Brief Reports

Distribution of Epi Info Software
An Evaluation Using the Internet

Braddee Harbage, MPH, Andrew G. Dean. MD, MPH

Introduction: Epi Info and Epi Map are computer programs for word processing, database management, epidemiologic statistics, and mapping designed for public health professionals. The programs are in the public domain, and distribution outside the Centers for Disease Control and Prevention (CDC) has been through a variety of informal channels.

Method: Individuals and organizations known to have distributed Epi Info or Epi Map since 1987 provided information. Distributors included CDC, the World Health Organization (WHO), commercial vendors, translators, instructors in university and public health settings, and other public health professionals. Reports documented a minimum number of 145,320 copies distributed.

Conclusions: Since 1994, the Internet has become a major means of propagation, accounting for 66% of the copies for which the method of distribution was known. The Internet also was a major information source for this study.

Medical Subject Headings (MeSH): software, epidemiology, diffusion of innovation, evaluation, Internet, market research, Epi Info (Am J Prev Med 1999;16(4):314–317) © 1999 American Journal of Preventive Medicine

Introduction

The first versions of Epi Info,[1,2] beginning in 1985, were designed to allow individual epidemiologists to enter and analyze data and produce statistical results without the assistance of computer professionals. Versions 5 and 6 added programming features for public health surveillance and other permanent applications. Companion programs include Epi Map for representation of epidemiologic counts or rates on geographic maps; DoEpi, a series of interactive computer exercises for teaching epidemiology and computing; and Statistical Software for Public Health Surveillance (SSS1) containing programs for time-series modeling and other surveillance statistics. All the programs were developed for the DOS operating system, but can be run as DOS programs under Microsoft Windows™ 3.1, Windows 95™, and Windows NT™.

Development of Epi Info and related programs was supported by Centers for Disease Control and Prevention (CDC) and World Health Organization (WHO). CDC provides technical support by telephone, fax, and e-mail. Epi Info or its manual has been translated into 13 non-English languages, primarily by volunteer translators in other countries.

Because Epi Info is produced by the U.S. government, it is in the public domain, and may be freely distributed, copied, translated, or sold by others. CDC provides copies to its own employees and makes the product available on the World Wide Web, but there is otherwise no formal relationship with distributors. Source code is provided to translators, who must purchase licenses to commercial source code libraries contained in the programs from their copyright owners. At least two small commercial publishers have reproduced and distributed copies of the programs and manual, and most translators have arranged for distribution of the product at no or modest cost.

Distribution of Epi Info over the World Wide Web of the Internet began in 1995. CDC's Web pages provide information on Epi Info, facilities for downloading copies, brief tutorials, and hypertext links to related Web sites outside of CDC. An Epi Info discussion group is maintained as a LISTSERV. Subscribers receive copies by e-mail of messages from other subscribers, including CDC staff.

To evaluate the distribution of Epi Info, we conducted a study of known distributors of the product. viewing transmission of Epi Info as the event of interest. and assessing barriers and aids to dissemination to the extent this is possible in a descriptive study. The investigation concerns the diffusion of innovation and ideas in the public domain, as in social marketing

From the Division of Public Health Surveillance and Informatics. Epidemiology Program Office, Mailstop C08. Centers for Disease Control and Prevention, Atlanta, Georgia.

Address correspondence to: Dr. Andrew Dean. Mailstop C08. Division of Public Health Surveillance and Informatics, Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta. Georgia 30333.

research,[3] but also resembles commercial market research in measuring the acquisition of a concrete product (3 diskettes and a 600-page manual, or their electronic equivalent).

The purpose of this study was to obtain a minimum estimate of the number of copies of the Epi programs throughout the world, to determine how copies are transmitted, and, if possible, what factors facilitate or retard their diffusion through public health circles. Knowledge of these factors will be used to design future distribution methods.

Materials and Methods

The study was performed during a 9-week period in the summer of 1997. The event of interest was defined to be reported or otherwise documented transmission of the programs and/or manual for Epi Info or Epi Map from 1985 through September 1, 1997, anywhere in the world. Persons known to have translated, distributed, copied, or sold copies of Epi Info or Epi Map were contacted by mail, e-mail, or telephone. Lists of contacts for the study included translators, Webmasters, and commercial distributors of public domain copies. Subscribers to the Epi Info World Wide Web Discussion Group LISTSERV facility (with about 360 members, mostly with public health backgrounds) were notified of the project, and information regarding the propagation of the product was requested. One hundred-one reports were received from persons who had distributed the programs or manuals. Suggestions by distributors concerning others distributing the program were followed up to obtain the broadest coverage possible, but resource limitations and regulations concerning government surveys did not allow systematic study of individual Epi Info users in the available time frame.

