Bundesgesundheitsbl - Gesundheitsforsch -
Gesundheitsschutz 2006 · 49:1179–1184
DOI 10.1007/s00103-006-0083-8
© Springer Medizin Verlag 2006

Eric J. Duell

Norris Cotton Cancer Center, Section of Biostatistics & Epidemiology, Center for the Evaluative Clinical Sciences, Department of Community & Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire, USA

Erratum

Bedauerlicherweise ist beim Umbruch dieses Beitrages in der Juli-Ausgabe des Blattes (DOI 10.1007/s00103-006-1293-9) ein technischer Fehler unterlaufen, d.h. die Aufeinanderfolge der Seiten ist dort fehlerhaft. Der vollständige Beitrag ist daher im Folgenden noch einmal abgedruckt.

The future of epidemiology

Methodological challenges and multilevel inference

Epidemiology and public health

A decade ago, there was considerable discussion and debate about the appropriate objectives and paradigms of modern epidemiologic research [1, 2, 3, 4]. These discussions raised several important issues that may have helped a new generation of epidemiologists acknowledge that “risk factor epidemiology” might be forcing our field to focus more on individuals and less on populations and public health. Ten years before these debates, Geoffrey Rose [5] discussed the difference between studying a collection of sick individuals and studying sick populations. He argued that important population-level disease determinants might be missed if we only study a collection of individuals within a population. However, the reverse is also certainly true: that disease determinants at the individual level will be missed if we only compare groups or populations.

Today, most epidemiologists acknowledge that public health is influenced by both population-level and individual-level determinants; however, the arguments raised 10 years ago still hold relevance, especially in this “post-genomics” era. Has the excitement about new statistical methods and molecular technologies enhanced our ability to understand disease and improve public health? Has epidemiology’s contributions led to improved disease prevention, detection, and treatment? Has the field of epidemiology moved even further from the study of populations?

I would argue that whether our data collection activities are conducted at the level of populations or individuals (or both) or whether we are studying interventions or etiology is not what defines our field. What defines our field is the ultimate goal of increasing knowledge toward improving public health, and to effectively communicate research results to the public health community [6]. Whether inferences will be made at the level of individuals or populations or both should be driven by the nature of the exposures and diseases under study. For example, the population prevalence of an exposure as “straightforward” as tobacco smoking is driven by a combination of social, behavioral, and genetic factors. Modern epidemiology can increase knowledge by considering all three factors and their interactions together rather than each factor separately.

In this commentary, I describe how both population-level and individual-level determinants can impact epidemiologic inference (and ultimately, public health); I identify some key methodological issues for new or continued research at the level of populations and individuals, particularly in the context of genetic factors, and finally discuss how we might begin to investigate multiple levels of inference (from genes to individuals to populations) simultaneously in an approach referred to as multilevel epidemiology. My intention is not to reinvent concepts previously proposed [4], but to reinvigorate these debates and inject current context into the argument for multilevel epidemiology.

Population-level inference

Ecologic studies

Ecologic studies can be complex conceptually and analytically, but they also can be very useful and economical. We should not view them as inherently flawed, or view those who conduct them as somehow inferior. Clearly, we must fully appreciate their shortcomings. One of the biggest concerns that is often cited is “ecologic fallacy” or the potential for bias in inferring individual-level or biological effects from group-level factors [7]. What is often missing in ecologic studies are data on the joint distribution of exposures and covariates within comparison groups, and information on the potential contextual (group-level) effects of exposures under study. Despite these warnings, biological inference may not be the goal of all ecologic studies. For example, recent studies have attempted to measure group-level effects of socioeconomic status (SES) on individual-level outcomes (contextual effects) by measuring and controlling for SES at the individual level [8, 9].

If the goal is to describe potential exposures or health events in certain groups or populations, then the ecologic study may be the perfect study design to begin with. An excellent example of an ecologic study that probably has no direct individual-level comparison was recently published by Austin et al. [10]. The researchers used geocoded addresses of restaurants and schools to examine spatial patterns of

food environments around kindergartens and primary and secondary schools in the Chicago area. The authors reported that three to four times as many fast food restaurants were found to cluster within walking distance to schools than would be expected by chance. This descriptive ecologic study provides the community with compelling hypotheses regarding the root causes of obesity in children: that some fast food restaurants target their locations to vulnerable populations, and that economic and marketing forces (group-level or social factors) affect individual-level choices and health. The results of this ecologic study are particularly compelling in light of the expanding obesity and diabetes epidemics among children and adolescents in North America and Europe. My point is that these hypotheses could not have arisen if the authors had chosen to study individual-level dietary factors in schoolchildren as measured using standardized food frequency questionnaires.

