Globalization, International Law, and Emerging Infectious Diseases

David P. Fidler, J.D.
Indiana University School of Law,
Bloomington, Indiana, USA

The global nature of the threat posed by new and reemerging infectious diseases will require international cooperation in identifying, controlling, and preventing these diseases. Because of this need for international cooperation, international law will certainly play a role in the global strategy for the control of emerging diseases. Recognizing this fact, the World Health Organization has already proposed revising the International Health Regulations. This article examines some basic problems that the global campaign against emerging infectious diseases might face in applying international law to facilitate international cooperation. The international legal component of the global control strategy for these diseases needs careful attention because of problems inherent in international law, especially as it applies to emerging infections issues.

The growing literature on new and reemerging infectious diseases often emphasizes the global nature of their threat; the U.S. Centers for Disease Control and Prevention (CDC) defines these diseases as "diseases of infectious origin whose incidence in humans has increased within the past two decades or threatens to increase in the near future" (1). The World Health Organization has asserted that emerging infections "represent a global threat that will require a coordinated, global response" (2). The threat is global because a disease can emerge anywhere on the planet and spread quickly to other regions through trade and travel. The global challenge of emerging infections has serious consequences for national and international law; a state's ability to deal with them is eroded because microbes do not respect internationally recognized borders (3). Experts grappling with these diseases no longer consider that the pursuit of a strictly national public health policy is adequate. The need for global cooperation increases the importance of international law in the public health arena. Part of the effort to create a global response to emerging infections should be an understanding of the problems that may arise from relying on international law in dealing with these diseases. This article outlines issues that will have to be confronted in using international law to combat emerging infections.

Address for correspondence: David P. Fidler, J.D., Indiana University School of Law, Third Street and Indiana Avenue, Bloomington, IN 47405-1001 USA; fax: 812-855-0555; e-mail: davidfidler@law.indiana.edu.

Globalization

The assertion that emerging infections are a global problem requiring a global strategy echoes observations made in other spheres of public policy: the traditional distinctions between national and international political, social, and economic activities are losing their importance (4). Globalization is eroding traditional distinctions between domestic and foreign affairs. Globalization has been defined as the "process of denationalization of markets, laws, and politics in the sense of interlacing peoples and individuals for the sake of the common good" (5). Globalization is distinguished from internationalization, which is defined "as a means to enable nation-states to satisfy the national interest in areas where they are incapable of doing so on their own" (5). Internationalization involves cooperation between sovereign states, whereas globalization refers to a process that is undermining or eroding sovereignty.

Globalization arises from the confluence of something old and something new in international relations. It involves the very old process of political and economic intercourse among sovereign states. The new element is the intensification and expansion of such intercourse made possible by technological advances in travel, communications, and computers. Encouraging such intensification and expansion is liberal economic thinking, which posits that economic interdependence makes all states economically better off and builds order and peace in the international system (6).

The changes wrought by new technologies unleashed in the receptive international milieu

created by liberal trade and economic policies have led to the belief that these developments are undermining sovereignty. Observers of international relations frequently note that governments no longer have control over economic forces at work within their countries. The speed and volume of international capital flows illustrate the denationalization of economics occurring through the process of globalization (7). Another example is the development of the global company—an enterprise that can no longer be considered national because of the global reach of its operations, financing options, markets, and strategies (7). The globalization of finance and business has ramifications for politics and law as leaders and legal systems adapt to the global era (8).

In public health, a similar combination of old and new factors can be seen. States have historically cooperated on infectious disease control, first through international sanitary treaties and later through the World Health Organization (WHO) (9). While international cooperation is not new, current global circumstances confronting the control of infectious disease are. Globalization is also at work in public health. The assertion that a country cannot tackle emerging infectious diseases by itself demonstrates that public health policy has been denationalized.

Globalization has affected public health in three ways. First, the shrinking of the world by technology and economic interdependence allows diseases to spread globally at rapid speed. Two factors contributing to the global threat from emerging infections stem directly from globalization: the increase in international travel (2, 10) and the increasingly global nature of food handling, processing, and sales (2, 10). HIV/AIDS, tuberculosis, cholera, and malaria represent a few infections that have spread to new regions through global travel and trade (10). The beneficial economic and political consequences of economic interdependence may have negative ramifications for disease control. In the European Union, for example, the free movement of goods, capital, and labor makes it more difficult for member states to protect domestic populations from diseases acquired in other countries (11).

Second, the development of the global market has intensified economic competition and increased pressure on governments to reduce expenditures, including the funding of public health programs, leaving states increasingly unprepared to deal with emerging disease problems. Industrialized as well as developing countries confront deteriorating public health infrastructures (12). Referring to the United States, one author described this deterioration as the "thirdworldization" of the American health care system (13).

Third, public health programs have also "gone global" through WHO and health-related nongovernmental organizations. Medical advances have spread across the planet, improving health worldwide. The worldwide eradication of smallpox in 1977 is a famous example. The global reach of health care advances has, however, a darker side. The globalization of disease control has contributed to the population crisis because people are living longer. Overpopulation creates fertile conditions for the spread of disease: overcrowding, lack of adequate sanitation, and overstretched public health infrastructures (2). Further, the widespread use and misuse of antibiotic treatments has contributed to the development of drug-resistant pathogens (1, 2). Finally, the success of control efforts in previous decades caused interest in infectious diseases to wane in the international medical and scientific communities and is now hampering emerging infectious disease control efforts (14).

International Solutions to Emerging Infections

International efforts are under way to respond to the threat of emerging infectious diseases. WHO and CDC have drafted action plans that stress the need to strengthen global surveillance of these diseases and to allow the international community to anticipate, recognize, control, and prevent them (1, 14, 15). WHO has also established a new unit to control and prevent emerging infections by mobilizing resources rapidly at the first signs of outbreaks (16). The Pan American Health Organization has also adopted a regional plan for controlling emerging infections in the Americas (17). Health authorities from Central American countries have adopted an emergency plan to control the epidemics of dengue and dengue hemorrhagic fever that recently swept through Central and South America (18). Physicians in the European Union recognize the need for better surveillance of infectious diseases (11). A U.S. government interagency working group has underlined the importance of international cooperation in dealing with the emerging infections threat (19). The U.S. Senate Labor and Human Resources Committee held hearings in October 1995 on "Emerging Infections: A Significant Threat to the Health of the

Nation" (20). At the Halifax Summit in 1995, the major industrialized countries adopted a pilot project called "Toward a Global Health Network" designed to help governments deal with emerging infections and other health problems (19) (Table 1). Clearly, the emerging infections threat and the need for action are on the international diplomatic and public health agendas.

Although international control plans would involve private organizations like universities and nongovernmental organizations, the primary actors on the emerging infections stage are sovereign states. The action plans are predominantly blueprints for cooperation among states and represent a call for the internationalization of responses to a problem caused by globalization. Put another way, the proposed solutions to the emerging infections threat rely on the sovereign state, while the threat feeds off the impotence of the state in addressing global disease problems. When it comes to public health activities, globalization erodes sovereignty, but the proposed solution makes sovereignty and its exercise critical to dealing with the threat of emerging infections.

The consequences of the unavoidable emphasis on international cooperation in the proposed action plans for emerging infections are troubling. To achieve the desired objectives (Table 1), states will have to agree on many issues and translate such agreement into guidelines or rules. International law becomes important to the effort for emerging infections control. Political leaders, diplomats, and scholars have long recognized the weakness of international law in regulating state behavior. At first glance, the prospect of having to rely on a notoriously weak institution of international relations as part of the global effort to combat emerging infections is unsettling.

Table 1. Some common elements of global emerging-disease control plans

Strengthen international surveillance networks to detect, control, and reduce emerging diseases.
Improve the international public health infrastructure (e.g., laboratories, research facilities, technology, and communications links).
Develop better international standards, guidelines, and recommendations.
Improve international capabilities to respond to disease outbreaks with adequate medical and scientific resources and expertise.
Strengthen international research efforts on emerging diseases, particularly with regards to antibiotic-resistant strains of diseases.
Focus attention and resources on training and supporting medical and scientific expertise.
Encourage national governments to improve their public health care systems, devote resources to eliminating or controlling causes of emerging diseases and coordinate their public health activities with WHO and the international community.

Sources: refs. 1, 14, 15, 19.

International Law and Infectious Disease Control

We might have been less unsettled if our experience with international law in controlling infectious diseases had been more positive. The success of WHO in globalizing disease control programs might suggest that the defects of international law have not hobbled its effectiveness in improving health care worldwide. However, despite having the authority to do so, WHO has been reluctant to use international law (21, 22). The International Health Regulations administered by WHO represent the most important set of international legal rules relating to infectious disease control, but the regulations only apply to plague, yellow fever, and cholera (23). The importance of health is mentioned in international declarations (for example, see the Universal Declaration of Human Rights, art. 25 [1]) and treaties (for example, see the International Covenant on Economic, Social and Cultural Rights, art. 12), leading some legal scholars to argue that international law creates a "right to health" (24); but this "right" does not directly address the control of infectious diseases. WHO has refrained from adopting rules on trade in human blood and organs, which does raise issues of infectious disease control as illustrated by the sale of HIV-contaminated blood in international commerce (25). Issues of disease control also appear in specialized treaty regimes outside WHO, such as treaties controlling marine pollution from ships (26). Other areas of international public health law, for example, rules about infant formula and guidelines on pharmaceutical safety, do not deal with the control of infectious diseases (25).

The effectiveness of existing international law on infectious disease control has been questioned. A 1975 WHO publication stated that the International Health Regulations have not functioned satisfactorily at times of serious disease outbreaks (27). More recently, WHO's efforts with the International Health Regulations have been called a failure, and noncompliance with these regulations

has increased in connection with reporting disease outbreaks (25). The HIV/AIDs crisis dramatically illustrated the weaknesses of the health regulations. Since AIDs was not originally (or subsequently) made subject to the regulations, states had, and continue to have, no notification requirements in connection with this new disease. Further, as HIV/AIDs spread globally, many states adopted exclusionary policies that, according to experts, violated provisions of the health regulations (25). In relation to one of the biggest disease crises of this century, parts of the International Health Regulations were irrelevant, and other parts were openly violated.

WHO's reluctance to apply international law has been attributed to its organizational culture, which is dominated by scientists, doctors, and medical experts. Perhaps the current weakness of international law on infectious disease control reflects WHO's nonlegal strategy rather than the inherent problems in international law itself. In connection with emerging infections, however, WHO is advocating an international legal strategy by recommending revision of the International Health Regulations (28). This recommendation suggests that WHO acknowledges the need for international legal agreement in dealing with emerging infections. The global threat posed by these infections represents in many ways a test case for international public health law.

The Challenge to International Law

The threat of emerging infectious diseases poses two challenges to international law: first, the emerging infections problem exacerbates basic weaknesses in the law. Second, these infections pose specific difficulties in the law, which are related to the nature of disease and its prevention.

Basic Weaknesses

The effectiveness of international law depends on the consent of states, which means that sovereignty and its exercise determine the fate of international legal rules (29). In adopting a legal strategy for its emerging infectious disease action plan, WHO has to convince its member states to take certain actions in response to disease emergence. The sovereignty of states looms large in formulating a global response to emerging infections, despite the fact that the process of globalization undermines the sovereignty of the state to deal nationally with these infections. In other words, the problem by-passes the state, but the solution has to rely on the state through the medium of international law. The central importance of the state and its sovereignty constitutes a basic weakness in international law because international legal rules tend to reflect the compromises necessary to achieve agreement and the unwillingness of states to restrict their freedom of action through international law. Part of the reason that the existing International Health Regulations cover only a few diseases might be the unwillingness of WHO member states to commit to more serious infectious disease control measures. The vagueness and lack of specificity in the so-called "right to health" also illustrate this problem. What is scientifically and medically necessary to combat emerging diseases may not be what states are willing to agree to undertake.

A second basic weakness follows from the "sovereignty problem"—the lack of effective enforcement of international law. States often agree to an international legal obligation without any serious intent of fulfilling it. The alleged failure of the International Health Regulations may be due to the failure of WHO member states to fulfill the duties they accepted. Neither the regulations nor WHO has any power to enforce compliance (25). An international legal regime on emerging diseases would also face this enforcement problem.

Specific Difficulties

The very nature of the emerging disease threat poses special difficulties for international law. The global scope of the problem necessitates agreement by most states to control emerging diseases. If any major country or group of countries does not participate, a gap in the global surveillance and control network threatens the efficacy of the entire effort. The negotiation of agreements involving many states is usually difficult, because each state knows that its nonparticipation threatens the success of the entire venture. This problem has occurred in international environmental law, where global regimes have been needed to deal adequately with environmental threats, such as ozone depletion.

A second specific difficulty arises from the extent of medical and scientific resources needed to establish an effective global surveillance and control network for emerging diseases. Fundamental aspects of the proposed action plans involve improving surveillance networks, public health infrastructures, scientific research, and

medical and scientific training (Table 1). Some states, particularly in the developing world, do not have the medical, scientific, and financial resources to undertake such measures. Unless more affluent countries provide the resources, developing states may use the inequity of wealth in the international system as an argument to complicate negotiating a global agreement. The so-called "North-South problem" has made the negotiation of international environmental agreements more difficult, as developing countries have bargained for more lenient treatment or a transfer of resources from affluent countries to help them improve environmental protection. A similar dynamic may appear in any negotiations for a global emerging disease effort. The U.S. interagency working group on emerging diseases has observed that major U.S. contributions to developing countries for emerging disease control purposes "is not a likely prospect during this period of deficit reduction and downsizing" (19), which suggests that resource availability will probably complicate international efforts in this area.

The problems associated with using international law in a global strategy to combat emerging diseases raise the question whether international law can provide an adequate foundation for the control of these diseases. The uncomfortable position of having no choice but to rely on international law when its weaknesses are substantial highlights the importance of thinking through the international legal aspects of a global emerging disease plan carefully.

