

数据录入及数据管理的免费专业软件
世界卫生组织及从事数据管理的专业公司推荐使用

EpiData 3.02

SHUJU GUANLI RUANJIAN

数据管理软件

实用教程

SHIYONG JIAOCHENG

主编 许军



军事医学科学出版社

EpiData3.02 数据管理软件 实用教程

主 编 许 军

军事医学科学出版社
· 北 京 ·

内 容 提 要

EpiData3.02 是一个专门用于数据管理的软件,充分吸收了 Epi Info 5.0、Epi Info 6.0、Epi Info 2000、EpiData2.0 及 EpiData3.0 等系列数据管理软件的精华。本书以新药临床试验数据管理为实例,介绍了 EpiData3.02 的运行环境和软件特点、调查表文件的建立、数据的录入与核对、数据的逻辑审查、数据的报告、数据管理程序的编写、系统参数的设置等。本书注重实用,通过一些示范性图片以及对某些操作的个人理解,有助于读者快速掌握 EpiData3.02 使用技巧。本书特别适合从事数据录入、数据核对、数据库的管理和数据报告等工作的科研人员学习 EpiData3.02 使用。本书还可作为大学生、研究生的教材,亦可以作为相关领域培训和继续教育用书。

图书在版编目(CIP)数据

EpiData3.02 数据管理软件实用教程/许军主编.

—北京:军事医学科学出版社,2006

ISBN 7-80121-776-4

I. E… II. 许… III. 数据管理—软件, EpiData3.02—教材

IV. C819

中国版本图书馆 CIP 数据核字(2006)第 005011 号

出 版:军事医学科学出版社

地 址:北京市海淀区太平路 27 号

邮 编:100850

联系电话:发行部:(010)63801284
63800294

编辑部:(010)66884418;66884402 转 6210,6213,6216

传 真:(010)63801284

网 址:<http://www.mmsp.cn>

印 装:京南印装厂

发 行:新华书店

开 本:880mm×1230mm 1/32

印 张:8

字 数:197 千字

版 次:2006 年 8 月第 1 版

印 次:2006 年 8 月第 1 次

定 价:36.00 元

本社图书凡有缺、损、倒、脱页者,本社发行部负责调换

前 言

EpiData 是由丹麦欧登塞(Odense, Denmark)的一个非盈利组织开发的。程序设计者为 Jens M. Lauritsen, Michael Bruus 和 Mark Myatt。EpiData 系列软件是在 Epi Info 6.0 软件基础上开发的,与 Epi Info 2000 软件相比,更加简单,有汉化功能,适用于各类科研人员管理实验数据。EpiData 软件虽不具备数据的分析功能,但有较强的数据转换功能,由该软件建立并存放的数据文件可以转化为比较常用的 dbf、Excel、SPSS、SAS 以及纯文本文件等多种数据库格式文件类型。该软件简单易学,方便实用,是一个免费使用的软件工具。目前该数据管理软件已有多种语言版本,如丹麦语、挪威语、荷兰语、意大利语、中文、法语、西班牙语、俄语、斯洛文尼亚语、塞尔维亚语、波兰语、葡萄牙语、阿拉伯语、德语、罗马尼亚语、英语等。

EpiData 软件可以在网上随时更新升级(<http://www.epidata.dk/>),目前已发展到 EpiData 3.02 版本。随着 EpiData 软件的不断升级,EpiData 在保留以前版本功能的基础上,不断充实新的内容,使该软件管理数据的功能不断得到加强和改进。本书注重介绍 EpiData3.02 软件管理数据的各项功能。作者凭借多年从事新药临床试验数据管理的实战经验,详细介绍了 EpiData 软件使用技巧和注意事项,通俗易懂,方便自学。

本书共有 10 章。第 1 章介绍了 EpiData3.02 软件的运行环境、特点及主要功能模块,第 2 章是关于操作 EpiData3.02 软件的菜单功能,第 3 章介绍了应用 EpiData3.02 的有关功能模块建立调查表文件,第 4 章介绍如何运用 EpiData3.02 建立及修改数据库,第 5 章是如何应用 EpiData3.02 建立数据核查程序及相关文

件,第6章介绍在 EpiData3.02 环境下如何进行数据库的数据录入,第7章是关于如何应用 EpiData3.02 软件对已建立的数据库进行管理和维护,第8章介绍了运用 EpiData3.02 的相关功能模块进行数据库的输入、输出及数据库文件的备份,第9章是一个具体的实例,第10章是关于 EpiData3.02 的系统设置及其他有关高级功能介绍。

本书适合多层次多专业的人士阅读,如统计学、数学、经济、金融及医学等研究领域的科研人员;可作为大学生、研究生的教材,亦可以作为有关部门进行培训和继续教育用书。

感谢暨南大学统计系王斌会教授,南方医科大学南方医院杨云滨教授对本书的出版给予的技术指导和关心!

