

The Epi Info™ Story

Key Events in the History of Epi Info™

1. 1976-1986
2. 1986 - Version 1
3. 1987-1988 - Versions 2, 3, 4
4. 1990-1992 - Versions 5 & 6, EpiMap, EpiTable & EpiNut
5. 1993-1998
6. 1998-2001 - Epi Info 2000
7. Computers and Epi Info Historical Summary to the Third Millennium
8. **All**

1976



The beginning of flexible computing in public health.

Auditorium A at CDC, converted to a war room for the Swine Flu crisis, is filled with epidemiologists and a Digital Equipment PDP 11 minicomputer the size of a refrigerator.

A program called SOCRATES, written in FORTRAN by programmer Rick Curtis, allowed an epidemiologist to define questions, enter data, and summarize the results in tabular form without the aid of a programmer.

1978-1984

The Epidemiologic Analysis System (EAS), a more flexible version of SOCRATES, was written in BASIC by Anthony (Tony) Burton of CDC and set up on a minicomputer for remote access by the Georgia state health department.

The Conference (now Council) of State and Territorial Epidemiologists formed a Computer Working Group, and, together with Drs. Keewhan Choi and Stephen Thacker of CDC, developed a five-point plan for development of epidemiologic computing.

1985 - 1986

Jeffrey A. Dean, a Junior at Paidea High School, began a field internship in computing at an insurance company, which required keeping a large printer supplied with paper. He expressed a desire for a more challenging experience to his Dad, who gave him specifications for a data entry program for epidemiology.



During March 1985, he wrote ENTER in Turbo Pascal, the language he had previously used to adapt a mainframe version of Dungeons and Dragons to a microcomputer.



Tony Burton, author of the BASIC version of the Epidemiologic Analysis System, provided ideas for the first version of ANALYSIS, programmed by Jeff during the summer of 1986.

Several years later, Tony moved to the Global Programme on AIDS, World Health Organization, Geneva, and he and Jeff continued to collaborate on later versions of Epi Info™ as Jeff traveled to Geneva during college vacations.

Dr. Richard Dicker, author of the Epidemiology in Action training manual and coordinator of the Epidemic Intelligence Service (EIS) course at CDC, provided advice on the statistics that were included with Epi Info™.



1984

Dr. Andrew Dean, Chairman of the CSTE Computer Working Group, served as President of CSTE and then was invited by Dr. Howard Ory to come to CDC and develop software for epidemiologists.

He produced plans for a series of programs called EPI-AID to assist epidemiologists in outbreak investigations.

Dr. Andy Dean, State Epidemiologist of Minnesota, became frustrated with the lack of flexibility of computer resources in the mid-1970s and vowed to join the microcomputer revolution.



Although, as Director of Disease Prevention and Control, he was unable to commit daytime hours to the effort, he built and programmed an IMSAI computer in the evening.

After several years, the project began to approach the useful stage. At about the same time, he was invited to come to CDC and produce software for epidemiologists.

1986 - Version 1

```
Epidemiologic Analysis System
Centers for Disease Control
Epidemiology Program Office
Jeffery A. Dean - Author
September 1, 1985

A. Data Entry and Editing (EnterInfo)
B. Data Conversion Routines (Convert)
C. Data Output and Analysis (Crosstab)
D. Return to Master Menu

Enter option letter desired:
C:\EAS>
```

Menu of the first version of Epi Info™,

implemented as a DOS batch file.

1986 - User Manual

EAS> TABLES VANILLA ILL

VANILLA		ILL	
	N	Y	Total
N	18	3	21
Y	11	43	54
Total	29	46	75

EAS>

A 27-page manual was added in February 1986, and the "Programs for Automatic File Definition, Data Entry, and Tabulation" were now called Epi Info™.

Distribution was informal; anyone who wanted the programs was given a photocopy of the manual and a 5 1/4 inch floppy diskette.

