Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.™

Morbidity and Mortality Weekly Report (*MMWR*)

History of Statistics in Public Health at CDC, 1960--2010: the Rise of Statistical Evidence

Supplements

October 7, 2011 / 60(04);35-41

Donna F. Stroup, PhD[1]

Rob Lyerla, PhD[2]

[1]*Data for Solutions, Inc., Decatur, Georgia*

[2]*Office of the Global AIDS Coordinator, US Department of State*

Corresponding author: Donna F. Stroup, PhD, Data for Solutions, Inc., Post Office Box 894, Decatur, GA 30031-0894; Telephone: 404-218-0841; Fax: 404-377-9696; E-mail: donnafstroup@dataforsolutions.com.

"A ... firm grasp of the statistical method was as essential part of the outfit of the investigator in that field [epidemiology] as was a grounding in bacteriology."---Anonymous, 1913 (1)

Introduction

It is difficult for us to imagine the report of an epidemiologic investigation without at least one 2×2 table, p value, or odds ratio. We now recognize that an understanding of mathematical methods and the use of statistics to assess data in epidemiology and public health are critical for identifying the causes of disease, modes of transmission, appropriate control and prevention measures, and for prioritizing and evaluating activities.

When CDC was established in 1946 (as the Communicable Disease Center), the U.S. Public Health Service borrowed statistical methods developed by Florence Nightingale and Edwin Chadwick, who had applied these techniques to implement sanitary measures in London (*2*). Based on William Farr's use of statistical induction to analyze death rates (*3*), Karl Pearson's creation of goodness-of-fit tests and correlation methods, and Bradford Hill's development of guidelines for establishing causal relationships (*4*), Nightingale employed statistics in her efforts to reform the British military health-care system through the founding of training programs and definition of sound professional standards (*5*).

During the 1950s, CDC's activities emphasized the work of sanitarians and laboratory scientists, and the analytic component of most epidemiologic investigations rarely went beyond descriptive analysis and 2×2 tables. However, with the establishment of the Epidemic Intelligence Service (EIS), rapid response to outbreak investigations, and involvement of mathematical experts, epidemiologic methods advanced (*6*). Case-control studies were used routinely by EIS officers. An investigation of *Staphylococcus* in a newborn nursery was the first CDC report to include a chi-square statistic and a p value (CDC, unpublished data, 1957). By the middle of the decade, an early dose-response analysis was included in an investigation of hepatitis in a housing project (CDC, unpublished data, 1956).

The 1960s

With the acquisition of *MMWR* in 1961 under Alexander Langmuir's leadership, CDC had a vehicle for influencing the practice of biostatistics. Langmuir's training under Wade Hampton Frost, the first professor of epidemiology in the United States at the Johns Hopkins University School of Hygiene and Public Health, led to Langmuir's emphasis on quantitative foundations for public health and the need to link data acquisition with practical application through the practice of public health surveillance (*7*).

During this decade, the first *t* test in an epidemic-assistance investigation (Epi-Aid) is found in Carl Norden's report of infectious mononucleosis in Kentucky (CDC, unpublished data, 1963). The first pie chart appears in James Bryan and Ron Roberto's Epi-Aid for suspected poliomyelitis in the Marshall Islands (CDC, unpublished data, 1963). During this period, the vast majority of requests for Epi-Aids collected data through convenience survey methods or used existing surveillance data. In only two of 502 Epi-Aids was the method of randomization reported. Calculations were restricted to those that could be done by hand or later on programmable calculators (Figure 1). Eventually, however, surveillance and other data analyses used mainframe computers and the punched card throughout the late 1960s.

The 1970s

In 1970, the Communicable Disease Center's name changed to the Center for Disease Control. Beyond semantics, this represented a broadening of the mission beyond communicable diseases. In 1971, the National Center for Health Statistics (not yet part of CDC) conducted the first National Health Assessment and Nutrition Examination Survey (NHANES). The National Institute for Occupational Safety and Health joined CDC in 1973 and brought use of methods for noninfectious conditions, such as large population-based studies.