WHO's Proposed Legal Strategy

WHO wants to revise the International Health Regulations as part of its global emerging disease strategy (28). WHO's proposal deserves some critical attention. It is not clear that the organization has adequate authority to incorporate comprehensive emerging disease control measures within the international regulations. Under Article 21 of the WHO Constitution, the World Health Assembly can adopt binding regulations in sanitary and quarantine requirements and other procedures to prevent the international spread of disease (22). The World Health Assembly adopted the International Health Regulations under Article 21. While Article 21 and the regulations are relevant to emerging disease control efforts, it is doubtful whether the regulations can serve as a foundation for a comprehensive emerging disease control plan. The disease-outbreak notification requirements in the regulations could be expanded to include more diseases, but nothing in Article 21 gives the World Health Assembly the authority to require WHO member states to strengthen public health infrastructures, which is considered critical in the emerging disease actions plans proposed to date (Table 1). It has been argued that attempting to address such infrastructure problems "is a solution which cannot be obtained by an international instrument but only by the improvement of the health conditions of the peoples of WHO's member states" (30). But, as the history of administering the International Health Regulations has shown, notification requirements have not worked satisfactorily and are weakened by the absence of adequate public health resources. Further, Article 22 of the WHO Constitution makes regulations promulgated under Article 21 automatically binding on WHO member states, except for member states that reject such regulations or make reservations thereto (31). Article 22 relates to the sovereignty problem and may deter WHO member states from agreeing to serious revisions of the regulations. Analysis of the regulations may question the wisdom of using the regulations as the legal basis for dealing with emerging diseases.

The World Health Assembly has the power to adopt conventions or agreements within WHO's competence (21). The Assembly could use this authority to address aspects of the global emerging disease control strategy that cannot be handled with a revision of the regulations. However, parceling up emerging disease control measures between the International Health Regulations and separate agreements would be legally complicated. Further, WHO has not used this power to adopt conventions or agreements, which explains its unwillingness to explore all legal options open to it.

Possible Alternative Legal Strategies

Alternative legal strategies to revising the International Health Regulations range from reliance on the development of customary international law to the adoption of multilateral treaties specifically on emerging-disease control (Table 2). An issue related to these alternative approaches is the substantive nature of the obligations contained in legal documents. We have to ask not only how states might agree on control rules but also what these states might agree to do. The proposed revision of the regulations

Table 2. Alternative international legal strategies to revising the International Health Regulations

Alternative legal strategies	Possible advantages	Possible disadvantages
1. WHA incorporates emerging disease control as part of the proposed World Health Charter scheduled for initial negotiations in 1997	Integrates emerging disease control measures into the overall WHO approach to international health issues	a. Emerging disease control would not be primary focus b. World Health Charter is likely to be more aspirational than obligatory
2. WHA adopts an emerging disease–specific convention under Article 19 of the WHO Constitution	a. Avoids IHR model b. Has potential to set out comprehensive global approach to emerging diseases	a. WHA has no experience with using Article 19 b. Large multinational treaties tend to contain general obligations rather than specific duties
3. States negotiate a framework multilateral treaty on general emerging disease obligations, accompanied by disease-specific or region-specific protocols containing detailed and specific commitments on emerging disease control	a. Takes emerging disease control out of WHO, eliminating problem of WHO's reluctance to use international law b. Allows for new protocols to be adopted for new diseases c. Framework-protocol approach has been used with some success in international environmental law on ozone depletion	a. WHO has to play central role in any emerging disease plan b. Framework-protocol approach might not be appropriate model for emerging disease control because the emerging disease problem differs from ozone depletion
4. Encourage regional arrangements and integrate them into global regime over time	a. Builds on strong regional systems of cooperation and coordination b. Offers "legal laboratories" to try various approaches to emerging disease control c. Avoids diplomatic headaches involved in trying to negotiate truly global legal regimes	a. Emerging diseases require a global approach not just a regional approach b. Amounts to emerging disease control for rich regions, leaving many developing countries outside legal regime c. Risks inconsistencies in how emerging diseases are handled by different regions
5. Encourage a bilateral approach in which individual countries negotiate detailed and specific commitments on emerging diseases and perhaps condition trade benefits and aid on emerging disease performance	a. Gives states flexibility in constructing legal obligations b. Permits possibility for sanctions for failure to live up to emerging disease obligations	a. Does not address global nature of emerging disease problem b. Sanctions element is unrealistic and might be unfair to developing countries lacking the resources necessary to implement adequate emerging disease control measures
6. Incorporate emerging disease control as part of international "right to health," making emerging diseases a human rights issue	a. Links emerging disease control with larger, powerful concepts of human welfare b. Builds on existing international law on the "right to health"	a. International "right to health" has no definitive meaning or scope and thus is a bad foundation for emerging disease control b. Human rights are inherently divisive in the international system; linkage with such a controversial area would hurt emerging disease control prospects
7. Rely on customary international law to develop emerging disease-control norms	Customary international norms on emerging disease control would be binding on all states except persistent objectors	a. It will be nearly impossible to develop general and uniform state practice recognized by states as legally binding in the emerging disease-control area b. Any customary norms that might form will probably be vague and hard to identify definitively c. Customary norms can take a very long time to develop

WHA = World Health Assembly; WHO = World Health Organization; IHR = International Health Regulations.

apparently would only apply the notification duties (currently found in the regulations) to more diseases. As indicated earlier, WHO cannot address in its revision of the regulations any of the improvements in public health infrastructures, surveillance networks, scientific research, or medical and scientific training at the heart of proposed emerging disease action plans. Further, it is not clear whether WHO intends to supplement expanded notification duties with any mechanism to monitor or enforce such duties.

International environmental law had to overcome some of the same obstacles encountered by WHO's international legal effort for emerging disease control. States realized that they could not handle global environmental problems without international cooperation and rules (32). Further, states knew that addressing environmental concerns would require changes for governments and companies within states and that developing states might have financial and technological difficulties implementing international agreements (32). In developing international environmental law, states, international organizations, and nongovernmental organizations did not rely on old approaches but instead crafted new international legal rules to deal with the global nature of the threats posed, the resource issue, and compliance and enforcement problems (33). Whether international environmental law has been successful is controversial; but it is important that states have not been willing to admit that improving environmental conditions within states is a solution that cannot be obtained by international agreements. Models and precedents from international environmental law are not in all respects helpful to the challenge of emerging-disease control; but, at the very least, those grappling with an international strategy for the emerging-disease threat could analyze international environmental law and other innovative legal responses to globalization to look for ways of making WHO's international legal strategy on emerging diseases as effective as possible.

Those currently designing global emerging-disease control strategies will eventually have to translate what is scientifically and medically needed to combat these diseases into international agreement and cooperation through international law. The movement from science and medicine into the realm of international law will not be easy. Relying on the International Health Regulations as the centerpiece of international law on emerging-disease control may not be the most effective international legal strategy. Whatever international legal approach is eventually taken will have to confront somehow a fundamental paradox: globalization jeopardizes disease control nationally by eroding sovereignty, while the need for international solutions allows sovereignty to frustrate disease control internationally. The combination of the process of globalization and the unavoidable need to rely on international law produces a most unattractive medium in which to wage potentially one of the most important medical and scientific endeavors in history.

Dr. Fidler is associate professor of law at Indiana University School of Law, Bloomington, where he teaches public and private international law. He has a master of philosophy (M.Phil.) degree in international relations from Oxford University, a J.D. from Harvard Law School, and a bachelor of civil law degree from Oxford University. Before joining the faculty at Indiana University School of Law, Mr. Fidler practiced law with Sullivan & Cromwell in London, England, and with Stinson, Mag & Fizzell, P.C., in Kansas City, Missouri.

Acknowledgments

I thank Jack Bobo, research assistant, for his valuable work during the preparation of this article, and Lane Porter and Allyn Lise Taylor for their helpful comments on earlier drafts of this article. I also thank my colleagues on the faculty of Indiana University School of Law for their helpful comments on an earlier draft presented at a faculty colloquium.

References

1. Centers for Disease Control and Prevention. Addressing emerging infectious disease threats: a prevention strategy for the United States. Atlanta: U.S. Department of Health and Human Services, Public Health Service, 1994.
2. World Health Organization. Communicable disease prevention and control: new, emerging, and re-emerging infectious diseases. WHO Doc. A48/15; Feb. 22, 1995.
3. Garrett L. The return of infectious disease. Foreign Affairs Jan.-Feb. 1996;66-79.
4. Aman AC. Introduction. Indiana Journal of Global Legal Studies 1993;I:1-8.
5. Delbrück J. Globalization of law, politics, and markets—implications for domestic law—a European perspective. Indiana Journal of Global Legal Studies 1993;I:9-36.

6. Gilpin R. The political economy of international relations. Princeton, NJ: Princeton University Press, 1987.
7. Kennedy P. Preparing for the twenty-first century. London: Harper Collins, 1993.
8. Shapiro M. The globalization of law. Indiana Journal of Global Legal Studies 1993;I:37-64.
9. McNeill WH. Plagues and peoples. New York: Doubleday, 1976.
10. Wilson ME. Travel and the emergence of infectious diseases. Emerging Infectious Diseases 1995;1:39-46.
11. Desenclos J-C, Bijkerk H, Husiman J. Variations in national infectious disease surveillance in Europe. Lancet 1993;341:1003-6.
12. Berkelman RL, Bryan RT, Osterholm MT, Leduc JW, Hughes JM. Infectious disease surveillance: a crumbling foundation. Science 1994;264:368-70.
13. Garrett L. The coming plague: newly emerging diseases in a world out of balance. New York: Penguin Books, 1994.
14. Emerging infectious diseases: memorandum from a WHO meeting. Bull World Health Organ 1994;72:845-50.
15. World Health Assembly. Communicable diseases prevention and control: new, emerging, and re-emerging infectious diseases. WHO Doc. WHA 48.13, May 12, 1995.
16. World Health Organization. Press Release. WHO/75, Oct. 17, 1995.
17. Epstein DB. Recommendations for a regional strategy for the prevention and control of emerging infectious diseases in the Americas. Emerging Infectious Diseases 1995;1:103-5.
18. World Health Organization. Press Release. WHO/72, Sept. 28, 1995.
19. National Science and Technology Council Committee on International Science, Engineering, and Technology Working Group on Emerging and Re-Emerging Infectious Diseases. Infectious disease—a global health threat. Washington, DC: CISET, 1995.
20. Emerging infections: a significant threat to the nation's health: hearing before the comm. on labor and human resources. 104th Cong., 1st Sess. 298 (1995).
21. WHO Constitution, art. 19.
22. WHO Constitution, art. 21.
23. World Health Organization. International health regulations. Geneva: World Health Organization, 1983.
24. Taylor AL. Making the world health organization work: a legal framework for universal access to conditions for health. Am J Law Med 1992;18:301-46.
25. Tomasevski K. Health. In: Scachter O, Joyner CC, editors. United Nations Legal Order. Cambridge, UK: Cambridge University Press, 1995.
26. Churchill RR, Lowe AV. The law of the sea. 2nd ed. Manchester, UK: Manchester University Press, 1988.
27. Delon PJ. The International Health Regulations: a practical guide. Geneva: World Health Organization, 1975.
28. World Health Assembly. Revision and Updating of the International Health Regulations. WHO Doc. WHA 48.7, May 12, 1995.
29. Brownlie I. Principles of public international law. 4th ed. Oxford, UK: Oxford University Press, 1990.
30. Fluss S. International public health law: an overview. In: Oxford Textbook of Public Health. In press.
31. WHO Constitution, art. 22.
32. Hurrell A, Kingsbury B. The international politics of the environment: an introduction. In: Hurrell A, Kingsbury B, editors. The international politics of the environment. Oxford, UK: Oxford University Press, 1992.
33. Birnie P. International environmental law: its adequacy for present and future needs. In: Hurrell A, Kingsbury B, editors. The international politics of the environment. Oxford, UK: Oxford University Press, 1992.

On Epidemiology and Geographic Information Systems: A Review and Discussion of Future Directions

Keith C. Clarke, Ph.D., Sara L. McLafferty, Ph.D,. and Barbara J. Tempalski
Hunter College-CUNY, New York, New York, USA

Geographic information systems are powerful automated systems for the capture, storage, retrieval, analysis, and display of spatial data. While the systems have been in development for more than 20 years, recent software has made them substantially easier to use for those outside the field. The systems offer new and expanding opportunities for epidemiology because they allow an informed user to choose between options when geographic distributions are part of the problem. Even when used minimally, these systems allow a spatial perspective on disease. Used to their optimum level, as tools for analysis and decision making, they are indeed a new information management vehicle with a rich potential for public health and epidemiology.

Geographic information systems (GIS) are "automated systems for the capture, storage, retrieval, analysis, and display of spatial data" (1). Common to all GIS is a realization that spatial data are unique because their records can be linked to a geographic map. The component parts of a GIS include not just a database, but also spatial or map information and some mechanism to link them together. GIS has also been described as the technology side of a new discipline, geographic information science (2), which in turn is defined as "research on the generic issues that surround the use of GIS technology, impede its successful implementation, or emerge from an understanding of its potential capabilities." Recently, GIS has emerged as an innovative and important component of many projects in public health and epidemiology, and this disciplinary crossover is the focus of this review.

Few would argue that GIS has little to offer the health sciences. On the other hand, like other new technologies, GIS involves concepts and analytic techniques that can appear confusing and can lead to misunderstanding or even overselling of the technology. In this article, we attempt to bridge the gaps between the principles of geographic information science, the technology of GIS, the discipline of geography, and the health sciences. Our intent is to introduce to the epidemiologist a set of methods that challenge the "visual" half of the scientist's brain.

Address for correspondence: Keith C. Clarke, Ph.D., Department of Geology and Geography, Hunter College-CUNY, 695 Park Avenue, New York, NY 10021, USA.;fax: 212-772-5268; e-mail: kclarke@everest.hunter.cuny.edu.