本书中如有不尽人意的地方,诚恳地希望得到大家的谅解并不吝赐教,以便今后再版时加以改进。

许 军

2006年6月于广州南方医院

E-mail: drugstat@163.com

目 录

第 1 章 软件概述	(1)
1.1 EpiData 的历史	(2)
1.2 EpiData 与 Epi Info 的兼容性	(2)
1.2.1 数据文件中的区别	(3)
1.2.2 EpiData 的 CHECK 语言与 Epi Info 的区别	(4)
1.3 软件界面	(5)
1.3.1 主程序界面介绍	(5)
1.3.2 菜单栏界面	(6)
1.3.3 工具条和按钮条	(7)
1.3.4 状态栏和页标	(7)
1.3.5 输入程序界面介绍	(7)
1.4 相关文件类型	(9)
1.5 一个简单的例子	(10)
第 2 章 菜单功能	(14)
2.1 文件菜单	(14)
2.2 编辑菜单	(18)
2.3 数据文件菜单	(26)
2.4 数据检查	(29)
2.4.1 选择输入字段并为它添加有效性规则的方法	(30)
2.4.2 检查功能窗口	(31)
2.5 数据输入/输出	(33)
2.6 有关文件	(51)
2.7 工具	(60)

2.8 窗口设置菜单和帮助菜单	(69)
第3章 建立调查表文件	(71)
3.1 编辑器(Editor)	(72)
3.1.1 自动缩进(Auto Indent)	(72)
3.1.2 对齐变量(Align Fields)	(73)
3.1.3 变量类型选择列表(Field Pick List)	(77)
3.1.4 编码器(Code Writer)	(85)
3.1.5 制表符@ (Tabulator Code @)	(86)
3.1.6 预览数据录入表格(Preview Data Form)	(89)
3.2 定义变量名(Field Names)	(90)
3.2.1 将第一个单词作为变量名(First Word in Question is Field Name)	(91)
3.2.2 自动定义变量名(Automatic Field Names)	(95)
3.2.3 变量标签(Variable Labels)	(98)
第4章 数据库的建立和修改	(99)
4.1 创建数据库(Create Data File)	(99)
4.2 修改数据库(Revise Data File)	(101)
4.3 重新定义变量名(Rename Fields)	(103)
第5章 建立核查文件	(105)
5.1 添加/修改核查命令(Add/Revise Checks)	(106)
5.1.1 基本操作	(106)
5.1.2 基本 CHECK 命令的设置	(109)
5.1.3 编辑当前变量的所有 CHECK 语句	(114)
5.2 CHECK 的文件结构	(116)
5.3 CHECK 命令列表	(117)
5.3.1 AFTER ENTRY	(117)
5.3.2 AFTER FILE	(117)
5.3.3 AFTER RECORD	(118)
5.3.4 AUTOJUMP	(119)

5.3.5	AUTOSAVE	(119)
5.3.6	AUTOSEARCH	(120)
5.3.7	BACKUP	(122)
5.3.8	BEEP	(122)
5.3.9	BEFORE ENTRY	(123)
5.3.10	BEFORE FILE	(123)
5.3.11	BEFORE RECORD	(123)
5.3.12	CLEAR	(124)
5.3.13	COLOR	(124)
5.3.14	COMMENT(*)	(126)
5.3.15	COMMENT LEGAL	(127)
5.3.16	CONFIRM	(131)
5.3.17	CONFIRMFIELD	(131)
5.3.18	COPYTOCLIPBOARD	(131)
5.3.19	DEFINE	(132)
5.3.20	EXECUTE	(133)
5.3.21	EXIT	(135)
5.3.22	GOTO	(135)
5.3.23	HELP	(136)
5.3.24	HIDE, UNHIDE	(138)
5.3.25	IF...THEN	(139)
5.3.26	INCLUDE	(140)
5.3.27	JUMPS	(140)
5.3.28	KEY	(142)
5.3.29	LABEL BLOCK	(143)
5.3.30	LABEL	(144)
5.3.31	LEGAL	(144)
5.3.32	LET	(144)
5.3.33	MISSINGVALUE	(146)