This expansion of activity to environmental and occupational problems brought expanded opportunities for the contribution of statistical and engineering methods to public health. One example is the use of NHANES data combined with data on lead in gasoline from the U.S. Environmental Protection Agency to develop a model to predict human blood lead levels (*8*). The results were used to provide evidence that subsequently led to a ban on the use of lead in gasoline in the United States.

In 1974, CDC assumed leadership of a major national immunization campaign. Although the theory behind herd immunity was developed during the 1920s, the development of vaccines coupled with advances in mathematical modeling in epidemiology found a new synergy in a paper written in 1971 (*9*). Four years earlier, in 1967, the World Health Organization had declared its intent to eradicate smallpox within 10 years, and the U.S. Public Health Service had declared its intent to eliminate measles from the United States within 1 year (*10*). Both of these tasks were theoretically to be achieved by the induction of herd immunity with vaccines.

The year 1976 saw the beginning of flexible computing in public health. To address the swine flu crisis (*11*), an auditorium at CDC was filled with epidemiologists and a Digital Equipment PDP 11 minicomputer the size of a large refrigerator. A program called SOCRATES, written in FORTRAN, allowed an epidemiologist to define questions, enter data, and summarize the results in tabular form without the aid of a programmer or a trip across campus to a mainframe computer. The SOCRATES program later formed the basis of another program, the Epidemiologic Analysis System, which was an early forerunner of Epi Info(tm), a suite of lightweight software tools for use in field epidemiology first released by CDC in 1985 (see below).

The 1980s

In the 1980s, public health saw an expansion of emphasis on statistical methods and more statistical sophistication among epidemiologists and analysts. The computer-punched card was gradually replaced as the primary means for data storage by magnetic tape, as better computers became available (Figure 2). Punched cards were still commonly used for data entry and programming at CDC until the mid-1980s, when the combination of lower-cost magnetic disk storage and affordable interactive terminals on less expensive minicomputers made punched cards obsolete. However, their influence persists through many standard conventions and file formats. For example, the terminals that replaced the mainframe card readers displayed 80 columns of text, the same amount of space on the punched card.

The first report in *MMWR* containing results from a logistic regression model appeared in 1982, only 3 years after the software package BMDP provided the LOGIT routine as part of its software (*12*). In this investigation of typhoid fever in Michigan, the model was unable to identify risk associated with any food item because of a small number of cases and little variation in food-consumption patterns. Since this first use, logistic regression has become a standard technique in public health and has contributed to policy formulation in many areas. For example, the results from a logistic regression analysis were used to implement a requirement that tobacco-control programs should include opportunities for community participation and interaction for maximal impact. (*13*).

In the early 1980s, CDC launched a major case-control study as part of the nascent investigation of human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) (*14*), which provided a platform for development of new statistical methods for surveillance and estimation of disease incubation periods (*15*). A major challenge for HIV/AIDS surveillance was poor data quality due to underreporting, reporting delay (*16*), and risk redistribution (*17*). To address these problems, statistical scientists adapted methods from correlation analysis (*18*) and developed a technique known as back-calculation (*19*).

Back-calculation uses the number of AIDS cases diagnosed per month or calendar quarter (which can be estimated from AIDS surveillance data) and the probability distribution of the incubation period (the time from HIV infection to diagnosis of AIDS) to estimate the number of persons infected with HIV. This incubation distribution must be estimated from cohort studies. On the basis of these data, back-calculation methods provide estimates of the number of persons infected with HIV during each month or calendar quarter necessary to account for the number of persons in whom AIDS has been diagnosed during those same periods. The number of persons in whom AIDS will be diagnosed in the future can then be projected from the estimated HIV epidemic curve and the incubation period distribution (*20*).

The back calculation method proved useful in navigating two major changes in the way HIV/AIDS surveillance was conducted. One was a 1993 change in the surveillance case definition for AIDS to include all HIV-infected persons who have <200 CD4+ T-lymphocytes/μL or a CD4+ T-lymphocyte percentage of total lymphocytes <14, or in whom pulmonary tuberculosis, invasive cervical cancer, or recurrent pneumonia has been diagnosed (*21*). The other was the development and widespread use of pharmacotherapy (zidovudine) (*22*). These and other statistical challenges in HIV/AIDS surveillance illustrated well the ability of statistical methods to respond to developing public health problems.

