Commentary

Epidemiology: Then and Now

Lewis H. Kuller*

* Correspondence to Dr. Lewis H. Kuller, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, 130 North Bellefield Avenue, Room 550, Pittsburgh, PA 15261 (e-mail: kullerl@edc.pitt.edu).

Initially submitted April 1, 2015; accepted for publication April 24, 2015.

Twenty-five years ago, on the 75th anniversary of the Johns Hopkins Bloomberg School of Public Health, I noted that epidemiologic research was moving away from the traditional approaches used to investigate “epidemics” and their close relationship with preventive medicine. Twenty-five years later, the role of epidemiology as an important contribution to human population research, preventive medicine, and public health is under substantial pressure because of the emphasis on “big data,” phenomenology, and personalized medical therapies. Epidemiology is the study of epidemics. The primary role of epidemiology is to identify the epidemics and parameters of interest of host, agent, and environment and to generate and test hypotheses in search of causal pathways. Almost all diseases have a specific distribution in relation to time, place, and person and specific “causes” with high effect sizes. Epidemiology then uses such information to develop interventions and test (through clinical trials and natural experiments) their efficacy and effectiveness. Epidemiology is dependent on new technologies to evaluate improved measurements of host (genomics), epigenetics, identification of agents (metabolomics, proteomics), new technology to evaluate both physical and social environment, and modern methods of data collection. Epidemiology does poorly in studying anything other than epidemics and collections of numerators and denominators without specific hypotheses even with improved statistical methodologies.

common source; epidemics; incubation period; person-to-person transmission

In 1991, as part of the 75th anniversary of the Johns Hopkins Bloomberg School of Public Health, I published a critique in the American Journal of Epidemiology entitled, “Epidemiology Is the Study of ‘Epidemics’ and Their Prevention” (1). I noted that in the 1960s, the Department of Epidemiology and Chronic Diseases at Johns Hopkins focused on the following ideas: 1) the study of epidemics in relation to epidemiology is the study of epidemics and their prevention; 2) epidemiology has a strong biological basis; 3) epidemiology is one of the basic sciences of public health and preventive medicine; 4) epidemiology training is much broader, and students are trained in both infectious and noninfectious disease epidemiology; and 5) the important role of clinical trials was firmly established.

Epidemiologic studies of chronic diseases and diseases with long incubation periods were drifting away from the basic concepts of epidemiology and the links with both public health and preventive medicine. In 2015, the situation has gotten much worse. Will long-incubation-period epidemiology or noninfectious-disease epidemiology survive as a discipline? Epidemiology is now defined as the collection of large sample sizes and measurement of numerous variables from stored samples to facilitate estimation of disease risk over time (2–4). Furthermore, new techniques to acquire dependent and independent variables for epidemiologic studies have become the cornerstones of epidemiology research (5, 6). These large samples and new technologies for measurement without specific hypotheses are likely to identify statistically significant irrelevant observations. An important issue is therefore how to get epidemiology “back on track.”

In this paper, I attempt to place the current incarnation of long-incubation-period epidemiology in the context of the traditional disciplines of host, agent, and environment; incubation periods; and modes of transmission of disease. These are the cornerstones of research and teaching of epidemiology at Johns Hopkins University. Epidemiology’s primary role was to identify the best approaches to advance public health.


The Department of Chronic Diseases included multiple disciplines, such as sociology, economics, psychology, and
Epidemiology is the study of epidemics

Epidemiology is defined as “the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems” (17, p. 3). I have previously expanded the definition (18). First, epidemiology is the study of epidemics in populations (18, 19). It does not study individual risk (i.e., “personalized medicine”). Second, epidemiology separates populations within epidemics into smaller and smaller groups at increasing risk of disease, for example, through the development of the Framingham Coronary Heart Disease Risk Score (20) or when using the extent of coronary artery calcification (21) to predict coronary heart disease.

