattsville, Maryland: Public Health Service. 1984.

5. National Center for Health Statistics. Vital statistics, ICD-9 ACME decision tables for classifying underlying causes of death. NCHS instruction manual; part 2c. Hyattsville, Maryland: Public Health Service. 1984.

6. Percy, Stanek E, Glockler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am J Public Health 1981; 71: 242-50.

7. Institut d'Estadística de Catalunya. Cens de població 1991. Barcelona: Generalitat de Catalunya, 1992.

Appendix: Codification Problems

1. Problems in Medical Entity Codification

Some flags are detected when assigning an ERN to medical entities. We found two specific situations:

1.1. (a) Medical entities without a specific ERN code, but another code can be assigned in order to achieve the ICD-9 codification because a similar nosological entity exists, or

(b) a different medical entity should be selected because of disagreement in the ICD-9 code

Example 1.1

Medical Entity	*ERN file*
(a)- Carcinoma bronchopulmonary	Lung bronchial carcinoma
- Diselectrolytemia	Electrolytic desarangement
- Atrial rupture	Ruptured heart
(b)- Vascular dementia	Multiinfartional dementia

1.2. No ERN can be assigned because neither code exists for this medical entity in the ERN dictionary nor a similar medical entity can be found.

Example 1.2

- Perinatal encephalopathy
- Metrorrhagia
- Homicidal submersion
- Motorcycle rider injured in collision with other and unspecified vehicle (traffic)

2. MICAR process: problems and rejected cases

When running MICAR-200, we detect some specific problems and differences comparing our assignment code process:

2.1. In the case of an accident with an unknown ERN (999999), MICAR output rejects, and tries to assign a code of an ill-defined entity (ICD-9, 799.7). Why is the manner of death (accident) not considered in order to assign a code of an accident not elsewhere classified?

2.2. Some cases are rejected although the coding is correct.

Example 2.2

Part I cardiorespiratory arrest/
complete auriculoventricular block/
digital intoxication

Output: 4275/4260/721 8583

2.3. When surgery is not indicated in a pacemaker patient due to a cardiopathy, MICAR considers the ERN cardiopathy is incorrectly positioned and erroneous because surgery is flagged and not found.

Example 2.3

Part I septic shock; cardiac arrest/
base pneumonia

Part II pacemaker patient; cardiopathy

Output: 7855 4275/481*8781 4299

Automatic Coding of Causes of Death by Means of Neural Networks

Xavier Roselló, M.D., Ph.D.[A]; Nuria Montellà, M.D.[B]; Josep M. Balaguer, M.S.[A]; Gloria Pérez, M.D., Ph.D.[B]; and Jaume Domènech, M.D., Ph.D.[B]

[A] Politechnical University of Catalonia. Barcelona, Spain.
[B] Information and Studies Service. Department of Health and Social Security. Generalitat de Catalunya. Barcelona, Spain.

Introduction

Mortality statistics are based on the causes of death reported on the Death Certificate (DC). The underlying cause of death[1] is selected from all medical entities listed on the DC. The use of all causes (multiple causes) is also recommended by the World Health Organization (WHO) in order to avoid the loss of additional information. Multiple cause management requires specific computer tools because of its magnitude and complexity. For processing cause of death, the U.S. National Center for Health Statistics (NCHS) developed automated computer systems to produce mortality statistics. One of these applications is the MICAR system (Mortality Medical Indexing, Classification, and Retrieval System), which assigns a specific code (the Entity Reference Number, ERN) to each reported condition on the DC. The MICAR system is based on a dictionary file and only accepts, to date, causes in English. The exclusive use of the English language limits its usefulness somewhat.

The Mortality Register and the Politechnical University of Catalonia have developed software that automatically performs the ERN coding process in order to establish the automatic codification and management of multiple causes. The methodology selected is that of Neural Networks because of its ability to recognize and automatically code bilingual patterns. This method operates on the basis of morphological and structural similarity.

The Neural Network Operation

Description of medical entities is constructed with word combinations. Thus, identification of words included in the medical entities may be the starting point for recognizing causes of death reported on the DC. Medical entities reported on the Catalonian DC may be written in either Catalan or Spanish—both languages are similar and share many of the same roots. Because there are two languages,

[1] The underlying cause of death is defined by WHO as "(a) the diseases or injury which initiated the train of morbid events leading directly to death or (b) the circumstances of the accident or violence which produced the fatal injury" (1).

the number of words must be smaller than the number of possible medical entities. In addition, words are grouped into families according to their roots.