5.3.34	MUSTENTER	(147)
5.3.35	NOENTER	(147)
5.3.36	QUIT	(147)
5.3.37	RANGE	(147)
5.3.38	RELATE	(148)
5.3.39	REPEAT	(149)
5.3.40	TOPOFSCREEN	(149)
5.3.41	TYPE	(149)
5.3.42	TYPE COMMENT	(150)
5.3.43	TYPE STATUSBAR	(152)
5.3.44	UNHIDE	(153)
5.3.45	WRITENOTE	(153)
5.4	运算符和函数	(155)
5.4.1	运算符(Operators)	(155)
5.4.2	函数(Functions)	(157)
5.4.2.1	算术函数(Arithmetic Functions)	(157)
5.4.2.2	字符串函数(String Functions)	(160)
5.4.2.3	日期和时间函数(Date and Time Functions)	(161)
5.4.2.4	其他函数	(166)
第6章	录入数据	(167)
6.1	在变量间转换	(167)
6.2	在记录间转换	(168)
6.3	在关联数据库间转换	(169)
6.4	查找记录	(170)
6.5	查找变量和关联变量	(172)
6.6	过滤器(Filter)	(173)
6.6.1	设置滤过规则	(173)
6.6.2	解除滤过	(173)

第 7 章 数据库的管理和维护	(174)
7.1 数据库的追加与合并(Append/Merge Data Files)	(174)
7.1.1 数据库的追加(Append)	(174)
7.1.2 数据库的合并(Merge)	(176)
7.2 双录入核对(Double Entry and Validation)	(178)
7.2.1 选择关键变量	(180)
7.2.2 选项设置	(180)
7.3 逻辑一致性核查(Logical Consistency Check)	(181)
7.4 根据数据库创建 QES 文件(Make QES File from Data File)	(183)
7.5 重新编码数据库(Recode Data File)	(183)
7.6 将年的表示方式从二位数转换为四位数	(185)
7.7 打包数据库(Pack Data File)	(186)
7.8 压缩数据库(Compress Data File)	(186)
7.9 打印数据录入表格(Print Data Entry Form)	(187)
7.10 将数据库存档(Archive)	(187)
7.11 数据库相关信息的管理	(189)
7.11.1 数据库结构(File Structure)	(189)
7.11.2 数据录入备忘录(Data Entry Notes)	(189)
7.11.3 数据库标签(Data File Label)	(190)
7.11.4 浏览数据(View Data)	(190)
7.11.5 数据列表(List Data)	(191)
7.11.6 基本的统计表格(Codebook)	(194)
7.11.7 分变量统计记录数(Count Records by Field)	(196)
7.12 数据库加密	(198)
第 8 章 数据库的输出和输入	(200)
8.1 数据库的输出(Export Data)	(200)

8.1.1	备份数据库(Backup of Data)	(200)
8.1.2	输出到文本文件(Export to Text File)	(200)
8.1.3	输出到 dBase III (Export to dBase III format)	(202)
8.1.4	输出到 Excel(Export to Excel)	(203)
8.1.5	输出到 SPSS(Export to SPSS)	(204)
8.1.6	输出到 SAS(Export to SAS)	(204)
8.1.7	输出到 Stata(Export to Stata)	(204)
8.1.8	输出到新的 EpiData 数据库(Export to New EpiData Data File)	(205)
8.2	数据库的输入(Import Data)	(205)
8.2.1	从文本文件输入(Import Text Files)	(206)
8.2.2	从 dBase 文件输入	(208)
8.2.3	从 Stata 文件输入	(209)
第 9 章	实例	(210)
9.1	建立调查表文件	(210)
9.2	建立核查文件	(218)
9.3	双录入核对	(222)
9.4	导出成 SPSS 等数据库文件	(222)
第 10 章	系统设置	(223)
10.1	选项设置(Options)	(223)
10.1.1	编辑器选项(Editor Options)	(223)
10.1.2	数据录入表格选项(Show Data Form Options)	(224)
10.1.3	创建数据库选项(Create Data File Options)	(225)
10.1.4	输出报表选项(Documentation Options)	(226)
10.1.5	高级选项(Advanced Options)	(227)
10.1.6	文件关联(File Associations)	(229)