During the mid-1980s, with the increasing availability of microcomputers, CDC epidemiologists first began using computers during field investigations, but no user-friendly software existed for the purpose. To remedy this problem, in the early 1980s, CDC began development of Epi Info, a general-purpose computer program that could be used for epidemic investigations and surveillance (Table). Early versions of Epi Info were used in field investigations on large "luggable" computers (*23*) (Figure 3). The widespread distribution of Epi Info and the responsiveness of its developers to the needs of epidemiologists in the field drove the application of statistical methods in field investigations throughout the world (*24*). A recent search of MEDLINE found >23,000 citations mentioning Epi Info in the peer-reviewed literature. Add to this countless other citations in reports not indexed, and the impact of its development on the field of statistics is apparent. In addition, Epi Info aided in early efforts to coordinate surveillance activities to reduce the workload of state health departments (*25*).

During this period, statistical methods for surveillance also advanced. The availability of methods of forecasting by using time series methods augmented previous regression results (*26,27*). An investigation in response to food poisoning in Peru was the first documented field investigation to implement a time series analysis (CDC, unpublished data, 1986). Use of these methods, developed during the 1920s, was aided by the availability of computers that allowed computations to be conducted in a reasonable amount of time.

More broadly, methods were developed to investigate changes in patterns of surveillance data to aid in epidemic detection and control (*28*). This development was further aided in 1987, when the National Center for Health Statistics became part of CDC and brought its expertise in vital statistics and surveys (*29*).

The 1990s

Innovations continued during the 1990s in such areas as the detection of statistical aberrations, and changes in patterns of data reported over time (*30--33*). A 1988 Symposium on Statistics in Surveillance (*34*) became the foundation for ongoing CDC symposia on the statistics of cluster investigations (*35*), statistics for rare events and small areas (*36*), statistics as a basis for public health decisions (*37*), emerging statistical issues (*38*), complicated designs and data structures (*39*), methods for decisions in uncertainty (*40*), methods for addressing health inequities (*41*), and use of multisource data (*42*). Over time, these symposia were accompanied by short courses to educate the public health community about statistical methods (*43*). In addition, CDC began giving awards for outstanding statistical work that had public health impact (Figure 4).

Despite considerable achievements in reducing smoking prevalence as the 20th century closed, tobacco use remained responsible for one of every five U.S. deaths. In 1999, CDC's Office on Smoking and Health created the National Tobacco Control Program to encourage coordinated efforts to reduce tobacco-related diseases and deaths (*44*). The National Youth Tobacco Survey measured the tobacco-related beliefs, attitudes, and behaviors of youth and was the first to gather data from both high school and middle school students. Findings were used to design strategies for youth-focused antitobacco campaigns (*45*). In 1994, economic methods were used to measure smoking-attributable costs (*46*).

In 1992, Anderson and May published *Infectious Disease of Humans* (*47*), documenting their work in mathematical modeling transmission of infectious diseases, which was critically important to understanding the ongoing work in fighting the global HIV epidemic, as well as malaria and tuberculosis. Subsequent work on modeling diseases has been used to monitor and model the impact of influenza outbreaks. During the 1990s, laboratory techniques improved enough so that strains of viruses could be mapped and links made to the epidemiologic investigation.

The 2000s

Although today the consequences of unhealthy dietary choices, sedentary lifestyles, and "supersized" food portions are familiar, during the late 1990s, their potential for harm was underestimated. Research published in 1999 documented the nation's rapidly increasing obesity rates in all U.S. states, regions, and demographic groups (*48*). In 2001, Congress appropriated $125 million for CDC to develop a national media campaign to change children's health behaviors. CDC responded through VERB, an innovative and expansive campaign based on behavioral science theory and contemporary principles of marketing, which produced measurable positive results (*49*). Once again, CDC epidemiologists were using statistical analytic methods that had previously been used in other disciplines. For example, Bayesian methods used by businesses and marketers to model personal and community decision making preferences (*50*) or cluster analysis and marketing segmentation methods were being used to inform health intervention and evaluation of health programs (*51*). Statistical methods in longitudinal analysis and mixed models used commonly in social research also contributed to the evaluation of results (*52*). Likewise, a method developed in 1896 for studies in biological sciences, capture-recapture analysis, was adapted for evaluating surveillance systems (*53,54*). This method facilitated the estimation of total number of cases from two surveillance sources, each of which might not be complete.

