

A.H. Burton, J. A. Dean, & A. G. Dean

# Software for data management and analysis in epidemiology

The collection of data is a lengthy and time-consuming process, but is not an end in itself. For the data to be useful, it is necessary for an appropriate analysis to be made and its results applied. This article describes a computer program for the analysis of epidemiological studies.

The results of the *World Health Forum* readership survey reported in this issue (pages 65-74) were processed on a microcomputer using the *Epi Info* software package (1). *Epi Info* is a family of general-purpose microcomputer programs designed to facilitate the management and analysis of epidemiological data. Versions 1 to 3 of *Epi Info* were produced by the Epidemiology Program Office, Centers for Disease Control (CDC), Atlanta, Georgia, USA. In 1988 the Surveillance, Forecasting and Impact Assessment Unit of the Global Programme on AIDS (GPA) of the World Health Organization modified *Epi Info* and distributed Version 4 for the management of data concerning AIDS and HIV infection. Version 5 is a joint CDC-WHO/GPA

product and is available from both the CDC and WHO.

Since 1987 more than 4000 copies of *Epi Info* have been distributed in more than 34 countries. It is used in routine communicable disease surveillance systems, for investigation of disease outbreaks, data management for HIV/AIDS surveillance, and nutritional surveys. Several universities in the USA and Europe use *Epi Info* in courses on biostatistics and epidemiology.

*Epi Info* contains modules for defining data entry screens; setting data entry validation checks; performing data entry and update; and producing summary analyses. It also includes utilities for importing from and exporting to other common data formats such as Lotus 1-2-3, dBase III, SPSS, and SAS. The box overleaf explains these functions in more detail.

---

Mr Burton, a Systems Analyst, was formerly with the Centers for Disease Control, Atlanta, GA, USA, and is now with the Surveillance, Forecasting and Impact Assessment unit of the Global Programme on AIDS, World Health Organization, 1211 Geneva 27, Switzerland. Mr J. A. Dean is in the Department of Computer Science at the University of Minnesota, Minneapolis, Minnesota, USA, and Mr A. G. Dean is with the Epidemiology Program Office at the Centers for Disease Control, Atlanta, Georgia, USA.

Version 5 includes missing value indicators, fixed decimal fields, European data fields, mathematical functions, linear regression, direct analysis of dBase III files, graphics in ANALYSIS, a report formatter, pop-up

### ***Epi Info* modules**

- EPI           – the main menu providing access to the other modules.
  
- EPED          – a text processor used to define data entry screens, *Epi Info* programs, *Epi Info* formats, and to create simple documents and memoranda.
  
- ENTER       – the data entry module. New data may be entered, previously entered data may be retrieved by browsing, by specifying the record number, or by indicating the contents of the record. Records may be updated and/or flagged for deletion.
  
- ANALYSIS   – reads both *Epi Info* and dBase III files. It allows the user to produce line lists, frequency distributions, cross-tabulations, odds ratios, relative risk, and Mantel-Haenszel summary statistics for stratified tables. It performs analysis of variance, multiple linear regression, and matched case-control analysis. A REPORT command allows formatting the output of tables and frequencies for standard reports. Multiple files may be related on a common key. ANALYSIS commands may be entered in the interactive prompt mode or command files may be created for batch execution.
  
- CHECK       – allows the user to define data entry validation and control checks. Ranges, legal codes, skip patterns, and repeating fields may be easily specified.
  
- STATCALC   – a calculator program used to compute standard epidemiological statistics on  $2 \times 2$  and  $2 \times 2 \times n$  tables. Odds ratios, relative risk, chi squares, Mantel-Haenszel summary statistics, and exact confidence intervals are calculated from tables entered at the keyboard. STATCALC also computes sample sizes and performs tests for trends.
  
- CONVERT     – exports *Epi Info* files to 12 other data file formats. In addition to fixed length and delimited ASCII files, it also converts *Epi Info* files to dBase III, Lotus 1-2-3, SAS, and SPSS PC formats.
  
- IMPORT      – converts fixed length and comma delimited ASCII files, dBase III and Lotus 1-2-3 worksheets to *Epi Info* files.
  
- MERGE       – combines *Epi Info* files. Files with a common structure may be appended to each other. Files with common keys may be merged or data from one file may be used to update a master file.

windows for data entry codes, relational linkage of several files, and an improved user interface for ANALYSIS.

Other features new to Version 5 include an improved editor with programming features for prompted text creation, mathematical operations and comparisons among multiple fields during data entry, indexed files, searching on multiple fields in ENTER, and relational data entry. Important additions for advanced system developers are the ability to include custom-written data entry and analysis routines.

Version 5 requires an IBM-compatible microcomputer running DOS 2.0 or later, 512K (512 000) bytes of random access memory (RAM), and at least two floppy disk drives. A graphics adapter is necessary to produce graphs. For more effective use, 640K of RAM, a hard disk, a colour monitor and a printer are recommended. While the complete *Epi Info* system can be loaded in 1.5 megabytes of disk storage, the largest program will fit on one 360K diskette.

Data files may contain up to 4 000 000 records of up to 300 variables. A questionnaire may consist of up to 500 lines or approximately 20 screens.

Version 5 is available in English, and a French version is planned for release later this year. Copies of *Epi Info* may be obtained free of charge by writing to the following address. Please specify if 5 ¼ inch (13.5 cm) or 3 ½ inch (9 cm) diskettes are required.

Surveillance, Forecasting  
and Impact Assessment Unit  
Global Programme on AIDS  
World Health Organization, 1211 Geneva 27,  
Switzerland.



#### Reference

1. **Dean, A. G., Dean, J. A., Burton, A. H., & Dicker, R. C.** *Epi Info Version 5: a word processing, database, and statistics program for epidemiology on microcomputers.* Atlanta, GA, USA, Centers for Disease Control, 1990.