An Internet search was performed for World Wide Web sites referring to and/or distributing Epi Info and related programs. This was accomplished through the use of metasearch engines at http://www.netseek.com/metaseek/homepage.html, and individual search engines at http://www.tomco.net/~cprags/search.htm. The metasearch engines included Netseek, Metacrawler, Starting Point, and Savvy Search. The single search engines included Excite, Infoseek, Lycos, Yahoo. Alta Vista, Galaxy, Hot Bot, and Open Text Index. Items entered for all searches included multiple spellings of the target phrases ("Epi Info," "Epi-Info," "EpiInfo," "Epi Map," "Epi-Map," "EpiMap," "Do Epi," "Do-Epi." and "DoEpi"). A search was also done using these names in combination with all words beginning with "epidemi." Similar combination searches were done for SPSS™ and SAS™—commercial statistical packages often used in epidemiology. Sites distributing Epi Info or related programs or manuals were identified from among the larger number that mentioned the target phrases.

Summary data were available for distribution from the CDC Internet site for the period from February 6, 1995, through September 1, 1996. Subsequent counts required processing log files for the entire CDC Internet site comprising hundreds of megabytes of files. Log files of approximately 10 megabytes each were provided by CDC's central Internet facility for May and July 1997. These were processed to count downloads of Epi Info and related programs. For programs contained in multiple files, counts of the last of the series were taken to indicate successful downloads. Mean monthly counts for these 2 months were 1,642 for Epi Info, 358 for Epi Map, and 446 for DoEpi. These rates were used to calculate downloads for the year ending September 1, 1997.

For each source of information, a record was created in an Epi Info, version 6.04b, database containing the number of copies distributed. A total of 102 records were created from 37 sources. Each record included starting and ending dates, the program and version, language, the identity of the distributor, and the format of the copies (hard copy vs. Internet). Downloaded copies were counted only from sites that offered independent downloading rather than links to another site.

Results

The total number of programs and manual transmissions documented was 145,320. The number of copies distributed by a single source varied from 1 (reported by 3 individuals) to 26,282 (from the CDC Internet site). Copies of Epi Info comprised 90% of the total 145,320 copies, the remainder being copies of Epi Map (9.8%, 14,185 copies), the teaching program DoEpi (1.6%, 2281 copies) or Statistical Software for Public Health Surveillance (SSS1) (0.7%, 1081 copies). The rest of the results will consider all four programs together as "Epi Info," except where noted.

Before 1994, all copies distributed were either copies of diskettes and/or a printed manual. Tabulating records that began in 1994 or later gave a different picture, with 66% of copies downloaded from the Internet among those copies where the distribution method was known. Over the entire period of the study, Internet distribution accounted for 29% of total distributions.

Translations of Epi Info and/or its manual have been produced in 13 languages: French, Spanish, German, Portuguese, Italian, Chinese. Arabic, Russian, Norwegian. Polish, Czech, Hungarian, and (reportedly) Farsi. The Arabic translation was produced by a private entrepreneur, and the Russian version was supported by funding through CDC; the other translations represent a number of person-years of effort on the part of dedicated volunteers from public and private institutions. The 145,320 copies reported included 120,636 in English (83.0%), 8692 in Spanish (6.0%), 6352 in

French (4.4%), 2500 in Portuguese (1.7%), 1400 in Italian (1.0%), 1300 in German (0.9%), 1160 in Norwegian (0.8%), 780 in Chinese (0.5%), and 2500 where language was not reported (2.0%).

CDC distributed 38% of the copies and WHO 9.3%. Hard copies distributed by CDC were for employees only; 85% of CDC's total distribution was via the Epi Info World Wide Web site, which serves internal and external clients. Commercial vendors distributed 23% of the copies, and translators, 25%. Individuals known to the staff or who responded to our inquiry on the Epi Info Worldwide E-mail Discussion Group LISTSERV provided 20 reports comprising 4.5% of total copies. One half of the individuals who responded had distributed 90 or more copies, and one reported distributing 1,000 copies.

CDC's Epi Info Technical Support Hotline handled an average of 701 inquiries per month during 1997. One third of the inquiries were via e-mail, and e-mail also provided a mechanism for transmitting files and programs for examination by the Hotline staff. A search of reference citations in the scientific literature disclosed 1207 that included "Epi Info" or one of its spelling variations.