1986 - 1st Field Trial

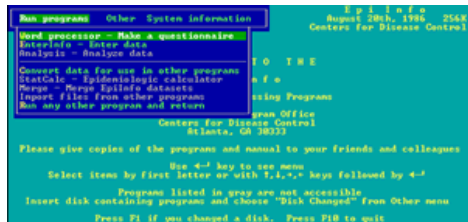
The first field trial of Epi Info™ and its companion "artificial intelligence" system, Epiaid, was conducted in the investigation of a syphilis epidemic in Fort Lauderdale, Florida, by Dr. Consuelo Beck-Sagué, then an EIS Officer.



The idea of high-level computer assistance in planning an investigation never caught on, and the technology of the time did not make the Epiaid programs easy to use.

Epi Info™, with its practical tools for constructing questionnaires and entering and analyzing data, became ever more popular over the next few years.

1987 - Version 2



Epi Info™, Version 2, had a bound manual of 112 pages,

a pull down menu, the EPED word processor for creating questionnaires, ENTER, ANALYSIS, an epidemiologic calculator called STATCALC, EPICURVE for plotting histograms, a MERGE program, and IMPORT and CONVERT programs for file conversions.

```
EPI> read c:\epi5\data\paluf.in
EPI> tables sexe palu

SEXE   0   PALU   1   Total
-----
1      354   62      416
2      458   94      552
9       17    4       21
Total  829   168     989

An expected cell is < 5. Chi square is not valid.
Chi square = 174738.22
2 degrees of freedom.
p value = 1.00000000

EPI>
```

The TABLES command began to offer statistical output. Note the negative chi square, probably a

result of limitations in integer values in an early version of Turbo Pascal.

The first versions of Epi Info™ were distributed by Andy, Jeff, and Tony from their office.

Suggestions came from users in the field and were rapidly incorporated into new versions of the software.



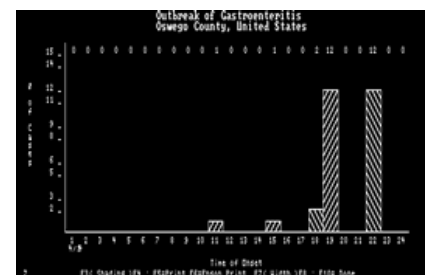
1988 - Version 3

Epi Info™, Version 3, came with a 150-page manual, enhanced statistics, and sample programs, including two outbreaks and a surveillance system.

The programs and manual remained in the public domain, but copies were also marketed by a commercial firm in Atlanta –USD, Inc.– without any special arrangement with CDC.

This provided a means of distribution that did not require copying by CDC office staff or permission to distribute through government channels.

The graphing of epidemic curves was a popular feature (despite an optical illusion caused by the diagonal line pattern).



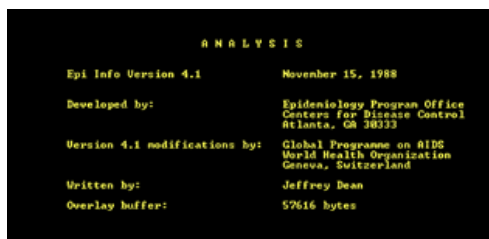
1988 - Version 4 CDC and WHO Collaboration



WHO, Geneva

CDC, Atlanta

Epi Info™, Version 4, was developed by Tony Burton, who had moved to the World Health Organization, Geneva, and by Jeff Dean, who worked there during summer vacations.



It had greatly enhanced features, but also a number

of bugs, and was used as the basis for Version 5, released from CDC.



Chris Smith



Karl Brendel



Tom Arner

1990 - Version 5

Epi Info™, Version 5, featured a manual of 384 pages, that described missing value indicators, linear regression, a variety of graphs, analysis of matched data, a report generator, a relational file system, and features for nutritional anthropometry.

It allowed European date formats and non-English characters in data fields, and became the first version to be widely distributed internationally.



The Turbo Pascal source code was sent to volunteer translators who recompiled and distributed the programs in French, Spanish, and Chinese. An Arabic translation was done as a commercial venture, and the Russian version was done under contract to CDC.