In response to the terrorism events of 2001, statisticians began to develop methods for use in defense and national security (*55*). The rise of spatial statistics and geographic information systems meant that epidemiologists could better map prevalence data to suggest gaps in response or impact of disease or injury (*56*). Economic data could be mapped for use in cost-effectiveness studies, and overlaying data types (prevalence, economic costs, demographics) could be used for better decision making and for evaluation of programs. Mapping the cholera outbreak in John Snow's time seemed to have come full circle.

Many of the techniques of spatial analysis depend on statistical measures and methods, including univariate statistical measures and directional analysis (*57*). Additionally, statistical methods have been developed to address the specific needs of spatial datasets. The nature of these extensions differs from the ways in which multivariate statistics are derived from their univariate counterparts because of concepts of distance, direction, contiguity, and scale. For example, classical hypothesis testing and inferential procedures might not be appropriate for spatial problems because the datasets do not satisfy classical independence or distributional requirements or because the sampling frame may be unknown or poorly specified.

The Future of Statistics

In the future, epidemiologists will continue to pursue new statistical techniques that can increase the impact of their analyses on public health. For example, the coming decades might bring innovations in new data collection modalities (e.g., hand-held data collection methods, cellular phones) and methods needed to evaluate new public health and medical interventions, and they will all be packed into a shrinking global village. A large body of methods (e.g., canonical correlations, factor analyses, exposure assessment, nonparametric statistics, infectious disease modeling) can be brought to bear on new public health problems. However, the use of these new technologies also comes with challenges.

For example, the introduction of parallel sequencing technologies (*58*) has led to an exponential increase in the amount of available DNA sequence information for epidemiologic investigations. Because sequence data are now produced faster than they can be meaningfully analyzed, new approaches to the analysis of this information is one of the most important recent challenges for epidemiologists, bioinformaticians, and statisticians. Beyond methods to carefully sample and organize the massive amount of data, challenges include development of quantitative methods and models to estimate errors for the various sequencing platforms; algorithms and mathematical estimates of the reliability of genomes assembled from short-gapped reads; approaches to distinguish sequence-determination errors from biological polymorphism and mutation; and means to distinguish among multiple genomes within a single dataset, particularly when the relative sizes of those different genomes vastly differ.

Challenges especially relevant to the area of biodetection include development of models for rapid identification of the differences between the genomes of individuals of a species and for distinguishing between naturally occurring biological heterogeneity and newly emerged or artificially produced pathogenic sequences in complex samples. Mathematical models and methods to estimate the significance of genomic variability currently exist, and the use of these models and methods will increase as they become easier to use. Nanotechnology, the understanding and control of matter at dimensions of roughly 1--100 nanometers (10--9 meter), where unique phenomena enable novel applications, presents specific challenges to statistical methods: in understanding high variation in experimental results, in developing sampling plans to model the nanofabrication process efficiently, and in helping to improve low-quality and unpredictable product reliability. As they have during the past 50 years, in the coming decades statistical methods will play a major role in strengthening the evidence base for decisions affecting the well-being of communities.

References

1. Anonymous. International Congress of Medicine. Med Officer 1913;10:104.
2. Satcher D. Public Health Service: on the job for 200 years. Public Health Rep 1998;113:201-- 3.