An epidemic can be defined in several ways. These definitions include 1) the differences in disease incidence or mortality among different populations (22); 2) an unusual distribution of a disease within a population, such as dementia or Alzheimer disease in young people (23); 3) an unusual distribution within or between populations of physical and biochemical characteristics, such as very low or high levels of apolipoprotein B lipoproteins due to mutations of the proprotein convertase subtilisin/kexin type 9 gene (PCS K9) (24, 25); 4) a change in the incidence of a disease over time, for example, the decrease in melanoma related to the use of sunscreen (26–28), or a decrease in a disease, for example, the disappearance of rheumatic fever as noted by Leon Gordis in the T. Duckett Jones Memorial Lecture (29); 7) a higher incidence of a disease or illness in a defined subgroup of individuals within a population that can be defined by ethnic, social, and behavioral characteristics (30), for example, vegetarians with a low risk of CHD or selected cancers (31) or persons with a genetic attribute that substantially increases risk of a disease (e.g., the increased risk of dementia and Alzheimer’s disease for both homozygote and heterozygote apolipoprotein E allele carriers) (32–34); and 8) an unusual distribution of exposure to an agent known to be related to a cause of disease, for example, high levels of arsenic exposure in populations and increased risk of diseases such as lung cancer (35, 36).

The identification and characterization of epidemics are the first and most important steps in epidemiologic investigations. Conversely, epidemiology is a very weak tool for studying events other than epidemics (e.g., collecting samples from individuals without considering a specific epidemic and trying to identify risk factors for disease) (18).

The bigger the differences in disease distributions within a population, the more likely that epidemiology will be successful in identifying the specific interrelationship among the agent, host (genomics), and environment (37, 38). If there is an identified epidemic, there is likely to be a causal pathway and a specific etiological agent(s) (39). It is only the lack of skills, technology, and resources that prevent identification of the etiological agent(s) and potential methods of prevention.
Investigations of epidemics are dependent on the evolution of new technologies for the study of human biology, including measurement of host (genomics) (40, 41), agents of disease such as proteomics (42, 43), and metabolomics (44–47), and for the measurement of exposure (48–50). However, the use of proper epidemiologic methods to test hypotheses, not the new technologies, is primary (46). Successful epidemiologic studies often depend on close collaboration with both basic and clinical scientists. Successful intervention and control of a disease does not solely depend on untangling the specific molecular biology and pathophysiology of a specific disease, although it is extremely important in the development of new preventive and therapeutic approaches (51–54).

Epidemiologic studies cannot explain every case of a specific disease within an epidemic. As George Davey Smith noted, there will be cigarette smokers who will survive to old age without getting lung cancer or coronary artery disease because of their unique genetic characteristics or, more likely, random chance (55). Epidemiologists chase a lot of “nonepidemics,” using newer complex statistical methods to search large data sets for significant associations of dependent and independent variables, which leads to many uninterpretable *P* values (56).

**IDENTIFYING THE CASUAL AGENTS**

Successful identification of the “agent(s)” in the causal pathway of disease is the key to epidemiologic research. Such agents will have very high relative and attributable risks. This is true both for studies of chronic disease and for infectious-disease epidemiology (e.g., studies of the association between cigarette smoking and lung cancer (57) and studies of the relation among blood cholesterol levels, blood pressure, and the risk of coronary heart disease (58)). Multiple causal agents with a high relative or attributable risk can be identified for the same disease (58, 59), such as coronary heart disease, likely because there is an association with stages of the pathophysiology of the disease (e.g., an association between heart disease and atherosclerosis, plaque changes, or thrombosis) or because there are highly correlated measurements of a common variables (59, 60).

**TIME, PLACE, AND PERSON**

All diseases have a specific distribution in the population in relation to time, place, and person. If there is no unique distribution of disease in relation to place, time, or person, it is likely that the definition of the disease consists of major subgroups of symptoms or diseases that have specific distributions in relation to time, place, and person. For example, in the 1920s and 1930s, heart disease consisted of rheumatic heart disease, syphilitic heart disease, coronary artery disease, and potentially other infectious or nutrition-related heart diseases, each with a specific distribution in relation to time, place, and person. As the incidence of rheumatic heart disease declined and that of coronary artery disease increased, the trends in heart disease among men and women began to change, with a substantial increase in coronary heart disease noted among men beginning in the 1920s and 1930s but only later among women, especially as they adopted cigarette smoking (61).