10.2 其他	(230)
10.2.1 INI 文件(The .INI File)	(230)
10.2.2 程序参数(Program Parameters)	(230)
10.2.3 语言包	(231)
10.2.4 数据库结构(Datafile Structure)	(232)
10.2.5 快捷方式/鼠标(Short – Cut Keys/Mouse) ...	(238)
参考文献	(242)

第1章 软件概述

EpiData 是一个免费的数据录入和数据管理软件,它主要用于数据录入、数据核对、数据管理和数据报告。该软件的功能是建立数字化的调查表格,使收集的资料信息录入计算机更加方便。该软件不但可以在数据录入过程中对数据中的错误进行核对,而且可以在数据录入完成后对数据进行核对,比如双录入的数据核对。

EpiData 的开发者是丹麦欧登塞(Odense, Denmark)的一个非盈利组织,即 The EpiData Association (<http://www.epidata.dk/>)。程序设计者为 Jens M. Lauritsen, Michael Bruus 和 Mark Myatt。该软件目前有多种语言版本,如丹麦语、挪威语、荷兰语、意大利语、中文、法语、西班牙语、俄语、斯洛文尼亚语、塞尔维亚语、波兰语、葡萄牙语、阿拉伯语、德语、罗马尼亚语、英语等。

EpiData 的工作原理源自 DOS 版本的 Epi Info 6.0,但是工作界面为 Windows 版。EpiData 的安装、运行不依赖系统文件夹中的任何文件,也不会系统在文件夹中安装或替代任何 DLL 文件。程序设置等参数被保存在 EpiData.ini 的文件中。可以通过 setup.exe 在计算机中安装这个程序;也可以直接拷贝 EpiData.exe 文件到计算机中,同样可以运行。

理论上,该程序对录入的记录数没有限制。但实际应用中,记录数最好不要超过 200 000 ~ 300 000 (曾经用 250 000 测试过)。整个录入界面不能超过 999 行。对数值或字符串编码进行解释的文字长度最多为 80 个字符,编码长度最多为 30 个字符。

1.1 EpiData 的历史

EpiData 的研发工作最早由丹麦的 Jens M. Lauritsen 发起。最初是作为 Funen 市开展的“预防意外伤害行动”(Initiative for Accident Prevention)中的一部分。但是为什么会着手研发一种新的数据录入程序呢?

Epi Info 6.0 具备所有需要的数据录入和管理功能。但是,随着视窗(Windows)程序的发展,很多使用者发现很难应付 1990 ~ 1995 年开发的 Epi Info 的 DOS 模块。而商业化程序一般不针对数据的管理,使用起来也没有那么简单,也不具备双录入核对的功能。美国 CDC 的 Epi Info 工作组将 Epi Info 6.0 更新到 Epi Info 2000,更新的 Epi Info 2000 采用了一种全新的工作策略,数据库不再采用过去简单的 ASCII 格式,而是启用 Access 数据库格式。

1999 年底, Jens M. Lauritsen、Mark Myatt 和 Michael Bruus 组成研发小组。Michael Bruus 是一位专业的 Pascal(帕斯卡语言,一种高级计算机编程语言)程序设计员,编程工作主要由他完成。工作小组希望将 EpiData 开发成一个简单、易用、独立的应用程序。这个程序不需要任何专门的数据库系统驱动(基于 dll)。同时,他们希望从社会团体、个人及其他捐助者那里得到资金支持,这样,就可以免费发布这个软件。

1.2 EpiData 与 Epi Info 的兼容性

EpiData 在原理和思路上是基于 Epi Info 的 DOS 版本,参见 <http://www.cdc.gov/epo/epi/epiinfo.htm>。开发 EpiData 的基本思路是要求它的数据文件与 Epi Info 兼容,EpiData 和 Epi Info 6.0

非常近似。很多应用 Epi Info 6.0 创建的数据库基本不需要什么调整就可以用于 EpiData,特别是当 Epi Info 6.0 中只设置了简单的 CHECK 命令(例如,允许值、重复、必须录入、跳转等)时。下面简单介绍两者间的区别。