3. Farr W. Report on the mortality from cholera in England, 1848--1849, London: Her Majesty's Stationery Office; 1852 [Also published as Farr W. Registrar General's report on cholera in England 1849--1850. London: W. Clowes & Son; 1852].
4. Susser M. Epidemiology in the United States after World War II: the evolution of technique. Epidemiol Rev 1985;7:147--77.
5. Cohen IB. Florence Nightingale. Sci Am 1984;3:128--37.
6. Thacker SB, Stroup DF, Sencer DJ. Epidemic assistance by the Centers for Disease Control and Prevention: role of the Epidemic Intelligence Service. Am J Epidemiol. In press, 2011.
7. Langmuir AD. The surveillance of diseases of national importance. The challenge of epidemiology: issues and selected readings. Washington, DC: Pan American Health Organization;13:855--67.
8. Pirkle JL, Brody DJ, Gunter EW, et al. The decline in blood lead levels in the United States. JAMA 1994;272:284--91.
9. Fox JP, Elveback L, Scott W, et al. Herd immunity: basic concept and relevance to public health immunization practices. Am J Epidemiol 1971;94:179--89.
10. Sencer DJ, Dull HB, Langmuir AD. Epidemiologic basis for eradication of measles in 1967. Public Health Rep 1967;82:253--6.
11. Sencer DJ, Millar JD. Reflections on the 1976 swine flu vaccination program. Emerg Infect Dis 2006;12:29--33.
12. CDC. Typhoid fever---Michigan. MMWR 1982;31:544, 549--50.
13. Wipfli HL, Fujimoto K, Valente TW. Global tobacco control diffusion: the case of the framework convention on tobacco control. Am J Public Health 2010;100:1260--6.
14. CDC. A cluster of Kaposi's sarcoma and *Pneumocystis carinii* pneumonia among homosexual male residents of Los Angeles and Orange Counties, California. MMWR 1982;31:305--7.
15. Jaffe HW, Choi K, Thomas PA, et al. National case--control study of Kaposi's sarcoma and *Pneumocystis carinii* pneumonia in homosexual men: part 1, epidemiologic results. Ann Intern Med 1983;99:145--51.
16. Jager JC, Heisterkamp SH, Brookmeyer R. AIDS surveillance and prediction of the HIV and AIDS epidemic: methodological developments. AIDS 1993;7(Suppl 1):S67--72.
17. Song R, Hall HI, Frey R. Uncertainties associated with incidence estimates of HIV/AIDS diagnoses adjusted for reporting delay and risk redistribution. Stat Med 2005;24:453--64.
18. Morgan WM, Curran JW. Acquired immunodeficiency syndrome: current and future trends. Public Health Rep 1986;101:459--65.
19. Gail MH, Brookmeyer R. Methods for projecting course of acquired immunodeficiency syndrome epidemic. J Natl Cancer Inst 1988;80:900--11.
20. CDC. HIV prevalence estimates and AIDS case projections for the United States: report based upon a workshop. MMWR 1990;39(No. RR-16).
21. CDC. 1993 Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR 1992;41(No. RR-17).
22. Rosenberg PS, Gail MH, Schrager LK, et al. National AIDS incidence trends and the extent of zidovudine therapy in selected demographic and transmission groups. J Acquir Immune Defic Syndr 1991;4:392--401
23. 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;7:178--82.
24. Nieves E, Jones J. Epi Info(tm): now an open-source application that continues a long and productive "life" through CDC support and funding. Pan African Medical Journal 2009;2:6. Available at http://www.panafrican-med-journal.com/content/feature/2/6/ .
25. CDC. National Electronic Telecommunications System for Surveillance---United States, 1990--1991. MMWR 1991;40;502--3.
26. Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962--1979. I. Time series

forecasts of expected pneumonia and influenza deaths. Am J Epidemiol 1981;113:215--26.