New technologies can improve the classification of disease and lead to subclassifications with unique epidemiology. Examples include the recent recognition and identification of the presence of anticyclic citrullinated peptide antibody (62, 63) in patients with diagnosed rheumatoid arthritis. Approximately 70% of patients with rheumatoid arthritis have anticyclic citrullinated peptide antibodies. Rheumatoid arthritis is an epidemic among women when compared with men, as are most autoimmune diseases. The epidemiology of rheumatoid arthritis that is positive for anticyclic citrullinated peptide antibody appears to be very different from that for rheumatoid arthritis that is negative for anticyclic citrullinated peptide antibody in relation to host (genetics), human leukocyte antigen shared epitope, cigarette smoking, and elevated levels of inflammatory cytokines (64). The identification of amyloid plaque in the brain using positron emission tomography imaging in studies of the association between heart disease and atherosclerosis, plaque changes, or thrombosis or because there are highly correlated measurements of a common variables (59, 60).

**INCUBATION PERIOD**

All diseases have a specific incubation period, which is the time from initial exposure to the onset of clinical disease. Philip Sartwell, who was the Chairman of the Department of Epidemiology in the 1960s, noted that the incubation period was log-normally distributed and that the antilog of the log (standard deviation), which he called the “dispersion factor,” was a measure of the variability of time around the median incubation period (70, 71). Armenian and Lilienfeld (72) evaluated the incubation periods of neoplastic diseases for which the time of the initial exposure could be measured, such as in cancer related to occupational hazards, therapeutic exposure to radiation, etc. Recently, incubation period was evaluated in relation to radiation exposure after the atomic bomb (73) and to the risk of acquired immunodeficiency syndrome among patients with hemophilia (74). Armenian and Lilienfeld reported that the incubation period was log-normally distributed. The incubation period varied from 4.6 years for intrauterine exposure to radiation and acute leukemia to 36 years for occupational exposure to asbestos and the risk of lung cancer.

Recently, there has been less effort to study incubation periods in noninfectious diseases with long incubation periods primarily because the likely agents are continuous, such as specific dietary factors and persistent environmental agents, rather than single exposures of disease. It is important to have an estimation of the incubation period when trying to determine whether the introduction of new agents into the population is contributing to the changes in incidence or distribution of disease, new risk factors, etc. In infectious diseases with short incubation periods (e.g., the recent Ebola epidemic (75) and severe acute respiratory syndrome (76)),
such associations are much easier to identify. In the Pittsburgh Youth Study, Dr. Rolf Loeb et al. (77) determined the incubation period for adverse behavioral outcomes in adolescence and young-adult life that were tracked to childhood experiences and that led to homicide. Similarly, lifespan studies can be used to study whether early-life brain injuries secondary to poor nutrition, infection, trauma, and environmental exposures are in the causal pathway for chronic diseases with later onset and long incubation periods (78).

A recent interesting approach to studying the incubation period of Alzheimer disease was performed by the Dominantly Inherited Alzheimer Network; 88 carriers with a mean age of 39 years were evaluated (79). Age at onset of Alzheimer disease in the index cases was used to estimate the relationship between the disease biomarkers and the onset of clinical disease, that is, the incubation period. If the “epidemic” increase of autism is real, as many think, then the cause must be an exposure to a specific environmental agent, probably very early in life and perhaps before or during pregnancy, and studies in the postpartum period may be unsuccessful in identifying the agent(s) (80).

Misinformation about the incubation period can lead to spurious interpretation of the results of epidemiologic studies. For example, levels of low-density lipoprotein cholesterol, apolipoprotein B, or low-density lipoprotein particles are weak predictors of the risk of coronary heart disease among older individuals (81–83), most of whom already have extensive atherosclerosis and are past the incubation period. Some investigators, unfortunately, have concluded that lipid levels are not important in older individuals (84). The development of atherosclerosis begins in childhood (85).

New epidemics of diseases (both those with short incubation periods and those with long incubation periods) in the United States are likely to begin in persons who are better educated and in the upper levels of socioeconomic status because those people are more likely to travel outside of the United States and introduce into the population new toxic chemicals, or novel lifestyles. They are also more likely to try new expensive drug therapies that have an adverse effect, which causes what is known as iatrogenic disease. New epidemics that began in persons with higher socioeconomic statuses or educational levels might move fairly rapidly to those with lower socioeconomic statuses and less education. The rate at which the epidemic spreads depends on the incubation period, pathogenicity, and virulence of agents. The determinants and modeling of the spread of the epidemic remain very important components of current epidemiologic research, especially as it is applied to preventive medicine and public health. A high priority of epidemiologic research must therefore be vigilant surveillance of changes in the incidence or morbidity of a disease, subclinical disease, or predisease markers in the population, which can be clues to new epidemics. Such studies require a representative sample(s) and good surveillance methodology. The new approaches to measuring the incidence of diseases in large populations include the use of health maintenance organization networks (88).