Ecologic studies can incorporate both place and time in their design and analysis. Ecologic studies that incorporate spatial relationships between exposures and outcomes (e.g., rates of disease) may fall under the general rubric of environmental epidemiology, disease mapping, or “spatial epidemiology.” While the incorporation of place and time does not eliminate problems of bias in ecologic studies as mentioned above, it can serve to refine or improve definitions of the variables and covariates under study [11]. For a simple example, with regard to mapping incidence of cancer, it is unlikely that smoking prevalence or SES-related factors measured during the past year are as relevant as these factors measured three or more decades in the past. Factors measured or occurring in more recent time periods are more likely to influence disease mortality.

Geographic information systems (GIS)

The use of GIS or computerized systems of spatial data layers (health data, environmental data, economic, political, and meteorological data, etc.) has probably contributed to a recent resurgence of interest in ecologic and disease mapping studies [12, 13]. Many ongoing case-control and

cohort studies could benefit by simply collecting accurate residential street address information or global positioning satellite (GPS) coordinates for all study participants. Handheld GPS units are relatively inexpensive and easy to implement in field work. Lifetime residential and occupational histories also could be obtained using standardized data collection instruments. A novel application of GIS would be to combine individual-level biomarker, genetic and exposure data (from case-control or cohort studies) with multiple spatial data layers using GIS applications. As GIS and spatial data applications in epidemiology and environmental health begin to incorporate more sophisticated statistical modeling (e.g., WinBugs, GeoBugs, MLWin), ecologic studies and environmental epidemiology can only benefit. A concern for studies using GPS longitude and latitude data is that of participant privacy. Producing maps that pinpoint cases, especially in rural areas, could violate participant confidentiality. A few strategies at avoiding this are to provide aggregate data in grids of a large enough scale to protect confidentiality, or to map rates of disease in as fine a geographic unit as possible. In rural areas with low population density and unstable rates, incorporation of spatial or statistical smoothing is recommended [14].

Prognosis, outcomes, health services

These lines of research, especially health services research, often combine group-level data and individual-level data. For example, in health services research in the USA, we may assign “exposure” to the individual based on medical claims data (e.g., receipt of a specific medical test or procedure), and use group-level (zip-code or county-level) data for covariates such as education, income, and behavioral factors (smoking, diet), or even worse, to just ignore these covariates all together.

Health services and disease mapping studies would be greatly improved with the systematic collection of income, education, and behavioral data. Because medical claims data exist for billing and not research, it may take outside forces (such as government regulation) to include

these variables in future medical claims data sets. Perhaps for good reason, individual privacy concerns currently outweigh the needs of the health services research community. Alternatives for health services researchers may include the use of systematic, population-wide surveys of income, education, and behavioral factors outside of the usual census surveys but at a sufficiently fine geographic scale (census blocks or groups) for use in these types of studies. Currently, such systematic and widespread surveys do not exist.

Individual-level inference

Traditional risk factor studies and genomic studies

Traditional risk factor studies at the individual level have been the mainstay of epidemiology during the past 50 years. Many important exposure-disease relations have been identified; one of the most important was tobacco smoking and lung cancer (also heart disease and cancers other than lung). The concept of multifactorial disease (or “etiologic heterogeneity”) arose from the study of multiple factors, usually at the level of the individual. Breast cancer and colorectal cancer are multifactorial diseases, and “risk factor epidemiology” helped to establish this important concept. Probably the most debated individual-level risk factor is the role of diet (as measured using questionnaires) in the etiology of complex diseases such as cancer and cardiovascular disease. Improvements in dietary assessment are needed [15] and further research on biomarkers of dietary constituents such as isothiocyanates and other compounds also would help. In fact, research that supports the discovery and development of biomarkers to improve the measurement of exposures and outcomes is still greatly needed.