27. Stroup DF, Thacker SB, Herndon JL. Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age, 1962--1983. Stat Med 1988;7:1045--59.
28. Stroup DF, Williamson GD, Herndon JL, Karon JM. Detection of aberrations in the occurrence of notifiable diseases surveillance data. Stat Med 1989;8:323--9.
29. CDC. Notice to readers: NCHS National Center for Health Statistics joins CDC. MMWR 1987:36;390.
30. Stroup DF, Thacker SB. A Bayesian approach to the detection of aberrations in public health surveillance data. Am J Epidemiol 1993;4:435--43.
31. Nobre FF, Stroup DF. A monitoring system to detect pattern changes in public health surveillance data. Int J Epidemiol 1994;23:408--18.
32. Simonsen L, Clark M, Stroup DF, et al. A method for timely reporting of influenza-associated mortality in the United States. Epidemiology 1997;8:390--5.
33. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80:i89--96. Available at http://www.syndromic.org/syndromicconference/2002/Supplementpdf/Hutwagner.pdf .
34. Symposium on statistics in surveillance. Stat Med 1989;8:251--400.
35. CDC. Guidelines for investigating clusters of health events. MMWR 1990;39(No. RR-11):1--16.
36. Smith SJ, ed. Symposium on small area statistics in public health: design, analysis, graphic and spatial methods. Stat Med 1996;15:1907--86.
37. Falter KH, Betts DR, Rolka DB, Rolka HR, Sieber WK, eds. Symposium on Statistical Bases for Public Health Decision Making: from exploration to modelling. Stat Med 1999;18:3159--375.
38. Sieber WK, Green TA, Haugh GS, et al., eds. Symposium on emerging statistical issues in public health for the 21st century, Stat Med 2001;20:1307--561.
39. Lin LS, Conn JM, Green TA, Johnson CH, Odencrantz JR, Sieber WK Jr, eds. 8th Biennial CDC and ADSTR Symposium on Statistical Methods: issues associated with complicated designs and data structures. Stat Med 2003;22:1359--626.
40. Lipman H, Cadwell BL, Kerkering JC, Lin LS, Sieber WK, eds. Study design and decision making in public health. Proceedings of the 9th Biennial U.S. Centers for Disease Control/Agency for Toxic Substances and Disease Registry (CDC/ATSDR) Symposium on Statistical Methods. January 27---29, 2003. Atlanta, Georgia, USA. Stat Med 2005;24:491--669.
41. Davis KE, O'Connor KS, eds. Statistical issues in addressing health inequalities. Stat Med 2008;27:3925--4144.
42. Utilization of multi-source data: InfoFusion. Available at http://www.cdc.gov/sag.
43. Sieber WK, Green T, Williamson GD. Statistics and public health at CDC. MMWR 2006;55(Supl 2):22--4.
44. CDC. Ten great public health achievements---United States, 1900--1999. MMWR 1999;48:241--3.
45. Nelson DE, Kirkendall RS, Lawton RL, Chrismon JH, et al. Surveillance for smoking-attributable mortality and years of potential life lost, by state---United States, 1990. MMWR 1994;43(SS-1):1--8.
46. Nelson DE, Lawton RL, Chrismon JH, et al. Surveillance for smoking-attributable mortality and years of potential life lost by State -- United States, 1990. MMWR 1994;43(SS1):1-8.
47. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. New York, NY: Oxford University Press; 1992.
48. Mokdad AH, Serdula MK, Dietz WH, Bowman BA, Marks JS, Koplan JS. The spread of the obesity epidemic in the United States, 1991--1998. JAMA 1999;282:1519--22.
49. Huhman ME, Potter LD, Duke JC, Judkins DR, Heitzler CD, Wong FL. Evaluation of a national

physical activity intervention for children: VERB campaign, 2002--2004. Am J Prev Med 2007;32:38--43.
50. Grover R, Vriens M. The handbook of marketing research: uses, misuses, and advances. Thousand Oaks, CA: Sage Publications; 2006.
51. Böcker F. Marketing science around the world. Journal of Marketing 1981;45:169--72.
52. Onweugbuzie AJ, Collins KMT. A typology of mixed methods sampling designs in social science research. Qual Rep 2007;12:281--316.
53. Hall HI, Song R, Gerstle JE 3rd, Lee LM; HIV/AIDS Reporting System Evaluation Group. Assessing the completeness of reporting of human immunodeficiency virus diagnoses in 2002-2003: capture--recapture methods. Am J Epidemiol 2006;64:391--7.
54. Southwood TRE, Henderson P. Ecological mortality. 3rd ed. Oxford: UK: Blackwell Science; 2000.
55. Elbert Y, Burkom HS. Development and evaluation of a data-adaptive alerting algotithm for univariate temporal biosurveillance data. Stat Med 2009;28:3226--48.
56. Robinson TP. Spatial statistics and geographical information systems in epidemiology and public health. Adv Parasitol 2000;47:81--128.
57. De Smith MJ, Goodchild MF, Longley PA. Geospatial analysis---a comprehensive guide. 3rd ed. London;Winchelsea Press;2009. Available at http://www.spatialanalysisonline.com/ .
58. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008;299:1335--44.