MODE OF TRANSMISSION OF DISEASE

The mode of transmission of the agent within person-to-person, vector-borne, and common-source epidemics determines the research methodology. There have been few recent epidemiologic studies in which investigators described the mode of transmission of the agent(s) in the population. Helen Abbey, Professor of Biostatistics at Johns Hopkins University, used novel approaches, including playing with marbles, to teach students the basis of Reed-Frost methods for evaluating epidemics with person-to-person transmission (89). Such techniques are now being applied to investigations of diseases with long incubation periods (90).

Cervical cancer is a venereal disease caused by infection with specific strains of human papilloma virus (91), and the risk of liver cancer is higher in people with hepatitis B transmitted vertically (e.g., from mother to child) (92). Both are examples of diseases with person-to-person transmission and long incubation periods. Studying the possible person-to-person transmission of chronic diseases is challenging but very important for etiological research and for the creation of better preventive strategies. The agents for possible person-to-person transmission of a disease such as multiple sclerosis (93) might not be identifiable despite extensive epidemiologic studies, likely because of an absence of the technology necessary to identify the agent(s).

Examples of common-source epidemics include those resulting from exposure to environmental toxins at the workplace, such as cancers caused by asbestos; the increased risk of asthma attack and hepatitis A due to air and water pollution; the higher incidence of CHD and rising prevalence of obesity associated with higher intakes saturated fat and increased total kilocalories; and lung and other cancers caused by cigarette smoking (94–98). The study of common-source epidemic, such as the hepatitis A epidemic associated with a contaminated water supply in Washington County, Maryland, was a key component of basic epidemiology training at Johns Hopkins University in the 1960s.

The obesity epidemic has been the most recent important common-source epidemic. The obesity epidemic is due to an increase in caloric intake and a decrease in energy expenditure. Unfortunately, many epidemiologists, clinicians, and public health practitioners mistook this common-source epidemic for a person-to-person transmission epidemic, which led to little success in controlling it (99, 100). It is also possible that changes in specific types of foods or methods of preparation could have affected gut microflora, which in turn could have increased the individual risk of obesity (i.e., increased host susceptibility) (101).

In a common-source epidemic, the incidence of the disease is initially higher among those who were exposed than among those who were not. For example, consider a food-born epidemic with a short incubation period caused by ingesting toxic *Escherichia coli* and resulting in diarrhea. The incidence would be higher in those who had consumed the contaminated food than in those who did not. There is an important host susceptibility or genetic component. It is often not difficult to determine who was exposed and who was not. Successfully measuring the dose of the exposure among individuals who have been exposed to the agent is often far more difficult,
especially in epidemiologic studies of diseases with long incubation periods in relatively homogenous human populations. Each individual’s dose of exposure is a function of the within-individual variability over time rather than of the between-individual exposures (102, 103). New technologies in fields such as metabolomics, proteomics, and epigenetics might greatly enhance our ability to measure within-versus between-individual exposures to a common-source agent by providing biomarkers of specific exposures (41, 95, 104–106).

HOST, AGENT, AND ENVIRONMENT

All epidemics are defined by the interaction of the host, the agent, and the environment. The measurement of host susceptibility (genomics) has had the biggest recent impact on epidemiologic studies. The Department of Chronic Diseases and later the Department of Epidemiology at Johns Hopkins University was one of the first to include a program on genetic epidemiology; it was directed by Dr. Bernice Cohen. The Moore Clinic began to change its focus in the 1960s under the direction of Dr. Victor McKusick, and it became a world leader in clinical genetics and later in population genetics. There was a close link between the Moore Clinic and the School of Public Health at Johns Hopkins. Each year in the epidemiology course, Dr. McKusick would draw a line across the blackboard. He would note that at one extreme end was the small percentage of the population affected by pure Mendelian genetic disorders and at the other extreme were the relatively few environmental diseases for which there was little or no genetic component. In between was the interaction between genes (host susceptibility) and the environment.

Unfortunately, the reductionist model for the study of the genome and disease failed to recognize the importance of the agent and environmental exposures, as well as the critical need to define phenotypes in epidemiologic studies (107–111). The genome is the scaffolding for the responses to specific agents and environmental exposures (112).