Individual-level studies also encompass clinical trials and many intervention studies, as well as genetic epidemiology studies of inheritance patterns in pedigrees and association studies of common genetic variants in the general population. Clearly, all of these lines of inquiry are important and will continue in the near future. With the sequencing of the human genome now complete, genome epidemiology has tak-

Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2006 · 49:1179–1184
DOI 10.1007/s00103-006-0083-8
© Springer Medizin Verlag 2006

Eric J. Duell

The future of epidemiology. Methodological challenges and multilevel inference

Abstract

A decade ago there was considerable debate about the appropriate objectives and paradigms of modern epidemiologic research. One concern put forth in these debates was that “risk factor epidemiology” might be forcing our field to focus more on individuals and less on populations and public health. Today, most epidemiologists acknowledge that public health is influenced by both population-level and individual-level determinants. Ecologic studies are valuable tools for generating hypotheses and addressing group-level determinants of disease risk. Traditional risk factor studies and genomic studies have helped establish the multifactorial concept of disease causation. Individual-level studies also have provided the biomedical community with hy-

potheses that have stimulated research into disease mechanisms that have led to reductions in morbidity and mortality for diseases such as HIV/AIDS, cardiovascular disease, and cancer. Current debates about the role of genomic data in epidemiology and public health mirror the debates about risk factor epidemiology one decade ago. Genomic variation is measured at the individual level, but how this variation is maintained in human populations is a group-level (population) phenomenon that is worthy of epidemiologic investigation in its own right. Multilevel epidemiology seeks to understand multiple levels of inference, from genes to individuals to populations and could combine hypothesis-driven research with aspects of data mining. Multilevel epi-

demiology calls for the study of health and disease determinants defined at the population level and individual level for a more comprehensive strategy to understanding human disease etiology. With the continued development of multilevel statistical methods and the advent of data mining, the technical constraints of the past will become less relevant to the next generation of epidemiologists who wish to embrace a more multilevel epidemiology.

Keywords

Epidemiology · Methods · Multilevel · Risk factors · Gene-environment interactions · GIS · Geography

Die Zukunft der Epidemiologie. Methodische Herausforderungen und Multilevel-Interferenz

Zusammenfassung

Vor 10 Jahren gab es eine beachtliche Diskussion über die Ziele und Paradigmen moderner epidemiologischer Forschung. Eine in diesen Diskussionen vorgebrachte Sorge war, dass die „Risikofaktor-Epidemiologie“ unser Fachgebiet zwingen könnte, den Schwerpunkt mehr auf Einzelpersonen und weniger auf Populationen und die öffentliche Gesundheit zu legen. Heutzutage erkennen die meisten Epidemiologen an, dass die öffentliche Gesundheit von Determinanten sowohl auf Populations- als auch auf Individualebene beeinflusst wird. Ökologische Studien sind wertvolle Instrumente, um Hypothesen aufzustellen und sich mit Determinanten des Krankheitsrisikos auf Gruppenebene zu befassen. Herkömmliche Untersuchungen zu Risikofaktoren und genetische Studien haben dazu beigetragen, das multifaktorielle Konzept der Verursachung von Krankheiten zu etablieren. Auch durch Studien auf Individual-

ebene entstanden in der Biomedizin Hypothesen, die die Erforschung von Krankheitsmechanismen gefördert haben, was wiederum zu einer Verminderung der Morbidität und Mortalität von Krankheiten wie HIV/AIDS, Herz-Kreislauf-Krankheiten und Krebs geführt hat. Die aktuellen Diskussionen über die Rolle genetischer Daten in der Epidemiologie und im öffentlichen Gesundheitswesen spiegelt die Diskussionen über die Risikofaktor-Epidemiologie vor 10 Jahren. Die genetische Variation wird auf der Individualebene erfasst, aber wie diese Variation in der menschlichen Population erhalten wird, das ist ein Phänomen auf Gruppenebene (Population), das allein schon eine epidemiologische Untersuchung lohnt. Ziel der Mehrebenen-Epidemiologie ist das Verständnis mehrerer Ebenen von Inferenzen, von den Genen über Einzelpersonen bis hin zu Populationen. Dabei könnte hypothesenbasierte Forschung mit Aspekten

der gezielten Datensuche kombiniert werden. Die Mehrebenen-Epidemiologie erfordert für eine umfassendere Strategie zum Verständnis der Ätiologie menschlicher Krankheiten die Untersuchung von Gesundheits- und Krankheitsdeterminanten, die auf Populations- und Individualebene definiert sind. Mit der fortgesetzten Entwicklung statistischer Mehrebenen-Analysen und dem Aufkommen des Data Mining werden die technischen Einschränkungen der Vergangenheit für die nächste Generation von Epidemiologen weniger relevant sein. Sie werden sich eine mehr auf den Mehrebenen-Ansatz ausgerichtete Epidemiologie zu Eigen machen.