The preceding ideas allow for the development of a method that recognizes similar words. The methodology is based on word proximity, defined as the rate between corresponding letters and the total number of letters of the longest word, or similarity, which is determined by comparing the first vowel and the sequence of the consonants. With word identification, medical entities are recognized and assigned a specific ERN.

Neural Network (NN) can be used as a tool for recognizing patterns. The name comes from its similarity to the operation of neural networks in the brain. An artificial NN consists of a set of processing elements or neurons arranged in layers and connected from layer to layer. The number of layers and the number of neurons in each layer depends on the data entry information and the expected output. An artificial neuron carries out three operations:

1. *Input combination*—an input comes from outside the network or from a previous layer through a connection. Each connection has a weight that represents the strength of the link between the two neurons. The input combination used is a linear combination of inputs where the coefficients are weighted.

2. *Transfer function*—transforms recombined input into the output. It is a monotonically increasing function as the sigmoidal or hyperbolic tangent.

3. Neuron output value—the output value from the transfer function.

A NN, "DECES" has been designed so that the input layer neurons are associated with words that have similar semantic content called *terms* (articles and prepositions are dropped). The output layer neurons are associated with medical entities that are correctly written in Catalan, called *descriptors*. These output layer neurons use a sigmoidal transfer function. The cases where abbreviations have been used or where Catalan and Spanish words are very different is handled by replacing the original word with a synonymous Catalonian word that corresponds to the previous input layer.

A NN will only work if its weights have suitable values. It uses a learning process to modify weights in such a way that for every input presented to it, the NN yields the desired output. It is implemented by an iterative algorithm whose objective function is the minimization of quadratic error between observed and desired outputs for every input used. During the learning process, the knowledge is stored in the weights of the NN connections.

When a medical entity reported on a DC is considered, synonymous substitutions are made. The medical entity is then split into components that correspond to the consonant sequence. Each activated consonantic sequence in turn activates descriptors containing a word with the same consonantic sequence. The neurons activated in the descriptor layer yield an output value between 0 and 1; and the ERN

associated with the highest level of activation is selected as the code for the medical entity problem. The NN is not able to assign an ERN if:

1. One or more words are unknown and are not similar to any of the input layer terms. If this is the case, then it is necessary to add a new descriptor and corresponding ERN to the NN to be able to code the medical entity.
2. Multiple descriptors, with different ERNs, are activated at similar levels so that the system has difficulty assigning the correct code. In this case, the NN must go through the learning process.

In either case, the result is a descriptor without an ERN or with an incorrect ERN assignment.

Neural Network Accuracy

The accuracy of the DECES system in assigning the correct ERN was evaluated before using it in the system. We compared the global agreement and the sensitivity, specificity, and the positive and negative predictive values to manual codification.

First, we included the 2,000 most frequently used medical-entity ERNs on the NN and began the learning process. Then, when new cases of medical entities without ERNs were detected in the DECES application these were also added to the NN. At this time, the accuracy study includes 3,612 ERNs. From 29,946 DCs, a random sample of 1,067 was drawn, which included 3,471 causes (3.25 causes per DC). The number of medical entities coded by DECES was 3,141 (90.49%) while 3,174 medical entities were manually coded(91.45%). Codes were correctly assigned in 86.60% of the cases and the global agreement between those that were correctly coded was 93.17%. The accuracy indicators are as follows:

	Value	CI 95%
Sensitivity	94.71	93.9–95.5
Predictive positive value	97.75	97.2–98.3
Specificity	76.76	71.9–81.5
Predictive negative value	57.57	52.7–62.4

Reference

1. World Health Organization. Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, based on the recommendations of the Ninth Revision Conference, 1975. Geneva: World Health Organization. 1977.

Test of MICAR-TRANSAX-ACME System: Coding of Israeli Death Notifications

Pnina Zadka, M.Sc., M.P.H., State of Israel, Prime Minister's Office, Central Bureau of Statistics-Health Division

Background

Israel had planned to adopt the U.S. NCHS software for automated coding of causes of death for over a decade. The initiative to change into a fully automated coding system evolved from three main objectives: to study multiple causes of death, improve uniformity in coding, and possibly to decentralize the coding of causes of death.