Better measures of exposure to the agents in the physical and social environments are critical to understanding the interrelationship among the specific genetic attributes, pathophysiology, and disease. Better evaluation of exposures to agents of disease and determinants of the phenotype, as well as newer, lower-cost genomic analyses, will more likely lead to advances in genetic epidemiology (113–117). Again, these studies will greatly benefit from new technologies and better and lower-cost data-collection methods for the investigation of epidemics.

EPIDEMIOLOGY, PUBLIC HEALTH, AND PREVENTIVE MEDICINE

Epidemiology is the basic science of public health and preventive medicine (118). The focus of epidemiology is on reducing disease in the population by identifying the causal pathways and best approaches for prevention. Many of the epidemiology faculty at Johns Hopkins University in the 1960s had prior experience in public health agencies. George Comstock, Professor of Epidemiology, had been a leader in clinical trials for the prevention and early treatment of tuberculosis (119, 120). He was a strong believer in the key role of epidemiology in policy and in public health (119), and he noted the demise of the close links of epidemiology with public health, especially in US studies of chronic diseases and diseases with long incubation periods. Comstock stated that if the epidemiology was not closely linked with policy, it was not worth doing (119).

Lilienfeld was actively involved in the evaluation of the associations of cigarette smoking and radiation exposure with cancer (121). The Commission on Chronic Illnesses in Baltimore in the 1950s, of which he had been a part, had led one of the original efforts to evaluate screening and secondary prevention measures (122). Lilienfeld was also actively involved in the evaluation of screening techniques, especially for lung cancer. These early approaches ultimately led to the successful trials of computerized tomography screening to identify early lung cancer and to reduce mortality (123). Leon Gordis, who later became the chair of the Department of Epidemiology, played a decisive role with Dr. Milton Markowitz in studies of penicillin prophylaxis to reduce the risk of recurrent rheumatic fever among children in the Baltimore city school system (124), which became a model for school-based prevention programs.

The movement of epidemiology away from public health has continued and become exaggerated in recent years, especially for the study of chronic diseases with long incubation periods. Recently, Sandro Galea, Chair of Epidemiology at Columbia University, wrote about the importance of consequential epidemiology and its relevance to public health (125, 126).

Almost all diseases have a specific cause or causes with high attributable risk. Rarely can a disease be substantially reduced in a population without controlling those causal agents. The key roles of epidemiology are not only identification of the determinants of disease but also careful evaluation of the interventions to reduce the extent of disease. Doing good is not necessarily consistent with achieving beneficial results. I refer to this phenomenon as “Popeye Epidemiology”: performing repeated studies to support a belief (e.g., that eating spinach will solve all of one’s problems) that is not based on science (127–129).

In summary, the key steps in the future of epidemiology are as follows: 1) study epidemics and be innovative in your hypotheses, study designs, and most importantly, populations of interest to find those that are most likely to help solve the puzzles (130, 131); 2) have specific hypotheses related to likely agents, vectors, modes of transmission, incubation periods, and host susceptibility (genomics); 3) utilize new technologies to define the disease or pathology (132) of interest and work closely with clinical and basic science investigators to improve measures of phenotype and of identification of causal agent(s) and physical and social environments; 4) use high-quality measures of the genome to determine host susceptibility (133); 5) use new methods to evaluate and monitor epidemics, including both data collection and measurement of outcomes within the context of studying epidemics; 6) evaluate epidemic models for the best approaches to prevention; and 7) avoid repetitive studies of the same associations (i.e., circular epidemiology) (4).

The success of epidemiology should be judged as it was in the past, by its relevance to public health and preventive medicine, and not by the number of P values <0.05 in a paper or
the number of publications or citations (134). The prevention of a specific disease in a population depends in great part on epidemiologic studies to identify the agents and host and to elucidate unique environmental interactions that determine the transmission of the disease within a population. Don’t worry, there is no lack of epidemics to be found, and the causal pathway and methods of prevention can be identified by smart epidemiologists (135–137). The high-hanging “fruit” just requires a slightly longer ladder, not a wider one that likely captures “bad apples.” If epidemiology’s primary role becomes large data collections or repeated studies of small risks, many of which might yield false associations (138–140), in the hope of identifying causal associations or effectiveness of clinical therapies, the field is doomed and opportunities to improve the health of populations will be lost.

ACKNOWLEDGMENTS

Author affiliation: Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Lewis H. Kuller).

Conflict of interest: none declared.

REFERENCES