FIGURE 1. Statistician at CDC using MonroMatic desktop calculator, Model 8N-213. circa 1958

Photo: CDC

Alternate Text: The figure is a photo of a CDC statistician using a MonroMatic desktop calculator in 1958.

FIGURE 2. Computer workstation at CDC, 1980s

Photo: CDC

Alternate Text: The figure is a photo of a CDC employee at a computer workstation in the 1980s.

TABLE. Examples of software systems developed by CDC in the 1980s and 1990s

Sofware system name	Primary use	Reference
IDEAS (Interactive Data Entry and Analysis System)	Support hospitals' participation in CDC's nosocomial infection surveillance activities	Horan TC, White JW, Jarvis WR, et al. Nosocomial infection surveillance, 1984. MMWR 1986;35(No. SS-1).
SAMEC (Smoking-Attributable Mortality and Economic Costs)	Allow states and local areas to estimate the impact of smoking-attributable illness and mortality	CDC. State-specific estimates of smoking-attributable mortality and years of potential life lost---United States, 1985. MMWR 1988;37:689--93.

Software for Congenital Syphilis Surveillance	Assist states in reporting cases of congenital syphilis	Dunn RA, Webster LA, Nakashima AK, Sylvester GC, Surveillance for geographic and secular trends in congenital syphilis---United States, 1983--1991, MMWR 1993;42(No. SS-6).
ARDI (Alcohol-Related Disease Impact)	Estimate the impact of alcohol consumption	CDC. Deaths and hospitalizations from chronic liver disease and cirrhosis---United States, 1980--1989. MMWR 1993;41:969--73. CDC. Alcohol-Related Disease Impact (ARDI). Available at http://apps.nccd.cdc.gov/ardi/homepage.aspx.
SURVTB	Support state health departments in TB case surveillance and prevention	CDC. Expanded tuberculosis surveillance and tuberculosis morbidity---United States, 1993. MMWR 1994;43:361--6.
STELLAR (Systematic Tracking of Elevated Lead Levels & Remediation)	Support state activities in prevention of elevated blood lead levels	CDC. State activities for prevention of lead. MMWR 1993;42:165,171--2.
PHLIS (Public Health Laboratory Surveillance System)	Support reporting from state public health laboratories	Bean NH, Martin SM, Bradford H, Jr. PHLIS: an electronic system for reporting public health data from remote sites. Am J Public Health 1992;82:1273--6.
Epi Info	Support data collection and analysis from field investigations; to support state surveillance activities	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;7:178--82.

FIGURE 3. "Luggable" Osborne computer, circa 1981

Photo: CDC

Alternate Text: The figure is a photo of a "Luggable" Osborne computer in 1982.

FIGURE 4. CDC's Statistical Achievement Ceremony 1993: Award for statistical methods to Investigation of 2,3,7,8-tetrachorodibeno-*p*-dioxin half-life heterogeneity in Veterans of Operation Ranch Hand. Claire V. Broome (presenter), James Pirkle, Samuel Caudill, and Mitchell Gail (National Institutes of Health)

Photo: CDC

Alternate Text: The figure is a photo at CDC's Statistical Achievement Ceremony in 1993. Pictured are winners of the Award for Statistical Methods to Investigation: Clair V. Broome (presenter), James Pirkle, Samuel Caudill, and Mitchell Gail.

Page last reviewed: October 07, 2011
Page last updated: October 07, 2011
Content source: Centers for Disease Control and Prevention