Schlüsselwörter

Epidemiologie · Methoden · Multilevel-Analysen · Risikofaktoren · Gen-Umwelt · Interaktion

en root, but healthy skepticism exists [16, 17, 18]. Some of the concerns about genomic information include a lack of information on the function of most genes, weak effects of most genes studied to date, a lack of available interventions and treatments, genetic discrimination, and data overload (particularly for so-called whole-genome scans).

Interestingly, a new series of debates over the goals of genome epidemiology and its role in public health mirror the debates about risk factor epidemiology one decade ago [17, 18]. As Merikangas and Risch (2003) point out in their commentary [18], interventions at the level of individuals or families are likely to be more successful for rare genetic diseases, while population-level interventions are likely to be successful for more common complex diseases. This certainly seems logical. For the epidemiology of common complex disease, it is likely that genomic information primarily will be important in the context of exposure (gene-environment interactions). The field of pharmacogenomics is actually the study of gene-environment interactions [19]. At least two examples display the power of this relatively new subfield: HER2/neu gene amplifications, trastuzumab, and breast cancer treatment response and survival [20]; and EGFR mutations, gefitinib, and lung cancer treatment response [21]. By characterizing and understanding these gene-environment interactions, the medical community may be better poised to target treatment and intervention programs to the groups who would benefit the most.

The study of specific gene-environment (statistical) interactions for common complex diseases may provide clues about underlying etiologic pathways involving complex exposures [e.g., MTHFR, diet (folate metabolism) and neural tube defects; NAT2, red meat (heterocyclic amine metabolism) and colorectal cancer; PPAR γ , dietary fat (adipocyte regulation, inflammation) and type 2 diabetes] [19]. If we believe these examples, then it should follow that the study of exposures while ignoring the genetics (and vice versa) runs the risk of missing important clues to human disease causation.

In her excellent commentary [22], Rockhill states that “the epidemiology perspec-

tive speaks to averages in large aggregates of individuals. It provides only limited information, if any, about biological mechanisms and causal pathways underlying specific cases of disease.” Based on current epidemiologic data, it is certainly true that we can’t point to a single case of disease and explain their entire pathway to illness, or accurately predict a healthy individual’s risk for developing a specific disease by a specific age. I believe our field’s contribution to these issues is a bit more subtle. Our field can provide hypotheses that basic scientists and others can test in animal models, laboratory assays, or in individual patients. There are numerous examples of exposures and diseases for which epidemiology has provided clues to basic scientists whose own research has furthered the understanding of basic disease mechanisms: HIV/AIDS and immune system function; *Helicobacter pylori*, gastric disease, and inflammatory processes; ionizing radiation, cancer, and DNA repair pathways; diet, birth defects, and folate metabolism; and nonsteroidal anti-inflammatory drugs (NSAIDs), cancer, and prostaglandin signaling and inflammation. Without individual-level (reductionist) epidemiology, some of these important exposure-disease mechanisms may not have been pursued in a timely fashion.

Genetic factors as multilevel attributes

In epidemiologic research, genetic factors such as polymorphisms (base pair changes, insertions, deletions, and repeats) should be thought of as having both individual-level and population-level attributes. We measure genetic markers in individuals, but genetic marker behavior (how they are maintained in populations) is a group-level phenomenon. The very definition of a polymorphism (alleles that are polymorphic in 1% or more of a population) is tied to the language of populations. Haplotypes are the linear arrangement of alleles on the same chromosome inherited as a group. Just as alleles show population-level characteristics, haplotypes do as well. The International HapMap Consortium project was formed to describe haplotype patterns in various populations around the world with the goal of identifying common hap-