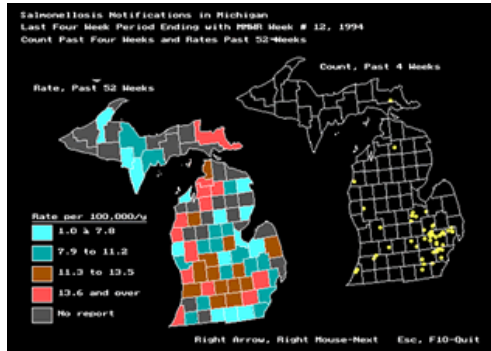
1992 - Version 6

Epi Info™, Version 6, had a 600 page manual, many more programming features, improved graphics, exact statistics, extensive examples, a



1992 - Epi Map, Version 1

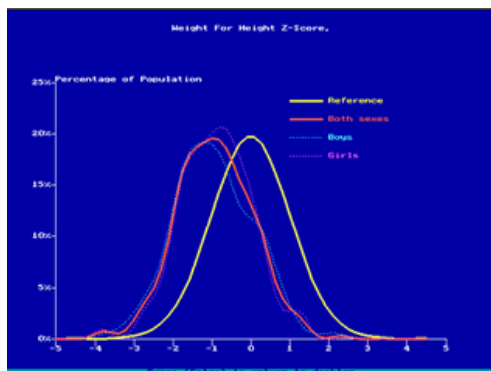
Epi Map, Version 1, was developed by Jeff Dean and Tony Burton in Geneva, with debugging and enhancements by Karl Brendel at CDC. Version 2, developed at CDC, added features for programming what happens after mouse clicks on geographic areas.



1992 - EpiTable & EpiNut

Dr. Denis Coulombier contributed two new programs, EpiTable and EpiNut, for statistical processing of tables and nutritional anthropometry, and many programming examples to Version 6. This version and several minor upgrades provided a solid basis for individual public health computing through the decade of the 90's and for worldwide distribution.

Graphs of nutritional values of a population compared with an international reference population in Nutstat.



programmable menu system, and a mini reportable disease surveillance system for a state health department.

```

CTRL [F1]-Begin [F2]-End [F3]-Copy [F4]-Move [F5]-Delete [F6]-Find [F7]-Repl [F8]-Format [F9]-Menu
Questionnaire for Church Supper - Ouzage
Age:### Sex:(M) Time of Supper(24 hour):####
1117 (Y) (Onset Date) (MM/DD) (mm/dd) (Onset Time) (24 hour):####
Food History (Mark Y or N):
Baked Ham:(Y)
Eggs:(Y)
Marbled Potatoes:(Y)
Cabbage Salad:(Y)
Jello:(Y)
Beef:(Y)
Bread:(Y)
Milk:(Y)
Coffee:(Y)
Water:(Y)
Cakes:(Y)
Ice Cream: (Vanilla): (Y) Chocolate:(Y)
Fruit Salad: (Y)
    
```

An Epi Info™, Version 6, questionnaire. Text defines the questions or prompts. Special symbols like ### and indicate numeric or Yes/No fields for data entry. Text fields are indicated by underline characters, but there are none in this questionnaire.



Robert Fagan, systems analyst at CDC, used Epi Info™ to develop reportable disease software eventually used in 42 states and territories in the US.

```

CONFIDENTIAL MORBIDITY REPORT
ID: 00000001 UPDATED: 05/23/94
LAST NAME: SMITH FIRST: FRED
ADDRESS: 1102 PARK PLACE SE CITY: ATLANTA
COUNTY: DEKALB COUNTY CODE: 83 REGION:
ZIP CODE: 30105 BIRTHDATE: 02/22/24 AGE: 69 AGE TYPE:
RACE: ETHNIC: SEX:
EVENT NAME: SHIGGONOSIS EVENT ICD9: CASE COUNT:
EVENTTYPE: EVENT DATE: 02/22/94 OUTBREAK ID:
MMWR WK: 10 YR: 94 REPORTED: 05/23/94
OTHER DATA: IMPORTED: REPORT STATUS:
LASTNAME: (You must enter data) All entries allowed
COPY: New COPY: Find COPY: Print COPY: Delete COPY: Choices COPY: Done COPY: Help
    
```