Computers were first applied to geography as analytical and display tools during the 1960s (3). GIS emerged as a multidisciplinary field during the 1970s. The discipline's heritage lies in cartography's mathematical roots: in urban planning's map overlay methods for selecting regions and locations based on multiple factors (4); in the impact of the quantitative revolution on the discipline of geography; and in database management developments in computer science.

Several factors combined in the 1970s to reinforce GIS development. First, computers became more accessible and less costly. Second, mainframe computers gave way to minicomputers and then workstations, which gave great power to the user and included the access to networks that has led to its own revolution in technology. Third, the types of user interface required to operate technical software changed from batch, command-line, and remote access to windowing systems and "point and click" graphic interaction. What had been expensive, slow, and difficult has rapidly become inexpensive, fast, and easy to use. A final but essential precondition to GIS development was the broad availability of public domain digital map data, in the form of maps of the landscape from the U.S. Geologic Survey and for census areas from the U.S. Census Bureau. The current GIS World Sourcebook (5) lists hundreds of system suppliers and sources of information and catalogs system capabilities. In short, GIS has now come of age, to the extent that the contributions of a growing number of parallel disciplines have both influenced and been influenced by GIS. Other disciplines now affecting GIS include forestry, transportation planning, emergency services delivery, natural

hazards planning, marketing, archeology, surveying, and criminal justice. A wide array of capabilities and information awaits the health scientist ready to pursue an interest in GIS.

In this article, we consider the functional capabilities of GIS and how they can relate to epidemiology. We then review studies in epidemiology and health science where GIS has already made a contribution and introduce the technologic and analytic background. We review spatial analytic methods and concepts of use in epidemiology and conclude by examining what the near future holds for technologic changes and what these changes mean for the study of emerging infectious diseases and other health applications.

GIS Functional Capabilities

GIS definitions usually focus on what tasks a GIS can do rather that what it is. GIS functional capabilities follow the standard GIS definitions; therefore, GIS can bring together the elements necessary for problem solving and analysis.

Data capture implies that 1) data can be input into the GIS from existing external digital sources; this is particularly the case when no data exist for a project, and the base data must be assembled from other studies, public domain datasets, and images. This usually means that GIS must be able to import the most common data formats both for image-type (raster) and line-type (vector) maps. 2) GIS can capture new map data directly; this means either that the user can scan the map and input it into the GIS or trace over a map's features using a digitizing tablet and enter them into the GIS map database. 3) The GIS can accomplish everything that a regular database system can, such as enter and edit data and update information in the existing database.

Data storage implies storage of both map and attribute data. Attribute data are usually stored in a relational database management system contained within the GIS and accessed by a spreadsheet or query-driven user interface. For storage, map data must be encoded into a set of numbers so that the geometry of the map is available for query, but also so that the map is stored digitally in one or more files. Image maps are usually stored as gridded arrays. Line maps are encoded by any one of several systems, but usually by using both the coordinate information and encoded topology, so that the relationships between points, lines, and areas, such as the adjacency of regions or the connectedness of lines, are known in advance. The more efficient and flexible these data formats or structures, the more operations can be performed on the map data without further processing.

Data records in GIS can be retrieved in one of two ways. The relational database manager allows searching, reordering, and selecting on the basis of a feature's attributes and their values. For example, the user may wish to select out and order alphabetically the names of all health clinics that had positive results in more than 10% of their tests. GIS also allows spatial retrieval. The user could select all clinics by region, by their latitude, or by their distance from the capital. The user could also select all clinics that are more than 10 km from a major road and within 100 m of a river or lake. In addition, combining searches is possible. There could be several data "layers," for example vegetation, rivers, transportation, and population of villages. A single retrieval could combine data from each of these layers in a single query. Layers can also be weighted, so that rivers, for example, are twice as important as roads in selecting villages with a population under 500 surrounded by forest.

Display functions include predominantly the making of maps. Tools must exist for constructing many types of maps, such as contours, symbols, shading or choropleth, and sized symbols. Formal map display often follows a series of more temporary map images, usually without a strict map composition, and the result of a test, an analysis, or a query. In addition, the GIS must be able to output finished format of maps to a medium, such as PostScript, on a plotter or printer, or onto photographic film.

Many tools exist to support field data collection. Tasks in which ancillary demographic information needs to be input and coregistered are simple. Habitat associated with a vector (e.g., a snail or a mosquito) may need remotely sensed data, such as vegetation cover or weather data. If these data are georegistered, integration is possible. One of the most useful functions is called address matching, in which street addresses with house numbers and street names are automatically placed into an administrative unit or placed as a dot on the map. Thus a digital phone list or mailing list of patients can be merged with the remainder of the data. In the United States, the Census Bureau's TIGER files can usually match 70% to 80% of unedited address records, and higher percentages if the address files are proofed and/or the more detailed

and up-to-date commercial street files are used. In some field projects, the GIS's ability to make maps became the mainstay of the effort, allowing planning of truck and jeep routes, sequencing field clinics for optimal routes for visits, and even for local navigation. The ability to display maps often goes far beyond their final or use in the laboratory. Often a GIS image map is more accurate and up to date than anything available locally.

Existing Applications of GIS in Epidemiology

Epidemiologists have traditionally used maps when analyzing associations between location, environment, and disease (6). GIS is particularly well suited for studying these associations because of its spatial analysis and display capabilities. Recently GIS has been used in the surveillance and monitoring of vector-borne diseases (7-9) water borne diseases (10), in environmental health (11-13), modeling exposure to electromagnetic fields (14), quantifying lead hazards in a neighborhood (15), predicting child pedestrian injuries (12), and the analysis of disease policy and planning (16).

In a recent study in Baltimore County, Maryland, GIS and epidemiologic methods were combined to identify and locate environmental risk factors associated with Lyme disease (7). Ecologic data such as watershed, land use, soil type, geology, and forest distribution were collected at the residences of Lyme disease patients and compared with data collected at a randomly selected set of addresses. A risk model was generated combining both GIS and logistic regression analysis to locate areas where Lyme disease is most likely to occur.

GIS allows analysis of data generated by global positioning systems (GPS). Combined with data from surveillance and management activities, GIS and GPS provide a powerful tool for the analysis and display of areas of high disease prevalence and the monitoring of ongoing control efforts. The marrying of GIS and GPS enhances the quality of spatial and nonspatial data for analysis and decision making by providing an integrated approach to disease control and surveillance at the local, regional, and/or national level.

GIS is being used to identify locations of high prevalence and monitor intervention and control programs in areas of Guatemala for onchocerciasis (9) and in Africa for trypanosomiasis (17). Spatial and ecologic data are combined with epidemiologic data to enable analysis of variables that play important roles in disease transmission. This integration of data is essential for health policy planning, decision making, and ongoing surveillance efforts. For example, as part of the guinea worm eradication effort, the United Nation's Children's Emergency Fund placed pumps in villages most infected with the disease to ensure access to a safe water supply (18). GIS enabled researchers to locate high prevalence areas and populations at risk, identify areas in need of resources, and make decisions on resource allocation (16). Epidemiologic data showed a marked reduction in prevalence in villages where pumps were introduced.

GIS was used in designing a national surveillance system for the monitoring and control of malaria in Israel (19). The system included data on the locations of breeding sites of *Anopheles* mosquitoes, imported malaria cases, and population centers. The GIS-based surveillance system provided means for administrative collaboration and a network to mobilize localities in the case of outbreaks.

In 1985, the National Aeronautics and Space Administration (NASA) established the Global Monitoring and Disease Prediction Program at Ames Research Center in response to the World Health Organization's call for the development of innovative solutions to malaria surveillance and control (20). A major aspect of the program was to identify environmental factors that affect the patterns of disease risk and transmission. The overall goal of the program was to develop predictive models of vector population dynamics and disease transmission risk using remotely sensed data and GIS technologies.

Remotely sensed data have been used in many vector disease studies (8,17,21-24). Remote sensing and GIS were used to identify villages at high risk for malaria transmission in the southern area of Chiapas, Mexico (8). An earth environmental analysis system for responding to fascioliasis on Red River Basin farms in Louisiana was developed by integrating LANDSAT MSS imagery with GIS (22). In Kwara State, Nigeria, a temporal analysis of Landsat Thematic Mapper (TM) satellite data was used to test the significance of the guinea worm eradication program based on changes in agricultural production (21).

Spatial Analysis and GIS

GIS applications show the power and potential of such systems for addressing important health issues at the international, national, and local levels. Much of that power stems from the systems' spatial analysis capabilities, which allow users to

examine and display health data in new and highly effective ways. Spatial analysis refers to the "ability to manipulate spatial data into different forms and extract additional meaning as a result" (25). It encompasses the many methods and procedures, developed in geography, statistics, and other disciplines, for analyzing and relating spatial information. Spatial relationships, those based on proximity and relative location, form the core of spatial analysis.

Gatrell and Bailey (26) describe three general types of spatial analysis tasks: visualization, exploratory data analysis, and model building. These range in complexity from simple map overlay operations to statistical models such as spatial interaction and diffusion models. The value of maps for public health analysis has long been recognized; John Snow's now classic maps of cholera cases in relation to the Broad Street pump are a good example. However, with its extensive data management and display capabilities, GIS offers much more than simple mapping. Map overlay operations allow the analyst to compute new values for locations based on multiple attributes or data "layers" and to identify and display locations that meet specific criteria (27). For example, in targeting locations for mosquito vector control, one might want to identify areas that have low elevation, specific types of vegetation favored by mosquitoes, and are within 100 m of ponds or other water bodies. Each of these attributes comprises a distinct data layer. With GIS, one can create 100-m buffers around water bodies and then select areas meeting all three criteria. Display of these areas on a GIS-generated map has obvious benefits for planning vector control strategies.

As indicated previously, this general class of procedures for weighing and overlaying maps, also known as "suitability analysis," has been used in diverse health applications. Typically the criteria and weights attached to them are specified by the analyst based on expert knowledge or prior research. Using the computational and visual display capabilities of GIS, one can then explore the sensitivity of results to the weights and cutoff values used. Another approach is to employ regression analysis to generate the linear combination of factors that best explain spatial variation in disease prevalence. The weights from the regression model are used to create a composite index of risk which can then be mapped (7).

Visualization is also an important tool for showing the change in disease patterns over time. Animation, embedded within a GIS, is highly effective in depicting the spread or retreat of disease over space and time. A series of animated maps were created to show the advance of the AIDS epidemic in the United States as it moved from and within major cities (28). One could imagine a similar animated map sequence showing the retreat and eventual eradication of a disease like smallpox. Clearly much more research is needed in this area, especially research that links animation to theoretical models of disease diffusion, within a GIS environment.

Visualization can be used in novel ways to explore the results of traditional statistical analysis. Displaying the locations of outlier and influential values on maps and showing variation in values over space can add a great deal to epidemiologic research. Although such tools are being developed and explored, they would benefit greatly from a closer and more seamless link between statistical packages and GIS (25).

The second general class of GIS methods addresses exploratory spatial analysis. These methods allow the analyst to sift meaningfully through spatial data, identify "unusual" spatial patterns, and formulate hypotheses to guide future research (26). The quantity and diversity of spatial data in GIS can be overwhelming: exploratory methods help the analyst make sense of data and address "what if" questions. Advances in computing and graphics technology have made this one of the most active areas in GIS/spatial analysis research.

Among the most important exploratory methods for epidemiology and public health are methods for identifying space-time clusters or "hot spots" of disease. Openshaw's geographic analysis machine (GAM) was an early method that worked completely within a hybrid GIS. The GAM's many applications included an attempt to determine if spatial clusters of childhood leukemia were located near nuclear facilities in Britain (29). The GAM works with point data on disease cases and searches at regular intervals for statistically significant clusters of disease prevalence. Maps display the locations of significant clusters, showing the proximity of clusters to hypothesized environmental threats such as nuclear facilities. Although Openshaw's work was widely criticized on statistical grounds, it opened the door for an active body of research on exploratory spatial analysis of disease. Some of the new methods that have been developed as outgrowths of Openshaw's approach have been published (30).

Exploratory methods are also valuable in searching for zones or districts of high disease prevalence. Because areas may differ greatly in population size, prevalence rates have different levels of variability and thus reliability (31). Researchers have long used probability mapping to show the statistical significance of prevalence rates (32); however, probability mapping does not give a sense of the actual rates or the populations on which they are based. An alternative method is to smooth rates towards a regional or local mean value using empirical Bayes methods (33). Although GIS and empirical Bayes methods have developed separately, there is much scope for interaction. For example, GIS can be used to generate geographically based regional or local means to which actual rates are smoothed. These might be based on averaging rates for contiguous areas (33,34); or they might rely on more complex, multivariate, spatial clustering procedures that incorporate proximity as well as population attributes.

Many methods for exploratory analysis of disease patterns are not appropriate for infectious diseases because the methods are essentially static and assume independence. For infectious diseases, cases clearly are not independent and the diseases move through time and space. In these situations, one can use spatial autocorrelation methods and space-time correlograms to explore the spatial and temporal patterns of infectious disease spread (35).

These methods provide a general sense of the speed and geographic pattern of disease transmission. Although the methods have not typically been incorporated in GIS, there is great potential for doing so, especially with recent advances in computer animation.

Modeling, the final class of spatial analysis methods, includes procedures for testing hypotheses about the causes of disease and the nature and processes of disease transmission. In general, modeling involves the integration of GIS with standard statistical and epidemiologic methods. GIS can assist in generating data for input to epidemiologic models, displaying the results of statistical analysis, and modeling processes that occur over space. The first two points are evident in recent, regression-based analyses of disease risk, such as the study of Lyme disease (7). There GIS was used not only to integrate diverse datasets and calculate new variables, such as slope and distance from forest, but also to map geographic variation in disease risk, as predicted from a logistic regression model.