We found 52 sites from which Epi Info or related programs and manuals could be downloaded from the Internet. Some of these sites linked the user to CDC's Epi Info Web site, and others were mirror sites maintaining copies on a local server. Six Web sites, including CDC's, offered complete downloading of program, manuals, and information related to Epi Info. There were 1296 World Wide Web documents that mentioned "Epi Info" or related programs. Epi Info and related programs were referred to on the Web, in the same document with "epidemi" 525 times. "SAS" and "epidemi" appeared 1928 times, and "SPSS" and "epidemi" 580 times.

Discussion

The event under study was dissemination of a copy of Epi Info or Epi Map or a manual by an organization or individual. The study was limited to known users or distributors of Epi Info and those who could be identified through this group or located on the Internet. It is likely that the proportion of copies distributed by individuals would have been much larger in a representative survey of public health professionals, because the few individuals included reported distributing large numbers of copies, and no attempt was made to contact others who may have given a few copies to friends or taught classes during which a number of copies were given out. The true number of copies may be several times the number documented in this study, even if allowances are made for receipt of several versions by single individuals. No reports were received from translators in several languages, and the existence of a manual in Farsi was documented only through personal communication from a visitor to Iran.

Although the numbers obtained represent only a minimum estimate, it is clear that Epi Info and Epi Map have found a worldwide audience. One vendor's records documented sales in 117 countries. Although the searches for mention of various programs in combination with words beginning with "epidemi" suggest that Epi Info is mentioned almost as frequently as SPSS and about one fourth as often as SAS on Web pages likely to deal with epidemiology, this is only a crude indicator ("SAS" is also an airline, for example). It is also likely that the commercial programs appeal to a different audience. One of Epi Info's main selling points is its freedom from restriction on making copies, making it particularly useful in small health departments, in developing countries, and for ad hoc use.

Commercial vendors of Epi Info accounted for less than one third of the distribution, but their role in providing printed copies of the manuals and physical diskettes is extremely important. Although the Internet has assumed an important role in recent years for distribution and serves as a source of copies for other distributors, many users do not have ready access to the Internet or the skills needed to download and install programs. The availability of printed manuals and diskettes serves as a reservoir for additional distribution and copying and a source of updates when these occur.

The Internet has transformed the distribution of Epi Info. making updates and translations immediately available around the world. E-mail provides a forum through which much valuable advice and discussion is obtained and provided by users. Increasingly, the design and testing of programs is conducted through international collaboration via e-mail.

The Internet has many advantages for the distribution of public domain programs, for which mechanisms of recovering costs are difficult or absent. Internet distribution eliminates the need for upfront financing of printing runs; the storage of manuals and diskettes; the cost, uncertainty, and delays of international shipping; and the overhead of accounting and billing. More importantly, it distributes the cost of copying to the user. making it possible for a government agency to distribute software merely by supporting a Web page. Future versions of Epi Info should be designed to take advantage of this distribution method by minimizing the need for printed manuals and providing help files suitable for use on the computer screen rather than as printed manuals. More automatic mechanisms should be installed to document downloaded copies from the CDC Internet site.

The first few versions of Epi Info required that users understand English. This characteristic limited its spread to English-speaking countries and the relatively small group of public health workers elsewhere who are familiar with English. Translators soon created inten-

tional mutations of the programs to other languages, allowing spread to continue among those more comfortable with one of many languages other than English. Future versions of Epi Info should be designed for easy translation to facilitate this process.

For evaluation purposes, minimum values of the number of copies distributed can be obtained by methods such as those used in this investigation. Overcounting can result from including several copies obtained by the same user, and undercounting through missing the large number of copies distributed "hand-to-hand" by users. Commercial products can evaluate success by sales in cash terms and through registrations that offer or seem to offer an advantage to the user, such as technical support or a warrantee. Approximately 7% of our copies were documented through voluntary regis-

tration by users for purposes of notification concerning updates and future products. An easy and automatic registration method is needed to provide information on the number of copies distributed on a worldwide basis for future versions of Epi Info.

References

1. Dean AG, Dean JA, Burton AH, Dicker RC. Epi Info: a general-purpose microcomputer program for public health information systems. Am J Prev Med 1991 May; 7(3):178–82.
2. Dean AG, Dean JA, Coulombier D, et al. Epi Info, version 6: a word processing, database, and statistics program for public health on IBM-compatible microcomputers. Atlanta, GA: Centers for Disease Control and Prevention (CDC), 1995.
3. Walsh, DC, Rudd RE, Moeykens BA, Moloney TW. Social marketing for public health. Health Affairs 1993;12(2):104–19.