1.2.1 数据文件中的区别

在 EpiData 中使用 Epi Info 数据文件

这时要注意,EpiData 不支持 Epi Info 中的以下字段类型:

- 电话号码
- 电话分机号码
- 背景和单一录入变量的颜色编码

在 Epi Info 中使用 EpiData 数据文件

这时要注意,EpiInfo 不支持下列 EpiData 字段类型:

- 欧式当天日期 < Today - dmy >
- 日期格式 < yyyy/mm/dd > 和 < Today - dmy >
- 声索引(soundex)变量
- 制表符(@)
- 背景和录入变量等的颜色编码,EpiData 不保存这些设置

字段名中的差异

EpiData 生成的字段名最多为 8 个字符,而 Epi Info 6.0 可达 10 个字符。因为 EpiData 可以读取 Epi Info 生成的数据文件,所以即使数据文件中有的字段是 10 个字符长的字段名,Epi Info 在读取过程中一般来说也不会有太大问题。

然而,如果在 Epi Info 的输入模块使用“使用已修改的调查表修改数据文件的结构”功能,原来在 EpiData 产生的数据文件可能会出现问題。原因是 EpiData 和 Epi Info 都是根据 QES 文件产生

数据文件,而 EpiData 产生的字段名最多可达 8 字符,而 EpiInfo 最多字符可长达 10 个字符。为了保证字段名的一致性,需要在调查表文件中添加“| |”,以保证字段名最多 8 个字符长。

1.2.2 EpiData 的 CHECK 语言与 Epi Info 的区别

在 EpiData 中,条件语句 IF...THEN 中可以使用多个 AND 或 OR,必须用圆括号将每个条件语句括起来。例如:

IF (a1 > 4) AND (a1 < 9) THEN...

另外,EpiData 在一些计算和表达式中的语句与 EPI Info 略有不同。

- 下面这些核对命令在 Epi Info 中没有:确认字段、标记块、自动存盘、确认、标记。

- HELP 命令有一些差别。

- 在某些命令中 (TYPE、HELP 等),颜色码和屏幕一致性在 EpiData 中不一样。

- 日期常数在 EpiData 中必须是 10 个字符且是欧洲形式。

- EpiData 在计算中使用表达式和函数语法有些不同 (如 Copy 和 Paste 等)。

- EpiData 中不支持 Codefield 和 codes;取而代之的是 COMMENT LEGAL 和 TYPE COMMENT,但具有相同的功能。

- Epi Info 不支持 QUIT 和 COPY TO CLIPBOARD (复制到剪贴板)命令。

1.3 软件界面

1.3.1 主程序界面介绍

第一次运行 EpiData 程序,可看到一个欢迎界面,里面有关于 EpiData3.02 的介绍。如果不想在下次运行时再让它出现,可选中“下次不再显示此页”的复选框。

Epi Data 主程序界面分为菜单部分、工具条部分、按钮栏、工作区和状态栏等部分(图 1-1)。

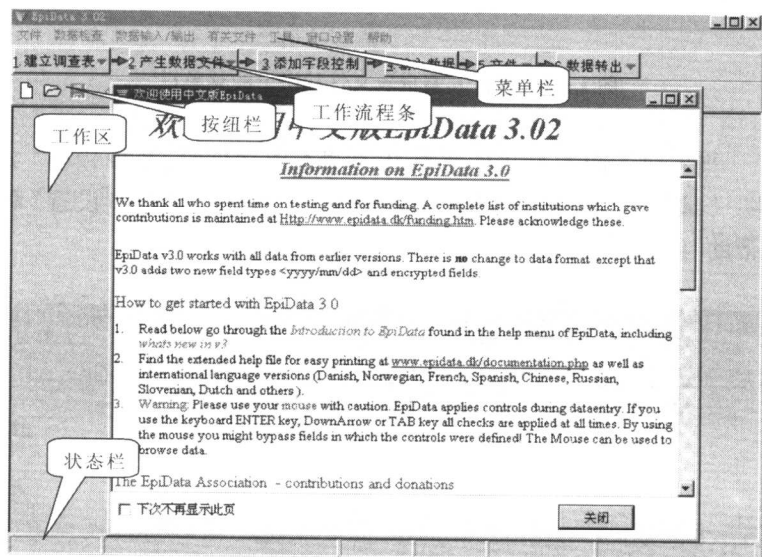


图 1-1 软件界面