Other GIS models are more explicitly spatial, expressing relationships or flows between people and places. Spatial interaction and spatial diffusion models are of particular relevance to the study of emerging diseases. Spatial interaction models analyze and predict the movements of people, information, and goods from place to place (36). The flows of people between rural areas, villages, cities, and countries are all forms of spatial interaction that are central to disease transmission. By accurately modeling these flows, it is possible to identify areas most at risk for disease transmission and thus target intervention efforts. Spatial interaction models reflect two general principles: that interaction decreases with distance and increases with population size or "attractiveness." Given actual flow data, one can estimate values that show the effects of distance and population size (or other "attractiveness" factors) on interaction. The models can then be used to predict spatial interaction patterns elsewhere. Although spatial interaction models and GIS developed separately, some GIS now have spatial interaction modeling capabilities (37).

Spatial diffusion models analyze and predict the spread of phenomena over space and time and have been widely used in understanding spatial diffusion of disease (38). Such models are quite similar to spatial interaction models except that they have an explicit temporal dimension. By incorporating time and space, along with basic epidemiologic concepts, the models can predict how diseases spread, spatially and temporally, from infected to susceptible people in an area (39) and aid in understanding the emergence of infectious disease (40).

Data

Important technical and logistic innovations in data and data access for GIS are under way and will come to fruition before the end of the century. First, and by far the most important, have been increased access to the Defense Department's global positioning systems (GPS), the availability of inexpensive hand-held devices for using the system, and the addition of direct-to-GIS data links to these systems. For a relatively modest investment, field users can add geographic coordinates to their data collection from anywhere in the world, at any time, and in any weather. These systems are so flexible that their antennas can be

placed on top of a car, and the logger can be connected to a portable computer on the dashboard, so that as the user drives along, the path of the vehicle is permanently recorded in the GIS's own data format and displayed on screen with a 1-s update. As these systems have become more common, they have also gained in precision and accuracy. It is not uncommon for fixes to be corrected using a process known as differential GPS, either after the fact by computer software or in real time, so that each point is recorded to the nearest meter on the ground. GPS and GIS together have permanently altered the relationship between field data collection and data analysis. Data collected in real time can be analyzed the same day and acted upon immediately.

Similarly, various devices used for capturing overhead images and photographs have undergone a similar revolution. First, technology has improved, allowing images in the infrared, thermal, radar and other wavelengths to be collected at higher and higher spatial resolutions. Second, massive changes in policy have resulted from the end of the Cold War. Formerly secret satellite data, such as the CORONA and Russian spy imagery, are now broadly available, even searchable on the Internet. In the United States, the National Air Photo program intends to remap the country every 5 years at a scale of 1:12,000 with 1-m resolution and publish the images as CD-ROMs. In addition, NASA's largest ever Mission to Planet Earth and its Earth Observation System will begin to return unimaginable amounts of information about the whole earth's geography and atmosphere well before the end of the century. The data will be available to any Internet user and distributed by a set of active archive centers.

Third, technical issues related to data transfer have been partially eliminated. This has come about by the convergence toward sets of industry standard formats such as GIF and TIF for images and new national and international digital map data standards. In addition, efforts are now under way to standardize reference information about datasets, termed metadata, so that the equivalent of a Library of Congress cataloging will be possible.

Finally, many datasets have become available that can form at least the skeleton of a new GIS project almost anywhere in the world. By combining public domain datasets, such as the Digital Chart of the World and satellite imagery, with GPS and field data, the claim that data collection and changes in format constitute 80% of the effort in a GIS project is rapidly being eroded and replaced by a mere morning spent surfing the Internet. Nevertheless, many of the world's nations are still poorly mapped at the more detailed spatial scales required for local analysis.

Hardware

GIS hardware has continued to improve. On the high end, workstations have both increased in power and dropped in price, making this platform the choice for large, laboratory-based GIS projects. As the GIS software packages have been modified for the workstation operating systems, most commonly UNIX with X-Windows, operations that were impossible because of computational complexity have now become commonplace. This trend will continue to the extent that few technical constraints like memory and central processing unit (CPU) power will exist for GIS. Some tasks, such as skilled visual image identification and interpretation, have been partly or wholly automated. On the low end, microcomputers have become immensely powerful and fast, easily capable of performing basic GIS operations even on portable computers. The theme of GIS mobility, added to satellite and cellular telephone communications, has permanently transformed the ability to operate with GIS in the field, and will lead to a new "data rich" era for epidemiologic study.

In addition, the next generation of systems will depend on network computing. Networks have allowed de facto parallel computing within a local area network. By supporting personal multitasking, they have allowed data to be held in a distributed way and retrieved for use on demand, and the network has built an immensely powerful support structure for information sharing. The World-Wide Web, for example, can deliver to a workstation-user free GIS software, data, and information on how to install and use the system, support for technical problems, and even an outlet to publish scientific results.

Software

GIS software has improved remarkably in the latest generation and will undergo still more changes. The basic tools of the computer programmer have undergone a transition from first generation to object-oriented database and programming languages, offering some benefits in program module reusability, improved data

handling, and ease of use as more and more packages are rewritten to take advantage of these tools. The WIMP (windows, icons, menus, and pointers) interfaces so common today owe their origins to this technology. Today, the GIS research community suggests that as the "desktop metaphor" becomes more commonly accepted, increasingly sophisticated metaphors will take over for organizing computing, including perhaps using maps themselves to manage the computer rather than vice versa.

Some changes are far more practical but still of great value. Most software systems now support context-sensitive help, electronic manuals, and automatic installation and update procedures. Each of these could benefit from intelligent software that uses an expert system base and continues to tailor the system around the GIS operator's revealed use. Such software, used over a network, has been termed an intelligent agent. Most GIS of the future will use these methods to seek out new data over the network that relate to your problem, alert you to mistakes in your data management and analysis, and perhaps automatically compose maps and reports at the completion of a project.

Multimedia and hypermedia are also rapidly becoming a component of GIS software. Multimedia allow simultaneous use of text, sound, animation, and graphics. GIS software has also developed the ability to interact in many spoken languages, under different operating systems, and on many different computers. The independence of the software and the tasks from particular computer platforms, or even vendors, are a highly desirable element in a distributed system.

GIS and Public Health

While it holds distinct promise as a tool in the fight against emerging infectious diseases and other public health problems; it is not simply the next widget to come into play. GIS can be seen as a new approach to science, one with a history and heritage, a finite and well researched suite of methods and techniques, and a research agenda of its own. It does not fit neatly into the health scientist's toolbox. It requires rethinking and reorganizing the way that data are collected, used, and displayed. It requires expense, training, and a climb up a learning curve. It needs maintenance and support and can be both overwhelming and threatening to the uninitiated.

On the other hand, the base of research and scholarship using GIS in the health sciences cannot be ignored. A first step would be to integrate instruction on GIS into college curricula in public health. An admirable body of experience in GIS education already exists, even a thoroughly tested national curriculum that can be easily adapted to a new set of demands (41). A second step would be to seek out more formal links between the research communities working with GIS. There are astonishing similarities for example in the field requirements for using GIS between forestry, ecology, archeology and epidemiology that could provide substantial benefits by the sharing of experiences and the pooling of resources.

Above all, GIS should be seen as improving the set of tools to promote public health. Good epidemiologic science and good geographic information science go hand in hand. The future of GIS has already retained a role for the geographically literate public health expert. Epidemiologists should seize the opportunity to set their own agenda and influence the technology and science toward the goal of public health.

Dr. Clarke is professor and chair, Department of Geology and Geography, Hunter College, and on the faculty of the Earth and Environmental Sciences Program at the Graduate School and University Center of the City University of New York. Dr. Clarke's most recent research has been on environmental simulation modeling, the impact of the Persian Gulf War on the technology of cartography, and mapping to support disease control programs in Africa.

References

1. Clarke KC. Analytical and computer cartography. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall. 1995.
2. Goodchild MF. Geographical information science. International Journal of Geographical Information Systems 1992; 6(1).
3. Tobler WR. Automation and cartography. Geographical Review 1959;49:526-34.
4. Steinitz C, Parker P, Jordan L. Hand-drawn overlays: their history and prospective use. Landscape Architecture 1976;66:444-55.
5. Geographic Information Systems. GIS world sourcebook. Fort Collins, CO: GIS World, Inc., 1995.
6. Gesler W. The uses of spatial analysis in medical geography: a review. Soc Sci Med 1986;23:963-73.
7. Glass GE, Schwartz BS, Morgan JM III, Johnson DT, Noy PM, Israel E. Environmental risk factors for Lyme disease identified with geographic information systems. Am J Public Health 1995;85:944-8.

8. Beck LR, Rodrigues MH, Dister SW, Rodrigues AD, Rejmankova E, Ulloa A, et al. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. Am J Trop Med Hyg 1994;51:271-80.
9. Richards FO, Jr. Use of geographic information systems in control programs for onchocerciasis in Guatemala. Bull Pan Am Health Organ 1993;27:52-5.
10. Clarke KC, Osleeb JR, Sherry JM, Meert JP, Larsson RW. The use of remote sensing and geographic information systems in UNICEF's dracunculiasis (Guinea worm) eradication effort. Prev Vet Med 1991;11:229-35.
11. Cuthe WG, Tucker RK, Murphy EA, England R, Stevenson E, Luckardt JC. Reassessment of lead exposure in New Jersey using GIS technology. Environ Res 1992;59:318-25.
12. Braddock M, Lapidus G, Cromley E, Cromley R, Burke G, Branco L. Using a geographic information system to understand child pedestrian injury. Am J Public Health 1994;84:1158-61.
13. Barnes S, Peck A. Mapping the future of health care: GIS applications in Health care analysis. Geographic Information systems 1994;4:31-3.
14. Wartenberg D, Greenberg M, Lathrop R. Identification and characterization of populations living near high-voltage transmission lines: a pilot study. Environ Health Perspect 1993;101:626-32.
15. Wartenberg D. Screening for lead exposure using a geographic information system. Environ Res 1992 Dec;59:310-7.
16. Tempalski BJ. The case of Guinea worm: GIS as a tool for the analysis of disease control policy. Geographic Information Systems 1994;4:32-8.
17. Roger DJ, Williams BG. Monitoring trypanosomiasis in space and time. Parasitology 1993; 106(Suppl):277-92.
18. World Health Organization. Dracunculiasis: global surveillance summary, 1989. WHO Bull 1990;68:797-8.
19. Kitron U, Pener H, Costin C, Orshan L, Greenberg Z, Shalom U. Geographic information system in malaria surveillance: mosquito breeding and imported cases in Israel, 1992. Am J Trop Med Hyg 1994;50:550-6.
20. Wood BL, Beck LR, Dister SW, Spanner MA. Global monitoring and disease prediction program. Submitted January 1994. Sistema Terra.
21. Ahearn SC, De Rooy C. Monitoring the effects of dracunculiasis remediation of agricultural productivity using satellite data. Accepted for publication 1996. International Journal of Remote Sensing.
22. Malon JB, Fehler DP, Loyacano AF, Zukowski SH. Use of LANDSAT MSS imagery and soil type in a geographic information system to assess site-specific risk of fascioliasis on Red River Basin farms in Louisiana. Ann NY Acad Sci 1992;652:389-97.
23. Washino RK, Wood BJ. Application of remote sensing to arthropod vector surveillance and control. Am J trop Med Hyg 1994;50(6 Suppl):134-44.
24. Zukowski SH, Wilkerson GW, Malone JB, Jr. Fasciolosis in cattle in Louisiana. II. Development of a system to use soil maps in a geographic information system to estimate disease risk on Louisiana coastal marsh rangeland. Vet Parasitol 1993;47:51-65.
25. Bailey, T. A review of statistical spatial analysis in geographical information systems. In: Fotheringham S, Rogerson P. Spatial analysis and GIS. London: Taylor and Francis. 1994.
26. Gatrell A, Bailey T. Can GIS be made to sing and dance to an epidemiological tune? Presented at the International Symposium on Computer Mapping and Environmental Health, Tampa, FL, February 1995.
27. Tomlin WR. Geographic information systems and cartographic modelling. Englewood Cliffs, NJ: Prentice-Hall, 1990.
28. Gould P. The slow plague: a geography of the AIDS epidemic. Cambridge, UK: Blackwell, 1993.
29. Openshaw S, Charlton M, Wymer C, Craft A. A mark 1 geographical analysis machine for the automated analysis of point data sets. International J Geographical Information Systems 1987;1:335-58.
30. Marshall R. A review of methods for the statistical analysis of spatial patterns of disease. J R Stat Soc 1991;154:421-41.
31. Cressie N. Smoothing regional maps using empirical Bayes predictors. Geographical Analysis 1992;24:75-95.
32. Choynowski M. Maps based upon probabilities. J Am Stat Assoc 1959;54:385-8.
33. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43:671-87.
34. Ord K, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. Geographical Analysis 1995;24:286-306.
35. Cliff A, Haggett P. Atlas of disease distributions: analytic approaches to epidemiological data. Oxford; UK: Blackwell Reference, 1988.
36. Haynes K, Fotheringham AS. Gravity and spatial interaction models. Beverly Hills, CA: Sage, 1984.
37. Ding Y, Fotheringham AS. The integration of spatial analysis and GIS. Computers in Environmental and Urban Systems 1992:3-19.
38. Cliff A, Haggett P, Smallman-Raynor, M. Measles: an historical geography of major human viral disease from global expansion to local retreat. Oxford, UK: Blackwell Reference, 1993.
39. Thomas R. Geomedical systems: intervention and control. New York: Routledge, 1990.
40. Haggett P. Geographical aspects of the emergence of infectious diseases. Geografiska Annaler, B 1994;76:91-104.
41. National Center for Geographic Information and Analysis. NCGIA Core Curriculum in GIS, 1990. URL: http://www.ncgia. uscb.edu/pubs/core.html.