lotypes for more efficient choice of genetic markers in association studies [23]. Many other genetic phenomena are attributes of populations rather than individuals. Hardy-Weinberg equilibrium, linkage disequilibrium, natural selection, migration, genetic drift, and bottlenecks are all population concepts, as is the concept of population substructure (stratification). If we can eliminate chance and genotyping error for a particular genetic marker, departures from Hardy-Weinberg equilibrium may be a phenomenon worth describing and pursuing in further research [24]. While simulation studies show that population stratification (a form of confounding in genetic association studies using unrelated controls) is unlikely to greatly bias most reported associations in non-Hispanic, European American populations [25], recent empirical evidence seems to suggest otherwise [26, 27]. Moreover, association studies in non-European and ethnically admixed populations could be more susceptible to this form of bias. Further research in different populations is clearly needed. In addition to population-level genetic phenomena discussed above, genome epidemiology can contribute knowledge for gene discovery (e.g., BRCA1/2), targeted intervention programs, as well as provide more refined estimates of gene penetrance [28]. Clearly, many epidemiologists have the opportunity to capitalize on our wealth of environmental and genetic data in this post-genomics era [16].

Genome-wide and gene-wide association studies

In the recent past, molecular epidemiologists conducting association studies of genetic polymorphisms and disease tended to study one or a few markers in just one or a few hypothesized genes (the candidate gene approach). As genotyping technology becomes more affordable and scalable, more epidemiologists will begin evaluating many thousands of markers through the use of genome-wide chips. Currently, there are several platforms and chips to choose from with varying genomic coverage. Ideally, a true genome-wide scan should include at least one marker for every 1,000 base pairs spread evenly across the entire genome. Currently, a true

genome-wide scan is still unaffordable for most epidemiologic studies of even modest size. Further, the statistical methods to analyze many thousands of markers are still under development. Model-free and nonparametric data mining methods such as multifactor dimensionality reduction (MDR) and combinatorial partitioning methods (CPM) hold promise for detecting gene-gene and gene-environment interactions [29, 30, 31], but are faced with computational challenges in the context of thousands of markers [32]. This is all likely to change. In the mean time, epidemiologists could utilize an interesting compromise between candidate gene and whole-genome association known as gene-wide association analysis [33]. With this method most or all of the sequence of genes hypothesized to be associated with the disease (candidate genes) would be evaluated by genotyping markers that are spread as evenly as possible across the gene including promoter regions and most exons and introns. Publicly available haplotype data (e.g., the HapMap) could be used to efficiently choose nonredundant markers across candidate genes for association analysis. However, the assumption that haplotype information obtained from the HapMap project is relevant for a wide range of markers in different populations needs to be evaluated [34]. Many epidemiologists have the data and training to test these assumptions and contribute to the ongoing debates. The study of genetic variation and gene-environment interaction has the potential to inform population scientists and basic scientists about the complexity of disease pathways in populations. Currently, inherited variation in genes is not amenable to change, but this may not be the case in the future as advances are made in gene therapy and we gain a greater understanding of epigenetic phenomena [35].

Multilevel inference and multilevel epidemiology

If we believe that risk factor epidemiology has led our field to focus more on individuals and less on populations and public health, the logical improvement is to conceptualize epidemiology as a “complex systems” approach [36] in which we

make inferences at multiple levels (e.g., molecules, individuals, groups, and populations). In 1996, Susser and Susser defined “eco-epidemiology” as the study of causal processes at the societal level with pathogenesis and causality at the molecular level [4]. In this commentary I describe “multilevel epidemiology” (or “systems epidemiology”) as the study of causal processes at all relevant levels of inference, both at the level of the population and the individual. (I prefer the term “multilevel” over “eco-epidemiology” as it avoids the somewhat negative spell of purely ecological studies.) Individual-level attributes can be conceptualized at the level of DNA, genes, haplotypes, RNA, enzymes, cells, tissues, individual lifestyles, and families. Population-level attributes can be conceptualized at the level of neighborhoods, communities, societies, cultures, economies, nations, ethnicities, and population genomics. Obviously, some factors will share attributes of populations and individuals to a varying degree (e.g., families are influenced by environmental, social, lifestyle, and genomic factors). Just as we ask personal and family histories using standardized questionnaires, we also would need to understand and consider local, regional, and societal population histories (e.g., cultural practices, politics, industrialization, and urbanization) [37]. The etiologic framework of multilevel epidemiology must be conceptualized at the individual level and population level. Such studies would help our field by moving from the study of abstract collections of risk factors in a convenience sample of individuals to the study of relevant health issues in populations [38]. These studies should describe and define complex systems, including interactions between potential causes at all levels of inference and should acknowledge the limits on generalizations about human and social systems and not seek to describe a “universal” dose-response relationship [4].