Analytic results from Version 6 of Epi Info™ include exact confidence limits for the odds ratio...

```

Dataset: C:\EPI6\000000_REC (7% records) Free memory: 292k
Criteria: All records selected. Output - Screen Time: 0.00 sec
----- tables vanilla ill
VANILLA + - Total
+ 43 11 54
- 3 18 21
Total 46 29 75
Single Table Analysis
Odds ratio
Cornfield 95% confidence limits for OR
May be inaccurate 5.07 < OR < 23.45
Maximum likelihood estimate of OR (MLE) 22.15
Exact 95% confidence limits for MLE
Command:
EPI6> tables vanilla ill
EPI6>
F1-Help F2-Commands F3-Variables F4-Browse F5-Printer on F9-DOE F10-Quit
    
```

...and an explanation of the Risk Ratio, as well as the statistics from previous versions. The Fisher Exact test is not displayed here because none of the expected values of the cells is less than 5.

```

Dataset: C:\EPI6\000000_REC (7% records) Free memory: 292k
Criteria: All records selected. Output - Screen Time: 0.00 sec
Exact 95% Mid-P limits for MLE
Probability of MLE >= 22.15 if population OR = 1.0 0.00000026
RISK RATIO(OR)(Outcome:ILL=1; Exposure:VANILLA=>) 5.57
95% confidence limits for RR 1.94 < RR < 16.83
Ignore risk ratio if case control study
Chi-Squares P-values
Uncorrected: 27.22 0.00000018
Mantel-Haenszel: 26.86 0.00000022
 Yates corrected: 24.54 0.00000073
Command:
EPI6> tables vanilla ill
EPI6>
F1-Help F2-Commands F3-Variables F4-Browse F5-Printer on F9-DOE F10-Quit
    
```


1993 - Conference

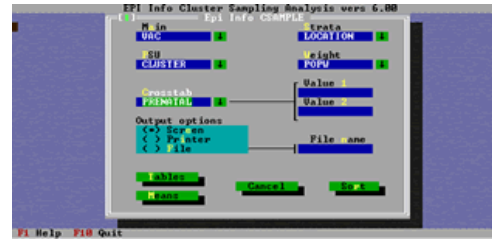
A conference on "Microcomputers and the Future of Epidemiology" convened 130 public health experts in Atlanta and provided the basis for designing a Windows version of Epi Info™.



Dr. Andy Dean on the left poses with

translators of Epi Info™ into Spanish (Juan Carlos Fernández Merino), Chinese, French (Dr. Robert Freund), Indonesian (Dr. Pandu Riono), and Arabic (Dr. Samy Sidki).

A new program in Version 6 processed data from complex sample designs, such as cluster samples used by WHO in immunization surveys, and stratified cluster samples frequently used in professionally designed surveys.



1994 - Internet Distribution & HelpLine
Internet distribution of Epi Info™ began, making it available to a wide international audience without charge.



"It is a generally accepted rule in the software business that producing a working prototype is only 20% of the cost of making a commercial product." (from a report of the Working Group on Biomedical Computing, Advisory Committee to the Director, National Institutes of Health, June 3, 1999).

Much of the work in making Epi Info™ useful on a range of computers worldwide in the hands of many different kinds of users is in debugging, testing, documenting, obtaining user feedback, and providing technical support to users.

An important feature in Epi Info's™ acceptance is the Epi Info™ Hotline. Two Hotline staff members respond to 20 to 40 inquiries per day by telephone, FAX, and e-mail. The Hotline not only provides support and builds confidence among users, but also is a channel through which problems and suggestions become available to the development team.

1996 - Translations & Global ListServ
Greg Fegan transferred the Epi Info™ Worldwide Discussion LISTSERV from Tulane University to CDC which brought about 400 interested Epi Info™ users in touch with each other and with the latest information.

Experienced users regularly responded to questions posed by others in the group. Collaboration via e-mail became commonplace, with users or translators in China, South and Central America, Europe, or India.



Top row: Epi Info™ manuals in German, French, English, Spanish, and Italian.

Middle row: A needle-stick surveillance system, an Arabic manual, Version 6 in English, Czech and Indonesian manuals.

Bottom row: Norwegian, Chinese, Portuguese, and Epi Map manuals.

The manual or programs for Epi Info™ for DOS became available in 14 languages.