The Evolution and Maintenance of Virulence in Microparasites

Bruce R. Levin, Ph.D.
Emory University, Atlanta, Georgia, USA

In recent years, population and evolutionary biologists have questioned the traditional view that parasite-mediated morbidity and mortality—virulence—is a primitive character and an artifact of recent associations between parasites and their hosts. A number of hypotheses have been proposed that favor virulence and suggest that it will be maintained by natural selection. According to some of these hypotheses, the pathogenicity of HIV, *Vibrio cholerae*, *Mycobacterium tuberculosis*, the *Shigella*, as well as *Plasmodium falciparum*, and many other microparasites, are not only maintained by natural selection, but their virulence increases or decreases as an evolutionary response to changes in environmental conditions or the density and/or behavior of the human population. Other hypotheses propose that the virulence of microparasites is not directly favored by natural selection; rather, microparasite-mediated morbidity and mortality are either coincidental to parasite-expressed characters (virulence determinants that evolved for other functions) or the product of short-sighted evolution in infected hosts. These hypotheses for the evolution and maintenance of microparasite virulence are critically reviewed, and suggestions are made for testing them experimentally.

How much of the emergence and reemergence of infectious diseases is due to evolution, rather than ecological, technical, and social change (1)? Under what conditions will attenuated vaccine organisms become virulent? Are hospitalized and immunocompromised hosts reservoirs for the evolution of virulent pathogens (2)? The answers to these and related questions require an understanding of the ecological conditions and genetic processes responsible for the evolution and maintenance of parasite-mediated morbidity and mortality in infected hosts—virulence, as we shall define it here.

At least since Darwin's time (3), evolutionary biologists have been interested in infectious diseases, but primarily with respect to the role of these diseases in the adaptation and evolution of humans and other species (4). A bit more than 15 years ago, this interest in infectious disease took a new turn, a focus on the microbes responsible for these diseases and the evolution and maintenance of their virulence. Here I offer a relatively brief and personal review of current theories of the evolution and maintenance of virulence in the bacteria, viruses, protozoa, and single cell fungi, "microparasites" (to use the term employed by population biologists), responsible for infectious diseases. I consider how these theories fit, what is known about the epidemiology of microparasite infections and the mechanisms of pathogenesis, and discuss procedures to test hypotheses derived from these theoretical considerations of the population biology and evolution of microparasites. For other recent reviews of this subject, see (5-8).

Address for correspondence: Bruce R. Levin, Ph.D., Emory University, EcLF, 1510 Clifton Road, N.E., Atlanta, GA 30322, USA; fax: 404-727-2880; e-mail: biobrl@emuvm1.cc.emory.edu.

The Conventional Wisdom

At one time, virulence was almost universally considered an artifact of recent associations between parasites and their hosts (9, 10), and to a fair extent, it still is (11). In accord with this view, which Bob May and Roy Anderson called "conventional wisdom" (12), parasite-host coevolution is necessarily in the direction of commensalism or, nicer yet, mutualism. The logic behind this view is pleasing to human sensibilities. A fully evolved parasite would not harm the host it needs for its survival, proliferation, and transmission. Indeed, the appeal of this view of nature of parasite-host coevolution was sufficient for its corollary to also be assumed valid. That is, pathogenesis is often taken as evidence of recent associations between parasites and their hosts.

Many observations are consistent with conventional wisdom about parasite-host coevolution. This is particularly so for most of the so-called

emerging diseases. For example, Legionnaires' disease, Lyme disease, and pneumonia caused by hantavirus are consequences of human infection by parasites and/or commensals of other species, rather than by organisms that have had a long association with humans. In fact, for these emerging diseases and some older microparasitic diseases, like Rocky Mountain spotted fever, anthrax, and rabies, humans play no (or at best a negligible) role in the transmission of the parasite and, in that sense, are an evolutionary dead end. While HIV is transmitted between humans, its association with our species is almost universally considered recent (13, 14).

Other observations can be interpreted as inconsistent with conventional wisdom. For some virulent pathogens, like *Shigella* and *Neisseria gonorrhoeae*, humans appear to be either the unique or the dominant host and vector for infectious transmission (15). For other lethal microparasitic diseases like malaria and tuberculosis (TB), there is evidence that these microparasites have had a long history in human populations and that humans play a major if not unique role in their infectious transmission. However, for the pathogens involved in both these diseases, animal origins have been implicated, and it is difficult to find clear evidence of their existence (or that of other extant pathogens) before the origins of agriculture (16-18).[1] One can always rescue conventional wisdom from these inconsistent observations by assuming that "long" is not long enough for these microparasites to evolve or coevolve with humans to a more amenable relationship. Then again, it may well be that some microparasites responsible for new infections in human hosts will evolve to become increasingly virulent human pathogens and be readily transmitted between human hosts.

Conventional wisdom is not based on hypotheses that can be readily tested and rejected. Microparasites that lead to the extinction of their only host face the same fate as the host. On the other hand, evolving and becoming gentle and prudent in treating their hosts (when natural selection operating at the level of individual microparasites favors profligate behavior like virulence) require some form of group-level or kin selection (8), and/or a host evolutionary response that unilaterally converts an otherwise virulent microparasite into a commensal. Conventional wisdom does not account for the actual mechanisms responsible for the evolution of benign associations between microparasites and their hosts.

Epidemiologic Models and the "Enlightenment"

In the early 1980s, at least among evolutionary biologists, conventional wisdom gave way to what, in an earlier consideration of this subject Catharina Svanborg and I satirically (but sympathetically) referred to as the "enlightenment" (24). In accord with this new view, natural selection could favor the evolution and maintenance of virulence as well as commensal and symbiotic associations between microparasites and their hosts. In other words, virulence could be the evolved as well as the primitive stage of these associations. The direction of natural selection in any given situation depends on the epidemiology and ecology of the microparasite and, in particular, the relationship between its virulence and its rate of infectious transmission in the host population.[2] This can be seen in the equation for the finite rate of increase of a directly transmitted microparasite in a wholly susceptible host population (12, 25, 26)

$$R_0 = \frac{ßN}{\alpha + b + \nu}$$

[1] The existence of genetic polymorphisms, like sickle cell, thalassemia, and glucose 6-phosphate dehydrogenase (G6PH) deficiency (19, 20), Duffy-negative blood groups (21), and specific HLA alleles (22) maintained by *Plasmodium*-mediated selection can also be interpreted as evidence for malaria's long association with humans. There is evidence for inherited resistance to TB among mammals (23), and arguments that TB epidemics have selected for inherited resistance in humans (16). However, that evidence is not as compelling as that for malaria.

[2] In at least the mathematical theory, the morbidity component of microparasite virulence is not treated explicitly (25). The symptoms and pain resulting from infection are implicitly incorporated in the rates of disease-associated mortality, recovery, and transmission. Moreover, while acknowledging the existence of microparasite-mediated selection and evolution in the host population, for the most part, the enlightened view of microparasite-host coevolution has concentrated on the changes in the microparasite population. The idea is that because of the relatively longer generation times, the rate of evolution in the host population is going to be low.

where ß is the rate constant of infectious transfer of the microparasite, N the density of the susceptible host population, α the rate of microparasite-induced mortality (virulence), b the rate of microparasite-independent mortality, and ν the rate of recovery. R_0 is the number of secondary infections caused by a single primary infection and serves as a measure of the fitness (here and elsewhere in a Darwinian sense) of the parasite in this naive host population. At any given host density, N, this measure of fitness of the parasite is directly proportional to its transmissibility, ß, and the term of its persistence in an infected host, the reciprocal of $\alpha + b + \nu$.

If the parameters of the R_0 equation were independent of each other, the predictions derived from this equation would be consistent with conventional wisdom: benign parasites would evolve. That is, natural selection would favor highly transmissible ($b \rightarrow \infty$), incurable ($\nu \rightarrow 0$), commensals ($\alpha \rightarrow 0$), or symbionts ($\alpha \rightarrow -\infty$). On the other hand, if transmission and virulence, the parameters ß and α in the R_0 equation, were positively coupled, natural selection could favor the evolution and maintenance of some level of virulence, $\alpha \rightarrow 0$, in the microparasite population.

In accord with the epidemiologic perspective implicit in the R_0 equation, an understanding of the evolution of virulence in microparasites comes down to elucidating the relationship between the rate at which the microparasite is transmitted between hosts and the rate of parasite-mediated mortality in individual infected hosts. If that relationship is positive, then some level of virulence may be favored. And, since the first statements of this new view of parasite-host coevolution (12, 26, 27), much of the research on the evolution of virulence has focused on the association between these two components of parasite fitness.

The most cited, and to me the single most compelling, evidence in support of this new interpretation of microparasite-host coevolution comes from the "experiments" using myxoma virus to control European rabbit populations in Australia and Europe (26, 28, 29). Within a relatively short time after the release of highly virulent myxoma, the viruses recovered from the then decimated and sometimes more resistant wild rabbit populations were less virulent and had lower rates of disease-induced mortality on control laboratory rabbits than those initially released. However, the extent to which myxoma virus from the wild became attenuated was substantially less than that which could be achieved experimentally (29). This was interpreted as evidence for a positive coupling between the rates of infectious transmission and rates of virus-induced mortality, a trade-off between virulence and transmission. Highly virulent forms of the virus had a disadvantage because they killed the rabbits too quickly and thus reduced the time available for them to be picked up by the insect (mosquito or flea) vectors required for their infectious transmission. Viruses that were too attenuated had a disadvantage because they generated fewer skin lesions and had lower densities of circulating virions, which presumably would reduce the rate at which they would be bitten by these insect vectors, the likelihood of biting vectors picking up myxoma, and the number of virions picked up at any given bite. Thus, in contrast to conventional wisdom and in accord with the enlightened interpretation, natural selection could favor and maintain the virulence of microparasites. This results when there is a positive coupling between a parasite's virulence and its capacity for infectious transmission.

The myxoma story is particularly compelling because the quantitative relationship between virulence and transmissibility inferred from the epidemiologic data and models was independently tested and demonstrated experimentally (30, 31). The myxoma story remains the only one for the microparasites of eukaryotic hosts where the predictions about transmission and virulence made from an interpretation of epidemiologic observations were tested experimentally. With few exceptions (32), inferences about the relationship between transmission and virulence and the trade-offs between these two attributes of a microparasite's association with its host have been derived from comparative evolution studies or retrospective interpretations of epidemiologic data. In some cases, these inferences are reasonably strong, e.g., in the study by Alan Herre (33) on fig wasps and a nematode parasite and by Deiter Ebert (34) on a planktonic crustacean with a protozoan parasite. The latter study is particularly convincing because it includes independent, experimental evidence of a positive correlation between the density of spores in infected hosts and the virulence and transmissibility of this protozoan parasite.

The enlightened view on the virulence of microparasites sometimes takes the positive association between the virulence of a microparasite and its transmissibility as axiomatic; therefore, it

assumes that a microparasite's virulence is constrained solely by the need to keep the host alive to facilitate its transmission to new hosts.[3] This is implicit in much of Paul Ewald's writing on this subject (6, 35) and is the basis of his main thesis that changes in rates of infectious transmission will select for microparasite strains or species with different levels of virulence.

By assuming a necessarily positive relationship between a microparasite's capacity for infectious transmission and the extent of morbidity and rate of mortality it causes in infected hosts, a positive "trade-off" (relationship) between transmissibility and virulence, Paul Ewald has generated scenarios for the evolution of virulence and changes in virulence for a diverse array of microparasites, including those responsible for cholera, influenza, dysentery, and AIDS (6, 35). While the details of Ewald's stories may differ, the plot is almost always the same: increases in the rates of transmission favor increases in virulence, and the reverse. For example, Ewald has postulated that the virulence of HIV observed in contemporary human populations, AIDS, is in large part due to evolution in this retrovirus responding to the increases in human-human transmission rate resulting from more promiscuous sexual behavior.[4] However, even when a direct relationship between the virulence and transmission rate of HIV is assumed, a deeper consideration of the epidemiology and course of this sexually transmitted disease shows that this simplistic conclusion about evolution and the virulence of HIV is chock full of caveats (36, 37). The relative contributions of transmission and virulence (as measured by the time before the onset of AIDS) to the fitness of HIV in the population of hosts depends on whether the disease is in an epidemic or endemic phase. Moreover, as I consider later, there are other, very different, hypotheses for the evolution of the virulence of this retrovirus and other pathogenic microbes that do not require the necessarily positive association between infectious transmission and virulence upon which Ewald has based his arguments for the evolution and maintenance of virulence in microparasites.

A corollary of the hypothesis of a positive trade-off between transmissibility and virulence is that if all else were equal, increases in the degree of vertical (e.g., from a mother to a fetus) transmission of a parasite, relative to its horizontal (infectious) transmission would favor reductions in its virulence (38). There is compelling, experimental evidence to support this corollary. However, the evidence is restricted to experiments with *E. coli* and its phage, *f1*, which can be transmitted vertically, in the course of cell division, or horizontally, by infecting susceptible, uninfected bacteria (39). While some, like me most of the time, may believe in the adage "what is true for *E. coli* is true for elephants, but only more so," other, less *coli*-centric souls, may want to see more experiments of this type with microparasites and vertebrate hosts. I certainly do.

Within-Host Population Dynamics and Virulence of Microparasites

There is a dearth of experimental investigations of the quantitative relationship between the transmission and virulence of microparasites. During the past few years, however, there has been a flurry of theoretical studies of the within-host population dynamics of microparasites that have specifically considered the relationship between the virulence and transmission rates of microparasites and their densities and/or rates of replication in infected hosts (40-44). In the simplest models developed in these theoretical studies of the within-host population dynamics of

---

[3] One way to experimentally augment the virulence of a microparasite, as, for example, measured by declines in its LD_{50}, is to artificially pass that microbe between hosts (15). From one perspective, this result is consistent with the trade-off hypothesis, as the effect of passage is to make the parasite's transmission independent of the host's survival, thereby allowing it to become more virulent without compromising its need to be transmitted to other hosts. However, increased virulence in a passage experiment is not sufficient evidence for that trade-off. (It may well be that the parasite's capacity for infectious transmission is impaired as a consequence of whatever increased its capacity for infectious transmission.) I know of no experiments that demonstrate that the increase in virulence generated during a passage experiment is also reflected as increased—transmissibility, as is necessary for the trade-off interpretation. Indeed, it may well be that an increase in the case-mortality rate or a reduction in the LD_{50} of a microparasite will be reflected as a reduction in its natural transmissibility.