Multilevel epidemiology could combine hypothesis-driven research with aspects of data mining. Methods for data reduction, such as machine learning, especially in the context of thousands of comparisons and multiple dimensions, hold promise but are in their infancy. Further development of epidemiologic applications of multi-

level and mixed model approaches for handling multiple levels of inference (and hierarchical data structures) also is needed, particularly for models that can incorporate clustering at more than two levels and for data in the form of counts, binary or multiple outcomes, and for survival analysis [38]. I am not necessarily calling for epidemiologists to include all variables at all levels simultaneously in a single model. Rather, I am calling for a more multilevel approach when formulating research questions with the hope that this will foster further research in multilevel modeling and force us to begin to think outside of our own subspecialties.

Multilevel epidemiology should seek to inform population-level interventions and individual-level interventions. Accountability for good health lies with individuals and with society. A rather simple example is tobacco cessation. In the USA, campaigns to change smoking behavior targeted to individuals have been much less successful than interventions targeted to groups (i.e., social and public policy changes) such as bans on smoking in restaurants, bars, and other public places [39]. In the European Union (EU), the rather striking geographic variation in lung cancer incidence rates is believed to reflect the underlying variation in cancer control programs (including anti-smoking campaigns), cultural attitudes about smoking, as well as regulatory policies directed at the tobacco industry [40].

Many epidemiologists have avoided a more multilevel concept of epidemiology in their research. The reasons for this may include the past criticisms of ecologic studies, the ease of conceptualization of individual-level factors over population-level factors, a lack of interest by funding agencies, and insufficient training, not to mention the sheer complexity of incorporating so many variables in one statistical model. Most funding agencies support epidemiologic research that mainly focuses on individual lifestyle or genetic factors and less on socioeconomic or other group-level factors. Moreover, most training programs in epidemiology focus on subspecialties (e.g., molecular, genetic, cancer, cardiovascular, environmental, perinatal, infectious, occupational, injury, social, psychiatric, and health services). Students of epidemiol-

ogy are often forced to identify one area of expertise, while gaining superficial knowledge or simply ignoring other subspecialties. Perhaps this is becoming more unavoidable as epidemiology continues to subspecialize (e.g., genome and proteome epidemiology), and, rather we should be teaching more courses in biostatistics, methods development, and even ways to initiate and maintain effective collaborations. The ultimate conceptualization of multilevel epidemiology would require multidisciplinary teams of researchers in addition to ambitious epidemiologists.

In conclusion, multilevel epidemiology calls for the study of health and disease determinants defined at multiple levels of inference. Multilevel epidemiology should seek to address public health problems in populations, and incorporate group-level (social, economic, environmental) data with individual-level (risk factor, biomarker, and genomic) data for a more comprehensive strategy to understanding human disease etiology. The incorporation of group-level data in individual-level studies (if done carefully) can only strengthen our already influential field. With the further development of multilevel statistical methods and data mining, the technical constraints of the past will become less relevant to the next generation of epidemiologists.

Korrespondierender Autor

Dr. Eric J. Duell

Norris Cotton Cancer Center,
Section of Biostatistics & Epidemiology,
Center for the Evaluative Clinical Sciences,
Department of Community & Family Medicine,
Dartmouth Medical School, Lebanon,
New Hampshire, USA
e-mail: Eric.Duell@Dartmouth.Edu

Acknowledgments

The author would like to thank Dr. Jason Moore and an anonymous reviewer for helpful comments on earlier versions of the manuscript. Dr. Duell was supported by NIH grant number 1 P20 RR018787 from the Institutional Development Award (IDeA) Program of the National Center for Research Resources and by NIH grant number CA98889 (to E.Duell).