Multiple Causes of Death

By coding all the listed diagnoses on the death notification, the knowledge on morbidity-related mortality will expand beyond the currently available underlying cause of death, to a more comprehensive system that will include all listed diagnoses on the death notification. The underlying cause of death has been used in the national cause-of-death statistics since the establishment of the state. The need to analyze co-morbidity has become essential as almost half of the deaths are due to heart disease-related causes.

Uniformity

Currently the coding of causes of death in Israel is carried out only within the Central Bureau of Statistics (CBS), with manual selection of the underlying cause of death according to the WHO International Classification of Disease rules. These rules were only slightly modified with each new classification. By using an automated selection system, the inter-coder selection bias will be reduced and the training period for new coders will be shorter.

Decentralization

Health reforms and the reorganization of the Ministry of Health have raised the need for the Regional Health Offices to provide updated information systems. Plans are being made to code causes of death in the Regional Health Offices. Decentralization of the coding of causes will make quality control of the registration and coding of causes of death very difficult. Implementation of an automated coding system will enable better quality control by the supervising authority, enhance uniformity in reporting to the central office, and avoid the need to duplicate the causes of death coding in the CBS.

The need for a system that will answer these needs led to a project that was carried out in the United States—testing the MICAR-TRANSAX-ACME system on original death notifications from Israel by an experienced coder from CBS, Israel.

Findings

Eighty-one death notifications were entered into SUPERMICAR, and the results are summarized in the following tables.

Table 1: Experimental run on SUPERMICAR

Number of records	30
Number of terms (codes)	184
Number of rejected records	14

In the experimental run, which included only 30 records that were processed by SUPERMICAR (the program that assigns numerical codes to the textual terms), the program terminated without processing almost half of the cases. These cases had at least one term that was not recognized by the software. The cases that were rejected by the software in the experimental run were mainly due to:

1. Spelling mistakes made by the person who filled out the death notification.

2. Unrecognized abbreviations.

3. Data entry mistakes, probably caused by lack of experience with the software.

The first run did not include the ACME part (the program that selects the underlying cause of death).

Table 2: First run SUPERMICAR

	Records	Terms
Processed	81	427
Completed	49	379
Terminated	32 (39%)	48 (11%)

Table 3: SUPERMICAR after processing of rejected records

	Records	Terms
Processed	81	427
Completed	58	396
Terminated	23 (28%)	35 (8%)

Table 4: Results from the ACME run

Ill-defined terms deleted from the ACME input	27
Codes changed prior to the ACME input	3
Codes changed during ACME input	24
External causes requiring reformation	17

Problems detected by comparing with manual processing:

1. A feminine cause assigned to male. For example, cervical stenosis (6,224) should have been spinal cervical stenosis.

2. Some cases had a wrong selection of the underlying cause of death due to overflow of the written text to Part II of the causes of death notification.

3. Small cell CA of pelvis (1,953) was omitted.

4. Entro-vesical (5,962) was miscoded as intestino vesical (5,961).

5. Ulcer in sacrum was omitted as there was no ";" preceding it.

6. Eisenmenger's syndrome was not selected as the underlying cause for pulmonary hypertension in an 8-year-old boy.

7. Ischemia inf. lat. wal. was unrecognized and deleted from the process and s/p CVA was selected.

8. Metastases general was unrecognized.

9. Cr. intoxication was unrecognized.

10. Cr. uteri was unrecognized.

11. Diabetic mellitus was unrecognized.

Discussion and conclusions

The software is useful in the expansion of multiple causes of death coding. It is now being modified for personal computers. Modifications are being made continuously by adding unrecognized terms to the dictionary. There will be a new version suited for the Tenth Revision of the ICD. The software seems an easy step in the transformation to multiple cause-of-death coding and analysis.