[4] I quote: "Severe immunodeficiency could develop in an old association [between a sexually transmitted virus or SIV and its host] as a result of increases in sexual partner rates causing evolution of increased virulence" (p. 143, reference 6). "If rates of unprotected sexual contact decline, so should the virulence of HIV" (p. 144, reference 6).

microparasites, the virulence of the microparasite, as measured by either the rate at which it kills its host or its LD_{50}, is assumed to be directly proportional to its rate of proliferation in that host, and its rate of infectious transmission is directly proportional to its within-host density (41). Under these conditions, in the absence of superinfection or mutation, selection favors microparasites with intermediate rates of within-host replication, i.e., intermediate levels of virulence. More complex situations, like the coexistence of microparasite lineages with different levels of virulence, result when virulence is proportional to the within-host growth rate of the parasite and single hosts can be infected with parasites of different growth rates (43) or when there are high rates of mutation to different levels of virulence within a host (45). Moreover, with superinfection and mutation, the theory developed in these two reports predicts that the average level of virulence of a parasite in an infected host can exceed that anticipated from models that do not allow for superinfection and/or assume that the parasite's level of virulence in an infected host remains invariant.

The Convergence of Theories

The predictions that can be made on the basis of the current view of the evolution of virulence differ from predictions that might follow conventional wisdom because the new view allows for natural selection in the parasite population to favor the evolution and maintenance of some level of virulence. Moreover, even when there is a positive association between a parasite's virulence and its transmissibility, under the conditions described in the following paragraph, the predictions of new methods can still converge with those of conventional wisdom.

If the density of the sensitive host population is regulated by the parasite, an extension of the enlightened theory predicts that natural selection in the microparasite population can lead to continuous declines in the level of virulence, possibly to immeasurable values (46). Although not stated in this general way, the same conclusion about declining virulence can be drawn from models of the epidemiology of HIV/AIDS (36, 37). During the epidemic phase of a microparasitic infection, when the host population is composed primarily of susceptible hosts, selection favors parasites with high transmission rates and thus high virulence. As the epidemic spreads, the proportion of infected and immune hosts increases and the density of susceptible hosts declines. As a result, the capacity for infectious transmission becomes progressively less important to the parasite's Darwinian fitness and persistence in the host population. Selection now favors less virulent parasites that take longer to kill their host and, for that reason, are maintained in the host population for more extensive periods. Analogous arguments have been made for the latent period of a bacteriophage infection (47), the evolution of lysogeny (48), the tradeoff between vertical and horizontal transmission (49, 50), and the advantages of microparasite latency in general (40).

Alternative Models for the Evolution of Microparasite Virulence

For any microparasite, the rate of transmission between hosts will always be a significant component of fitness, and, if all else is equal, parasites transmitted at higher rates in the host population have a selective advantage over less transmissible forms. On the other hand, there is no reason to assume that in general a microparasite's rate of infectious transmission will be positively associated with its virulence. Moreover, even when there is no relationship or a negative relationship between transmission and virulence, there are at least two ways by which natural selection can lead to the evolution and maintenance of virulence, coincidental evolution (24) and short-sighted within-host selection (51).

Coincidental Evolution

According to the coincidental evolution hypothesis, parasite-mediated morbidity and mortality are what Gould and Lewontin (52) likened to the spandrels of gothic churches. While these structural necessities may frame the frescos and paintings within, that is not the reason for their existence. They are architectural constraints. Analogously, the factors responsible for the virulence of a microparasite in an infected host may have evolved for some purpose other than to provide the parasite an advantage within a host or its transmission to other hosts.

It would be difficult to account for the evolution of botulism toxin by selection favoring *Clostridium botulinum* that kill people who eat improperly canned food. The same argument could be made for the toxins of *C. tetanae* and possibly for those produced by other free-living *Clostridia*. Although these organisms may proliferate in humans, they

are soil bacteria, and the effects of the toxin may not contribute to their capacity to colonize, proliferate, and be maintained in humans or to their capacity to be transmitted between human hosts. How many other microparasite-induced symptoms, and the resulting host morbidity and mortality, provide no advantage to that microbe in (or on) a host or its transmission between hosts? Did the lipopolysaccharides and other components of bacterial cell walls and cell membranes evolve because the fitness of bacteria expressing them is enhanced by "endotoxin"—induced overresponse of the immune system responsible for the morbidity and mortality of sepsis (53)? Do the toxins confer an advantage on *E. coli* O157 or *Staphylococcus aureus* (or the plasmids and phages that code for these toxins) because they produce, sometimes lethal, symptoms in infected hosts, hemolytic uremic and toxic shock syndromes, respectively? An earlier paper on this subject (24) argued that the adhesins produced by the *E. coli* responsible for the morbidity of symptomatic urinary tract infections evolved and are maintained to facilitate colonization of the gut. The painful symptoms of urinary tract infections generated by an inflammatory response to these adhesins may confer no advantage for the *E. coli* expressing them in the urinary tract and may in fact lead to the clearance of those bacteria (24).

Each of the symptom-inducing toxins and adhesins described above, as well as many other so-called "virulence determinants" (54) may indeed facilitate the microparasite's ability to colonize, proliferate, or be maintained in infected hosts, and/or be transmitted between hosts. This certainly sounds reasonable for many virulence determinants, e.g., the somatic cell invasiveness mechanisms of *Shigella*, the capsules of *Streptococcus*, the diarrhea-inducing toxins produced by *Vibrio cholerae*, and the sneezing and coughing induced by rhinoviruses. On the other hand, it is necessary to formally test this hypothesis that these symptoms have that effect and reject the alternative, that the morbidity and mortality generated by the expression of a specific virulence determinant provides neither a within- or between-host (infectious transmission) advantage to the parasite.

Short-Sighted Evolution

Natural selection is a local phenomenon. Characters that confer a survival or replication advantage on the individual organisms that express them at a given time or in a given habitat will be favored and evolve at that time and in that habitat. Whether the expression of those temporally or locally favored characters will increase or reduce the fitness of that organism at other times or in other habitats is irrelevant. Also irrelevant is whether a locally favored character makes the population better or less adapted to its environment at large or augments the likelihood of its survival in the future. This myopia is a fundamental premise of the theory of evolution by natural selection and the basis of the short-sighted evolution hypothesis for microparasite virulence (51).

Within an infected vertebrate host, microparasite populations go through many replication cycles and may achieve very high densities. They may also reside and proliferate in many different subhabitats (tissues and cells) and confront a variety of different and ever-changing constitutive and inducible host defenses which may, sequester, kill, or in other ways inhibit their proliferation. As a consequence of classic mutation, transposition, and recombination, genetic variability will be continually generated in the populations of infecting microbes. Mutant or recombinant microparasites that are better able to 1) avoid being done in or inhibited by the host's defenses; 2) proliferate in the host; or 3) invade and replicate in novel habitats, tissues, and cells where there is less competition from members of its species would have an advantage in that host. This would occur even when the expression of the characters responsible for that local advantage reduces likelihood of the transmission to other hosts. Stated another way, the morbidity or mortality caused by a microparasite infection could be the result of the within-host evolution that is short-sighted because that virulence actually reduces the rate at which that parasite is transmitted to other hosts.

Three examples of microparasite virulence that could be products of this mode of evolution can be considered (51). For two of these examples, bacterial meningitis and poliomyelitis, many human hosts are infected by the responsible microparasites, primarily *Haemophilus influenzae*, *Neisseria meningitidis*, and *Streptococcus pneumoniae* for meningitis and poliovirus for poliomyelitis, but very few manifest the symptoms of these infections. In the case of meningitis, the neurologically debilitating and sometimes fatal symptoms of the infection are a consequence of an inflammatory response against the bacteria entering and

proliferating in the cerebral spinal fluid. These meningitis-causing bacteria normally reside in the nasopharyngeal passages and are transmitted by droplet infection. The cerebrospinal fluid is, at least with respect to their infectious transmission, a dead end. On the other hand, bacteria capable of invading and proliferating in that habitat could have a local advantage as there are no other competing populations and only modest defenses. An analogous argument can be put forth for poliovirus. Symptomatic infections with this virus are caused by their invasion of and proliferation in the neurologic tissue of the central nervous system. Poliovirus normally replicates in the mucosal cells of the mouth, throat, and intestines and is transmitted by the oral-fecal route. Poliovirus virions proliferating in the central nervous system would almost certainly not be transmitted. The evidence in support of short-sighted evolution for the virulence of these specific microparasites is mostly circumstantial (51). On the other hand, short-sighted evolution for the virulence of specific microparasites is a hypothesis that can be tested. If the hypothesis is valid, the microparasites responsible for the symptoms would be genetically different from their ancestors that infected the host and better adapted for proliferation in the site of the symptoms than the ancestors themselves.

The third example of short-sighted evolution of virulence considered, HIV, is different from the other two in that virtually every human infected with this retrovirus that does not die of other causes, eventually manifests and succumbs to AIDS. However, although the case mortality of HIV infection may approach unity, as measured by the rate of mortality (deaths per unit of time), from an epidemiologic perspective, HIV is not a very virulent virus. There is substantial variation in the time between infection and the onset of AIDS. On average in industrialized countries, the term of this infection is 8 to 10 years (55). During the early phase of an HIV epidemic, most transmission of the virus occurs during the initial viremia, probably before seroconversion and certainly before the onset of AIDS (37, 56). It is not at all clear how the transmissibility of HIV virions during this early phase of the infection is related to the time of onset of AIDS. HIVs that are more transmissible early in the infection may lead to an earlier onset of AIDS. If this is the case and all else were equal, increasing opportunities for transmission during the epidemic phase would favor increases in HIV virulence (36, 37). However, there may be no association between HIV's capacity to be transmitted early and the time of onset of AIDS, or the time until the onset of AIDS may increase with the transmissibility of the virus during the early phase of the infection. Under either of these conditions, selection during the epidemic phase of the disease would favor more transmissible but less virulent HIVs.

In the course of HIV infection, the HIV population undergoes continuous genetic changes. In fact, in a number of hypotheses of HIV pathogenesis, AIDS is a consequence of mutation and selection in the HIV population that occurs during the course of the infection in individual hosts (57-60), i.e., short-sighted, within-host evolution. Albeit different in their details, all of these hypotheses are consistent with what is known about HIV infection, and all can account for the course of these infections and variable time of onset of AIDS.

Experimental Evolution Meets Experimental Epidemiology

Results of recent studies by population and evolutionary biologists predict at least three ways by which the virulence of microparasites can be favored and will be maintained by natural selection. 1) Direct selection: there is a positive relationship between the parasite's virulence and its rate of infectious transmission; 2) coincidental evolution: the parasite's virulence is due to character(s) favored and maintained by selection for some other function and the expression of those virulence determinants in an infected host does not confer a net advantage or disadvantage in the parasite population at large; and 3) short-sighted, within-host, evolution: the parasites responsible for the morbidity and mortality of an infection are selected for within the host because of a local advantage, and that evolution reduces the rate at which that locally adapted parasite is transmitted between hosts.

At this time, these predictions are based almost entirely on general theory and retrospective interpretations of epidemiologic and other observations about specific microparasites. Although this theory and these interpretations may be appealing, in a formal Popperian sense (61), almost all the mechanisms postulated for the evolution of virulence of specific microparasites are no more than untested hypotheses. However, unlike most evolutionary hypotheses, those about the evolution of microparasite virulence can be tested and rejected with prospective, experimental studies with

laboratory animal and plant hosts. These tests could be at two levels; first, tests of the validity of the assumptions behind these models of the evolution of virulence and second by tests of the predictions made from the consideration and analysis of these models.

For the direct selection hypothesis, it is essential to demonstrate a positive relationship between a microparasite's virulence and its rate of (or capacity for) infectious transmission. For mammalian hosts, protocols exist for determining this relationship (30-32). The object would be to estimate the densities of microbes at the sites of transmission (e.g., feces, nasal passages) during the entire course of the infection. Moreover, it would be useful to separately test the colonization ability and virulence of the microbes from these sites. According to the coincidental evolution hypothesis, it is possible that the virulence determinant responsible for morbidity and mortality in the host provides a local advantage to the parasite expressing it; whether it does or not could be tested with competition experiments between strains of that microparasite that are isogenic save for that virulence determinant. The genetic basis of many virulence determinants are known, and it should be possible to construct these strains. However, unlike in the direct selection hypothesis, in coincidental evolution, microbes expressing the virulence determinants should not be over represented at the sites of infectious transmission. Under the short-sighted evolution hypothesis, microbes isolated from the tissues and organs responsible for the symptoms of the infection (e.g., in the cerebrospinal fluid) should be better adapted for proliferation in those organs and tissues than the originally infecting strain from which they were derived. This could be tested with pairwise competition experiments between the original and potentially evolved strains injected at the site of the symptoms with a common, genetically marked competitor of that parasite. Here, too, it is necessary to demonstrate that the strain responsible for the symptoms is not overrepresented at the site of infectious transmission.

To test the prediction of the direct selection hypothesis and to exclude that mechanism in tests of the coincidental and short-sighted alternatives, it is necessary to study the epidemiology of the microparasites as well as their within-host properties. For bacteriophages and bacteria this is a relatively easy task, e.g., testing Abedon's hypothesis (47) about the direct relationship between the density of sensitive bacteria and selection for latent period length (a measure of virulence) and burst size (a measure of transmissibility). For eukaryotic hosts, this kind of study is going to be more difficult and, at this time, may not be possible. The basic protocols for experimental studies of the epidemiology of bacterial and viral infections of laboratory mice were developed and successfully employed a long time ago (32, 62).[5] However, experiments of these types are costly, labor-intensive, and time-consuming, and because of concerns about animal rights, it may be difficult to get permission to do these experiments with mice or other higher vertebrates. On the other hand, experiments of this type with insects and other invertebrate animal hosts as well as plants would be tenable and valuable as tests of the general theory, albeit less immediately relevant to the evolution and maintenance of virulence in human pathogens.