References

- Krieger N (1994) Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med* 39(7):887–903
- Savitz DA (1994) In defense of black box epidemiology. *Epidemiology* 5(5):550–552
- Susser M, Susser E (1996) Choosing a future for epidemiology: I. Eras and paradigms. *Am J Public Health* 86(5):668–673
- Susser M, Susser E (1996) Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *Am J Public Health* 86(5):674–677
- Rose G (1985) Sick individuals and sick populations. *Int J Epidemiol* 14(1):32–38
- Savitz DA, Poole C, Miller WC (1999) Reassessing the role of epidemiology in public health. *Am J Public Health* 89(8):1158–1161
- Morgenstern H (1995) Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health* 16:61–81
- Breeze E, Jones DA, Wilkinson P et al. (2005) Area deprivation, social class, and quality of life among people aged 75 years and over in Britain. *Int J Epidemiol* 34(2):276–283
- Huynh M, Parker JD, Harper S et al. (2005) Contextual effect of income inequality on birth outcomes. *Int J Epidemiol* 34(4):888–895
- Austin SB, Melly SJ, Sanchez BN et al. (2005) Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments. *Am J Public Health* 95(9):1575–1581
- Dreassi E (2003) A space-time analysis of the relationship between material deprivation and mortality for lung cancer. *Environmetrics* 14:511–521
- Krieger N (2003) Place, space, and health: GIS and epidemiology. *Epidemiology* 14(4):384–385
- Ricketts TC (2003) Geographic information systems and public health. *Annu Rev Public Health* 24:1–6
- Rushton G (2003) Public health, GIS, and spatial analytic tools. *Annu Rev Public Health* 24:43–56
- Kristal AR, Peters U, Potter JD (2005) Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev* 14(12):2826–2828
- Millikan R (2002) The changing face of epidemiology in the genomics era. *Epidemiology* 13(4):472–480
- Khoury MJ, Davis R, Gwinn M et al. (2005) Do we need genomic research for the prevention of common diseases with environmental causes? *Am J Epidemiol* 161(9):799–805
- Merikangas KR, Risch N (2003) Genomic priorities and public health. *Science* 302(5645):599–601
- Hunter DJ (2005) Gene-environment interactions in human diseases. *Nat Rev Genet* 6(4):287–298
- Slamon DJ, Leyland-Jones B, Shak S et al. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 344(11):783–792
- Shigematsu H, Lin L, Takahashi T et al. (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 97(5):339–346
- Rockhill B (2005) Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology* 16(1):124–129
- The International HapMap Project (2003) *Nature* 426(6968):789–796
- Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet* 76(6):967–986
- Wacholder S, Rothman N, Caporaso N (2000) Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 92(14):1151–1158
- Helgason A, Ingvaldottir B, Hrafnkelsson B et al. (2005) An Icelandic example of the impact of population structure on association studies. *Nat Genet* 37(1):90–95
- Campbell CD, Ogburn EL, Lunetta KL et al. (2005) Demonstrating stratification in a European American population. *Nat Genet* 37(8):868–872
- Thomas DC (2000) Genetic epidemiology with a capital "E". *Genet Epidemiol* 19(4):289–300
- Ritchie MD, Hahn LW, Roodi RE, Sing CF (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 69(1):138–147
- Moore JH (2004) Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn* 4(6):795–803
- Nelson MR, Kardia SL, Ferrell RE, Sing CF (2001) A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res* 11(3):458–470
- Moore JH, Ritchie MD (2004) STUDENTJAMA. The challenges of whole-genome approaches to common diseases. *Jama* 291(13):1642–1643
- Neale BM, Sham PC (2004) The future of association studies: gene-based analysis and replication. *Am J Hum Genet* 75(3):353–362
- Nothnagel M, Rohde K (2005) The effect of single-nucleotide polymorphism marker selection on patterns of haplotype blocks and haplotype frequency estimates. *Am J Hum Genet* 77(6):988–998
- Lyko F, Brown R (2005) DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. *J Natl Cancer Inst* 97(20):1498–1506
- Loomis D, Wing S (1990) Is molecular epidemiology a germ theory for the end of the twentieth century? *Int J Epidemiol* 19(1):1–3
- Schwartz S, Susser E, Susser M (1999) A future for epidemiology? *Annu Rev Public Health* 20:15–33
- Diez-Roux AV (2000) Multilevel analysis in public health research. *Annu Rev Public Health* 21:171–192
- Jha P, Chaloupka FJ (2000) The economics of global tobacco control. *Bmj* 321(7257):358–361
- Zielinski SL (2005) Diversity in cancer rates, control programs matches diversity of European Union. *J Natl Cancer Inst* 97(19):1398–1399