Nevertheless, despite the advantages of the automated cause-of-death coding, using the software developed by the United States, there are some problems with its use and implementation. These include problems that are related to misspelling of words, spelling of words in a different way (English or Latin), misorganization of terms or words, and the deletion of unrecognized terms or words. The software makes deletions without any on-line warning and without notifying which terms are omitted or deleted. Some of the selections of the underlying cause of death are not acceptable. Data entry persons are required to be able to reorganize the order of the listed terms and to make decisions in cases where the registration overflows to Part II. The current volume of rejections of 30%-40% is an obstacle in a decentralized coding system. The training and reorientation of coders with the software requires better instruction manuals that can be used without intensive individual instruction periods. In order to implement the software there is a need to adjust the dictionary for use with European terms and spelling, as well as to enable reorganization and modification of deleted terms.

Factographic Automated Information Reference System (FAIRS-"Potential")

Sergei P. Ermakov, D.Sc.; Vladimir V. Antonyuk, Ph.D.; Natalia S. Gavrilova, Ph.D.; and Galina N. Evdokushkina, Russian Institute of Public Health

The family of IBM PC-oriented information systems was developed for medico-demographic comparative analyses, for developing trends, and for the quantitative estimation of health priorities. This system is based on our methodological approach for constructing new demographic indices—the life potential and the working potential that link together some ideas from potential demography and a cohort consideration of the population (1).

The factographic automated information reference system (FAIRS-"Potential") can provide user-friendly population and mortality data input in compliance with Russian state statistic forms. It also calculates and compares many medico-demographic indices such as standard mortality ratios, age-adjusted mortality rates, the average age of death, and many other indices calculated for classes of diseases, and for individual categories of diseases, used in international statistical practice. This system also calculates integral complex indices of potential person—years lost due to premature mortality. These indices may be used for establishing priorities in regional health care planning processes and as specialized measures for estimating public health service development projects.

The system FAIRS2-"Potential" provides for: user-friendly input of primary death certificates in compliance with the Russian state registered system and user-friendly input of disability primary certificates. It also provides analyses of mortality data by classes of causes, individual causes within a class, by region and place of death, and accounting for sex, age, education, ethnicity, and social status, etc. Analyses of disability data by disability groups, disability causes, and regions are available as well.

Fairs-"Rayon" makes use of a new and original method of decomposing age-specific mortality patterns into endogenous and exogenous components and a specially-developed interaction algorithm using a "bootstrap - procedure" of assessing confidence intervals of obtained components by a simulation method. The system allows for a quantitative assessment of all external factors affecting mortality without measuring these factors by population size. This feature of the system creates a new opportunity for ecological monitoring.

System and decomposition methods allow the calculation of all the mortality indices mentioned above (see the first system, FAIRS-"Potential") and provide the opportunity for calculating the health potential for small areas.

This system was successfully used for preparing many articles and maps in the *Environmental and Health Atlas of Russia* (2).

The data presented below have already been calculated in the Department of Medical Demography of NPO "MEDSOCECONINFORM" and are ready for further distribution and dissemination.

Data Created by FAIRS-"Potential"

Description of Data—Age-specific mortality rates, 11 mortality indices, 8 indices of absolute losses of health potential.

Years—1980, 1981, 1989–96.

Regions—173 regions including 16 republics and the former USSR (1989-90 only), 72 regions for Russia (1980, 1981, 1989-96), 26 for Ukraine (1989-90 only), 18 for Kazakhstan (1989-91 only), 12 for Uzbekistan, 7 for Belorussia (1989-91 only), 5 for both Tadzhikistan (1989-91 only), and Turkmenistan (1989-90 only), 5 for Georgia (1989-90 only), and 4 for Kirgizstan (1989-91 only).

Causes of Death—195 individual causes of death (coded by ICD-9) and 18 classes of diseases.

Types of Settlement—Urban, rural, and both.

Data created by FAIRS2-"Potential" and FAIRS-"Rayon"
(Owned by the Sverdlovsk region medical information computer center.)

Description of Data—Death certificates, age-specific mortality rates, 11 mortality indices, and 4 indices of absolute losses of health and working potential.

Years—1990–95.

Regions—53 small areas in the Sverdlovsk region.

Causes of Death—195 individual causes of death (ICD-9) and 18 classes of diseases.

References

1. Ermakov, SP, Kiselev, AA. 1992. "Economical Aspects of Health," *World Health Statistical Quarterly*, vol. 45, no. 1, pp. 50-60.

2. *Environmental and Health Atlas of Russia*. 1995. Ed. by M. Feshbach Moscow, "PAIMS" Publisher House.