Acknowledgments

I thank Marc Lipsitch, Deiter Ebert, Jim Bull, David Thaler, and Tomoko Steen for reading this manuscript and for many insightful comments and suggestions, most of which I agreed with and a few of which I've incorporated. This endeavor was supported by a grant from the National Institutes of Health, GM33782.

Dr. Levin is professor of biology at Emory University and director of the Graduate Program in Population Biology Ecology and Evolution.

[5] Greenwood and colleagues (32), studied microparasites with different transmissibilities ("infectivity") and virulence.In one replica of their study of pasteurellosis (due to infections with a bacterium they call *Pasturella muriseptica* in experimental populations of mice, they report "the appearance of a variant that had gained infectivity and retained, or perhaps increased, its original virulence." However, in general, their study and Fenner's (62) provide little information about the direction of natural selection in these microparasite populations. With respect to evolutionary questions, these were wait-and-see experiments. Only one strain of microparasite was introduced into each population, and it was necessary to wait for mutations that changed their virulence or transmissibility. More information about the direction of selection and a better test of these evolutionary hypotheses could be obtained in these types of experiments if two or more genetically marked strains of microparasites with different virulence and transmissibility were introduced simultaneously and were allowed to compete.

He is a population and evolutionary biologist, who, like a number of others of his ilk, recently discovered infectious disease. Currently he and the postdoctoral fellows and students working with him are doing theoretical (mathematical modeling) and experimental research on the within-host population dynamics of bacterial infections and their treatment, and the epidemiology, population genetics and evolution of antibiotic resistance.

References

1. Schrag S, Wiener P. Emerging infectious diseases: what are the relative roles of ecology and evolution? Trends in Ecology and Evolution 1995; 10: 319-23.
2. Wallace B. Can "stepping stones" form stairways? American Naturalist 1989; 133: 578-79.
3. Darwin C. The descent of man and selection in relation to sex. New York: Random House, Modern Library, 1871 (reprinted 1960).
4. Haldane JBS. Disease and evolution. La Ricerca Scientifica 1949; 19:68-76.
5. Garnett GP, Antia R. Population biology of virus-host interactions. In: Morse SS, editor. The evolutionary biology of viruses. New York: Raven Press, 1994:51-73.
6. Ewald PW. The evolution of infectious disease. Oxford, UK: Oxford University Press, 1994.
7. Bull JJ. Virulence. Evolution 1994; 48:1423-37.
8. Frank SA. Models of parasite virulence. Q Rev Biol 1996;71:37-78.
9. Dubos R. Man adapting. New Haven, CT: Yale University Press, 1965.
10. Burnet FM, White DO. Natural history of infectious diseases. Cambridge, UK: Cambridge University Press, 1972.
11. Mims C, Dimmock N, Nash A, Stephen J. Mims' pathogenesis of infectious disease. 4th ed. San Francisco: Academic Press, 1995.
12. May RM, Anderson RM. Parasite host coevolution. In: Futuyama DJ, Slatkin M, editors. Coevolution. Sunderland, MA: Sinauer, 1983:186-206.
13. Essex M, Kanki PJ. The origin of the AIDS virus. Sci Am 1988; 259: 64-100.
14. Leigh Brown AJ. Holmes EC. Evolutionary biology of the human inmunodeficiency virus. Annual Review of Ecology and Systematics 1994; 25:127-62.
15. Davis BD, et al. Microbiology. 4th ed. Philadelphia: Lippincott, 1990.
16. Waters AP. Higgins DG, McCutchan TF. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proc Natl Acad Sci USA 1991; 88: 3140-4.
17. Allison MR, Mendoza O, Pezziam A. Documentation of a case of tuberculosis in pre-Columbian America. Am Rev Resp Dis 1973; 107: 985.
18. Bates JH, Stead WW. The history of tuberculosis as a global epidemic. Med Clin North Am 1993; 77: 1205-17.
19. Allison AC. Protection afforded by sickle cell trait against malarial infection. Br Med J 1954; 2:290-4.
20. Luzzatto L, Usanga EA, Shunmugam R. Glucose 6-phosphate dehydrogenase deficient red cells: resistance to infection with malarial parasites. Science 1969; 164: 839-41.
21. Miller LH, Mason SJ, David FC, McGinnis MH. The resistance factor to *Plasmodium vivax* in Blacks. N Engl J Med 1976; 295:302-4.
22. Hill AVS, Allsopp CEM, Kwaitkowski D, Ansty NM, Twumasi P, Rowe PA, et al. Common West African HLA antigens are associated with protection from severe malaria. Nature 1991; 252:595-600.
23. Lurie MB. Resistance to tuberculosis: experimental studies of native and acquired defensive mechanisms. Cambridge, MA: Harvard University Press, 1964.
24. Levin BR, Svanborg-Eden C. Selection and the evolution of virulence in bacteria: an ecumenical excursion and modest suggestion. Parasitology 1990; 100:S103-15.
25. Anderson RA, May RM. Infectious diseases of humans: dynamics and control. Oxford, UK; Oxford University Press, 1991: vii, 757.
26. Anderson RM, May RM. Co evolution of hosts and parasites. Parasitology 1982; 85:411-26.
27. Levin BR, Alison AC, Bremermann HJ, Cavali-Storza LL, Clarke BC, Frentzel-Beymem R, et al. Evolution of parasites and hosts (group report). In: Anderson RM, May RM, editors. Population biology of infectious diseases. Berlin: Springer, 1982:212-43.
28. Fenner F, Cairns J. Variation in virulence in relation to adaptation to new hosts. In: Burnet FM, Stanley WM, editors. The viruses: biochemical biological and biophysical properties. New York: Academic Press, 1959:225-49.
29. Fenner F, Ratcliffe FN. Myxomatosis. Cambridge, UK: Cambridge University Press, 1965.
30. Fenner FM, Day MF, Woodroofe GM. Epidemiological consequences of the mechanical transmission of myxoma by mosquitoes. Journal of Hygiene 1956; 54:284-303.
31. Mead-Briggs AR, Vaughan JA. The differential transmissibility of myxoma virus strains of differing virulence grades by the rabbit flea *Spilopsyllus cuniculi* (Dale). Journal of Hygiene 1975; 75:237-47.
32. Greenwood M, Hill AB, Topley WWC, Wilson J. Experimental epidemiology. London: Medical Research Council, 1936:209:1-204.
33. Herre EA. Population structure and the evolution of virulence in nematode parasites in fig wasps. Science 1993; 259:1442-5.
34. Ebert D. Virulence and local adaptation of a horizontally transmitted parasite. Science 1994; 265:1084-6.
35. Ewald PW. Host parasite relations, vectors, and the evolution of disease severity. Annual Review of Ecology and Systematics 1983; 14:465-85.
36. Lipsitch M, Nowak ML. The evolution of virulence in sexually transmitted HIV/AIDS. J Theor Biol 1995; 174:427-40.
37. Levin BR, Bull JJ, Stewart FM. The intrinsic rate of increase in HIV/AIDS: epidemiological and evolutionary implications. Math Biosci 1996; 132:69-96.
38. Levin BR, Lenski RE. Coevolution of bacteria and their viruses and plasmids. In: Futuyama DJ, Slatkin M, editors. Coevolution. Sunderland, MA: Sinauer Associates, 1983:99-127.

39. Bull JJ, Molineux IJ, Rice WR. Selection of benevolence in a host parasite system. Evolution 1991; 45:875-82.
40. Sasaki A, Iwasa Y. Optimal growth schedule of pathogens within a host: switching between lytic and latent cycles. Theor Popul Biol 1991; 39:201-39.
41. Antia R, Levin BR, May RM. Within-host population dynamics and the evolution and maintenance of microparasite virulence. American Naturalist 1994; 144:457-72.
42. Bonhoeffer SA, Nowak MA. Mutation and the evolution of virulence. Proc R Soc Lond B Biol Sci 1994; 258:133-40.
43. Nowak MA, May RM. Superinfection and the evolution of parasite virulence. Proc R Soc Lond B Biol Sci 1994; 255:81-5
44. Koella JC, Antia RN. Optimal pattern of replication and transmission for parasites with two stages in their life cycle. Theor Popul Biol 1995; 41:277-91.
45. Bonhoeffer S, Nowak MA. Intra-host versus inter-host selection: viral strategies of immune function impairment. Proc Nat Acad Sci USA 1994; 91:8062-6.
46. Lenski RE, May RM. The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. J Theor Biol 1994; 169:253-65.
47. Abedon ST. Selection for bacteriophage latent period length by bacterial density: a theoretical examination. Microbial Ecology 1989; 18:79-88.
48. Stewart FM, Levin BR. The population biology of bacterial viruses: why be temperate? Theor Popul Biol 1984; 26:93-117.
49. Lipsitch M., et al. The population dynamics of vertical and horizontally transmitted parasites. Proc R Soc Lond B Biol Sci 1995; 260:321-7.
50. Lipsitch M, Siller S, Nowak MA. The evolution of virulence in pathogens with vertical and horizontal transmission. Evolution 1996 (in press).
51. Levin BR, Bull JJ. Short-sighted evolution and the virulence of pathogenic microorganisms. Trends Microbiol 1994; 2:76-81.
52. Gould SJ, Lewontin RC. The spandrels of San Marco and the pangalossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B Biol Sci 1979; 205:581-98.
53. Whitnack E. Sepsis. In: Schaechter M, Medhoff G, Eisenstein BI, editors. Mechanisms of microbial disease. Baltimore: Williams & Wilkins, 1993: 770-8.
54. Finlay BB, Falkow S. Common themes in microbial pathogenicity. Microbiol Rev 1989; 52:210-30.
55. Fauci AS. Multifactorial nature of human immunodeficiency virus disease: implications for therapy. Science 1993; 262:1008-11.
56. Jacques A, Koopman JS, Simon CP, Longini IM. The role of primary infections in epidemics of HIV infections in gay cohorts. J Acquir Immune Defic Syndr 1994; 7:1169-84.
57. Nowak MA, Anderson RM, McLean AR, Wolfs TFW, Goudsmit J, May RM. Antigenic diversity thresholds and the development of AIDS. Science 1991; 254:963-9.
58. McLean AR. The balance of power between HIV and the immune system. Trends Microbiol 1993; 1:9-13.
59. Mittler JM, Antia R, Levin BR. Population dynamics of HIV pathogenesis. Trends in Ecology and Evolution 1995; 10:224-7.
60. Mittler JM, Levin BR, Antia R. T-cell homeostasis, competition and drift: AIDS as HIV-accelerated senescence of the immune repertoire. J Acquir Immune Defic Syndr Hum Retrovirol (in press).
61. Popper KR. The logic of scientific discovery. New York: Harper, 1965: 479.
62. Fenner F. The epizootic behaviour of mousepox (infectious ectromelia of mice) II. The course of events in long-continued epidemics. J Hygiene 1948; 46:383-93.

The Infectious Diseases Impact Statement: A Mechanism for Addressing Emerging Diseases

Edward McSweegan, Ph.D.
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

The use of an Infectious Diseases Impact Statement (IDIS) is proposed for predictive assessments of local changes in infectious diseases arising from human-engineered activities. IDIS is intended to be analogous to an Environmental Impact Statement. The drafting of an IDIS for specific activities, particularly in developing nations, would provide a formal mechanism for examining potential changes in local health conditions, including infected and susceptible populations, diseases likely to fluctuate in response to development, existing control measures, and vectors likely to be affected by human activities. The resulting survey data could provide a rational basis and direction for development, surveillance, and prevention measures. An IDIS process that balances environmental alterations, local human health, and economic growth could substantially alter the nature of international development efforts and infectious disease outbreaks.

A 1995 report by Aksoy et al. (1) describing the GAP (Turkish acronym for the Southeastern Anatolia Irrigation Project) irrigation project in Turkey suggests that anticipating the emergence or expansion of vector-borne and zoonotic diseases in a limited environment is a useful exercise. According to the report, a number of diseases (e.g., leishmaniasis, malaria, and schistosomiasis) are likely to increase in direct response to the expansion of irrigation and the increases in under water acreage and human population in the GAP region. The succinct overview of the disease and vector conditions in the GAP area could serve as a starting point for creating what will be referred to in this article as an Infectious Diseases Impact Statement (IDIS), a document that would be analogous to the Environmental Impact Statement (EIS) routinely used in the United States to assess the likely effects of construction, irrigation, agriculture, and similar activities on a local environment or region. An IDIS, however, would not assess the environment directly, but rather would predict changes in local disease patterns resulting from changes to the local environment.

Like an EIS, an IDIS would be a predictive and proactive assessment. Drafting an IDIS for a particular region or microenvironment would provide a formal mechanism for asking (and attempting to answer) specific questions about future changes in local health conditions. For example, what are the diseases and vectors in the given area? How are the proposed changes to the environment (e.g., dam-building, forest-clearing) likely to change the incidence and the prevalence of those diseases and vectors? What actions should be taken during the course of a given project and in the future to prevent potential increases in disease and vector populations? If an increase in human disease is likely, is the expense of the proposed project warranted? Will the economic benefits of a particular development or agriculture project be offset by increased costs in health care, vaccination, and vector control?

The 1969 National Environmental Policy Act was designed to provide a legal mechanism in the United States for evaluating potential impact to the environment from development activities and for permitting the public to participate in the evaluation process at the earliest stages (i.e., "scoping"). A Council of Environmental Quality in the executive branch of the federal government was established as a monitor, and the EIS process was implemented to inform decision makers and the public of potential environmental problems and reasonable alternatives to proposed actions. Environmental Protection Agency (EPA) requirements for environmental assessments are outlined in the Code of Federal Regulations (CFR). An EIS is intended to prospectively examine impact "upon the quality of the human environment of the United States and, in appropriate cases, upon the

Address for correspondence: Edward McSweegan, Ph.D., National Institute of Allergy and Infectious Diseases, National Institutes of Health, Solar Bldg., Rm. 3A34, Bethesda, MD 20892-7630, USA; fax: 301-402-0659; e-mail: em8p@nih.gov.

environment of the global commons outside the jurisdiction of any nation" (46 CFR sec. 504.7). The requirements of an EIS typically include descriptions of human populations in the designated area, current land use patterns, air quality, noise levels, locations of wetlands and coastal zones, sites of historical or cultural value, and non-point source pollution. The protection of human health is implied in the EIS process, but this concern is usually assumed to pertain to the location of industrial plants and dumps and to exposure to toxic chemicals, heavy metals, ionizing radiation, and pesticides. The EIS process contains no explicit reference to infectious diseases or disease vectors affecting human health in response to deliberate environmental changes. In 1995, the only published EIS references to diseases, infectious or otherwise, were for proposed control measures at two California plant nurseries. Yet past events suggest that attention should be directed toward changes in infectious disease patterns directly attributable to human-engineered events.