1997 - Distribution Evaluation

During the Summer of 1997, an evaluation of Epi Info™ distribution was performed by Braddee Harbage, a student intern, using e-mail and the Internet to contact as many distributors of Epi Info™ as possible. Results included:

- A minimum estimate of 145,320 copies of Epi Info™ and Epi Map distributed in more than 117 countries
- 1207 citations in the scientific literature
- 52 Internet sites providing copies of the Epi Info™ programs, manual, or related materials
- An estimate that 66% of copies were distributed by Internet

1998 - Year 2000 Compatibility

Version 6.04 b-to-c upgrade provided "Year-2000 Compatibility" for the DOS version of Epi Info™ by strengthening the 4-digit year features and providing a utility for upgrading previously created data files.

1998 - Epi Info™ 2000 for Microsoft Windows 95/98/NT



The Epi Info™ 2000 Development Team: Tom Arner, Karl Brendel, Andy Dean, Sireesha Sangam, and Godha Sunki.

2001 - Version 6 Bug Fix

Fast processors disclosed a bug in the original Turbo Pascal software. The programs were recompiled with a patched version, and are now available as Version 6.04d, consolidated on 3 diskettes. It is important to download and install the new version for use on any machine with modern processing speeds (more than about 400 mhz).

Computers & Epi Info, Historical Summary: Before 1980

Before 1980, a few epidemiologists used mainframe computers or struggled to assemble, program, and use microcomputers.

Minicomputer software allowed CDC epidemiologists to develop questionnaires and analyze data through terminals.

Computers & Epi Info, Historical Summary: The 80s

In the 80s, microcomputers and local area networks became standard public health tools.

Epi Info™ and other database and statistics programs became standard equipment for public health. Epi Info™ was used for surveillance systems, disease investigations, surveys, clinical research, nutritional anthropometry, and veterinary and dental public health.

Translations of the programs or manual appeared in 13 non-English languages.

Worldwide copies exceeded 145,000 in 117 countries.

1999 - Epi Info™ 2000 Beta

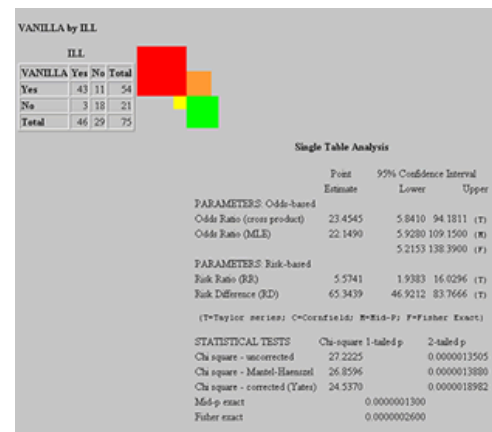
Roger Friedman and Matt Lantinga joined the team in 1999. Roger developed the Epi Info™ 2000 Analysis module, and Matt programmed the NutStat module.

May 13, 1999.

The Beta-Test Version of Epi Info™



2000 for Microsoft Windows 95, 98, and NT was released to a limited audience for testing.



Output from

ANALYSIS in Epi Info™ 2000 is in HTML for easy display or printing in Internet browsers.

Computers & Epi Info, Historical Summary: The 90s

In the 90s, remote areas of the world, starved for scientific information, suddenly began to share information with scientists in major capitals via the Internet.

The Internet became the major route for distribution of Epi Info™, and approximately 400 users shared information via the Epi Info™ LISTSERV Discussion Group. The Epi Info™ Hotline began to serve users worldwide.

Computers & Epi Info, Historical Summary: Epi Info™ 2000

Epi Info™ 2000 for Windows 95, 98, and NT will bring the benefits of hands-on questionnaire creation, database management, statistics, mapping and graphing to Windows, with Internet enhancements to follow.

Computers & Epi Info, Historical Summary: 1999

By 1999, EIS Officers and summer students at CDC regularly learn to process data after 1 or 2 days of instruction.

More than half of U.S. households have computers.

The Third Millennium

Will there be...

- Fast and information rich connections among scientists?
- Full-text access to scientific literature worldwide in all major languages?
- Methods for accessing and analyzing data from clinical sources for public health work?
- Communication with households via addressable video/computer services?
- Universal medical and public health records with voice input?