For example, the construction of the Aswan Dam in Egypt is widely believed to have precipitated the appearance of Rift Valley fever (RVF) in Egypt during the 1970s (2). Tens of thousands of RVF cases and hundreds of deaths followed. Similarly, completion of the Diama Dam in Senegal, in 1987, led to epidemics of malaria and RVF (3); impoundment of the Volta Lake in Ghana, in 1968, led to an explosive outbreak of schistosomiasis (4). Increased agriculture on the Argentine pampas and along the edges of Bolivian forests has contributed to frequent hemorrhagic fever outbreaks caused by Junin and Machupo viruses, respectively (5). Mining operations in the Brazilian jungles have led to outbreaks of malaria (6). Road-building projects under way in Papua New Guinea are likely to bring large numbers of susceptible human hosts into contact with rare and yet-to-be-discovered viruses. These epidemics and encounters with new diseases are the unforeseen consequences of critically altering the local environment. As a consequence, development and agriculture projects initiated to improve the lives of local populations can have the opposite effect by increasing disease prevalence and causing new epidemics. Embedding an IDIS requirement into the planning and execution of large-scale projects likely to alter local environments could prevent new epidemics and reduce infectious disease–associated morbidity and mortality.

How Would an IDIS Work?

An IDIS would first need to be established as an integral component of any activity likely to affect the health of a local population. In tropical and developing regions of the world, that would include a variety of national and international development activities. The area designated for large-scale alteration would be surveyed for current disease vectors, and the local populations would be examined for diseases likely to be affected by the project in question. The quality of the surveillance and the extrapolation of expected changes brought on by a particular activity would vary, depending on the knowledge of diseases, vectors, local host immunity, and other factors. Although the variables increase the margin of uncertainty in such extrapolations, these estimates would be expected to improve as the state of field and laboratory research improves and experience with preparing an IDIS increases. A retrospective examination of earlier projects in similar environments would also provide information for developing an IDIS. The standards of "existing credible scientific evidence" and "reasonably foreseeable" impact that current environmental impact assessments rely on could also be applied to the early stages of the IDIS process.

The resulting preproject assessment would provide a snapshot of conditions in a defined area, including the following: diseases likely to fluctuate in response to project activities, numbers of infected and susceptible hosts, existing control measures, and vectors likely to be affected by project activities. Such baseline data are frequently absent from development and agriculture activities (7). Knowing what diseases are already present, and how they might be changed, allows one to ask how anticipated changes in disease prevalence and distribution might be prevented or controlled through changes in the proposed project, improved case finding and treatment, changes in local sanitation and housing, increased vaccination or prophylaxis, or pest management programs. Some or all of the above health maintenance measures could then become components of the overall planning, budgeting, and execution of any major development or agricultural activity in the area. Health and health maintenance would become factors in the overall design and cost of the project. In many instances, local disease surveillance would become an ongoing part of the project, with supplemental assessments being made to refine the original IDIS.

Who Would Request an IDIS, and Who Would Respond to the Request?

Initial candidates would likely be donor organizations (the U.S. Agency for International Development and the World Bank, for example) that provide funding and oversight. In the absence of federal or international statutes, these organizations have the stature and financial capability to make infectious disease control an integral part of their development projects. Indeed, they should have an urgent interest in doing so because increases in diseases or new epidemics increase financial demands on them for medicines, vaccines, and pest control. In the end, more money would be spent beating back the outbreaks and epidemics that foresight might have prevented. National health ministries, state and territorial health departments, and local medical communities in developing countries might also request or initiate an IDIS. The practice of drafting an IDIS and implementing its recommendations might also rejuvenate underfunded areas of international health, vector biology, parasitology, and medical entomology as professionals in these fields are called on to conduct infectious disease assessments of development activities. The peer-reviewed literature and electronic services such as ProMED, Outbreak, and the World Health Organization (WHO) and Centers for Disease Control and Prevention World-Wide Web sites could provide the public "scoping" role that posting in the *Federal Register* and allowing a period for public comment provide in the EIS process in the United States.

The first application of an IDIS to a large-scale development or environmental activity could come from western donor organizations working in the developing world. The successful demonstration of an IDIS could encourage other organizations, national health officials, and health activists to push for the routine integration of public health with national development. This could happen in the United States, as well. The United States recently experienced the emergence of Sin Nombre virus in the Southwest and is theoretically open to the introduction of five vector-borne diseases: malaria, Rift Valley fever, yellow fever, dengue, and arbovirus encephalitides (15). Public health officials and citizen activists could initiate independent IDIS for activities perceived to threaten the balance between health, the environment, and domestic productivity.

What are the Strengths and Limitations of the IDIS Process?

A project-embedded IDIS would not be the same as an environmental management program, which seeks to control disease vectors through environmental modification and manipulation and through reduced human contact with vectors (8). An IDIS would, in fact, precede environmental management control measures by first postulating the likely emergence of specific pathogens and vectors. The usefulness of an IDIS lies in its ability to provide a conceptual framework for identifying potential disease problems, and, indeed, preventing them by altering or curtailing the very activities that could lead to disease emergence.

In an activist sense, an IDIS could be wielded as a tool of caution or prevention, much as an EIS is wielded in the United States to alter or halt some activity perceived to be a threat to the environment. That ability to influence potential changes and to affect health could be vital; public health concerns connected with agricultural and developmental projects are usually a low priority among foreign ministries, international donor organizations, and engineers (9); neglecting them can leave the full benefits of development unrealized.

Lest anyone imagines that an IDIS could be used solely as a tool of the political Left, as a kind of "liberation microbiology," it is important to point out that the same IDIS could be used to justify the use of pesticides and other organized control measures, including the relocation of local populations. Recently, for example, pesticide use has come under attack by various environmental groups, and donor organizations have become increasingly reluctant to fund such activities (10). In the United States, EPA's Endangered Species Act has also tended to thwart the use of pesticides because of potentially adverse impact on some birds and mammals (11). However, an IDIS describing the probable emergence of important disease vectors could be used to justify such use. Thus, a health care issue could be twisted into a health scare by either the political Right or the Left. The recent ratification of the North American Free Trade Agreement (NAFTA) was preceded in the United States by an effort to stall the treaty with an EIS requirement. If an IDIS had predicted new disease outbreaks from increased border trade and traffic, that concern might have had greater impact on the public imagination than more abstract concerns about atmospheric particulates in the border

region and could have been effectively used by anti-NAFTA forces. An IDIS should be not a political tool but rather a valuable information source that helps guide economic development and land use.

How Can an IDIS Complement Existing Surveillance Systems?

Almost half of the planet's five billion people are at risk for one or more vector-borne diseases (12, 13). Surveillance remains a key tool for monitoring these diseases and identifying new cases and outbreaks. Four types of surveillance are used in the control of vector-borne diseases: 1) recording human cases, 2) determining vector distribution and infectivity, 3) monitoring vertebrate reservoirs, and 4) tracking weather patterns to predict vector distribution (14). But throughout the developing world and across tropical boundaries, effective and continuous surveillance is extremely difficult, if not impossible. Cases are missed; outbreaks go unreported. Effective case reporting and continuous field monitoring are best conducted in limited, well-defined areas. Within the microenvironments of human activities, an IDIS could provide valuable baseline surveillance data before changes to that area occur and affect disease and vector distributions. This information could provide a rational basis and direction for ongoing monitoring and corrective measures (e.g., vaccination, relocation, pest control). Focusing on a limited area and a limited number of diseases in that area may also expand the use of promising but underutilized technologic methods such as remote sensing and geographic information systems (GIS).

Haines et al. (15) noted the importance of vector-borne disease monitoring and recommended that remote sensing and GIS be used to detect changes in ecosystems and vector populations. To a large extent, however, the advantages of satellite imagery and GIS have not been realized, in part, because of the frequent absence of "ground truth" (data on diseases, vectors, and other factors in the area) and of having to wait to observe natural environmental changes likely to affect disease and disease transmission (16-18). Satellite imagery for much of the planet has been collecting in databases since 1972 (16). By 1998, accumulated satellite data will be in the petabyte (1,000 terabyte) range, 1,000 times larger than the contents of the Library of Congress (B. Montgomery, NASA, pers. comm.). High-resolution, multispectral, multiyear images for many potential development and agricultural sites are available. Using a preproject IDIS to "ground truth" the project's environment with current satellite imagery, it may be possible to more completely describe local disease and vector conditions and make more accurate predictions about their plasticity during periods of construction, flooding, or farming. The result would be a firmer linkage of ground surveillance and satellite imagery to monitor public health changes within a well-defined and limited environment.

In recent years, the sudden emergence of rare or forgotten diseases such as Ebola virus infection, dengue, yellow fever, plague, and hantavirus (Sin Nombre virus) infection has attracted the attention of the public and inspired renewed commitments to surveillance and control. WHO recently formed a rapid response unit (the Division of Emerging, Viral and Bacterial Diseases Surveillance and Control) to deal with outbreaks of new and reemerging infections (19). Similarly, nine Southeast Asian countries held a meeting on emerging diseases and concluded that each country should also develop rapid response teams for epidemics (20). However, these disease control efforts are almost entirely passive, with staff, equipment, and budgets idling in anticipation of something eventually happening somewhere. It is difficult to maintain a high degree of public and financial support for such wait-and-see approaches to disease control. The United States has suffered a serious decline in national surveillance and outbreak investigations, in part, because of decreased support for passive monitoring programs (11).

Is an IDIS Really Needed When the Existing EIS Statutes Already Cover Human Health and Safety Concerns?

In the United States the need is not clear. Infectious diseases caused by environmental manipulation may be assumed to fall under the general EIS category of human health. However, infectious diseases have not often been considered in the past, and it is easy to imagine that if they were a factor in the EIS process, an environmental/infectious disease issue could be smothered under the weight of government regulations and the adversarial legal system. EPA operates under 16 federal statutes and 70 congressional committees and subcommittees and is engaged in some 600 lawsuits at any given time (21). Moreover, emerging infectious disease issues could bring EPA and the EIS process into conflict with the missions of

federal agencies and state and local health departments. Outside the United States, beyond federal statutes and informed public debate, the need for an IDIS is clearer. In the developing world, epidemics and substandard health care are common, and the national goals of healthy environment and healthy economy are usually at odds. An IDIS process that balances environmental alterations, local human health, and economic development could substantially alter the nature of international development efforts and infectious disease outbreaks.

To the ancient Greeks, the past appeared in front of them, real and visible; the future was behind them, unseen and unknowable. With that perspective, they were always glancing nervously backward, looking for a future that usually managed to creep up and tap them on the shoulder. In a sense, we have the same perspective for disease surveillance and control that the ancient Greeks had for time. Past epidemics and our responses to them are readily apparent; it is that unexpected tap on the shoulder by a hantavirus or an Ebola virus that is always so startling. We cannot know when and where such pathogens will emerge. Their appearance is often a chance event initiated by unpredictable changes in weather or the accidental encounter of a single person with a mysterious vector. These taps on the shoulder are an affront to our sense of control and understanding of disease. Moreover, it is unsettling to the public's sense of security and its faith in medical research. Although we cannot expect to eliminate the surprises of emergent pathogens in the near future, we can take control of situations in which our own actions directly lead to the emergence of diseases. Generating an IDIS in areas where human activities are likely to disrupt endemic-disease patterns would be an important step in controlling future outbreaks. Routine application of a preproject IDIS could improve local surveillance and health care planning by 1) providing baseline data on endemic-disease and vector prevalence and competence; 2) embedding projected health maintenance costs into the planning and cost of any project or activity likely to influence the environment and public health; and 3) providing a mechanism for instituting project alterations and health care measures to offset adverse effects on the health of local populations.

Acknowledgments

Special thanks to Dr. Michael Gottlieb, National Institute of Allergy and Infectious Diseases, Parasitology and International Programs Branch, and Donald C. Baur, Esq., Perkins-Cole, Washington, D.C. for critical comments and suggestions.

Dr. McSweegan is a member of the Parasitology and International Programs Branch at the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, Bethesda, Md. He has a long-standing interest in international science and health, and emerging diseases. He served in the U.S. State Department as an AAAS Fellow before joining NIAID's Tropical Medicine and International Research Office in 1988.

References

1. Aksoy S, Ariturk S, Armstrong MYK, Chang KP, Dörtbudak Z, Gottlieb M, et al. The GAP project in southeastern Turkey: the potential for emergence of diseases. Emerging Infectious Diseases 1995; 1:62-3.
2. Meegan JM, Shope RE. Emerging concepts on Rift Valley fever. In: Pollard M, editor. Perspectives in virology. New York: Alan R. Liss, 1981.
3. Lederberg J, Shope RE, Oaks SC, editors. Emerging infections: microbial threats to health in the United States. Washington, DC: Institute of Medicine, National Academy Press, 1992;71-2.
4. Scott D, Senker K, England EC. Epidemiology of human *Schistosoma haematobium* infection around Volta Lake, Ghana, 1973-75. Bull WHO 1982;60:89-100.
5. Morse SS. Emerging viruses: defining the rules for viral traffic. Perspect Biol Med 1991;34:387-409.
6. de Andrade ALSS, et al. High prevalence of asymptomatic malaria in gold mining areas in Brazil. Clin Infect Dis 1995;20:475.
7. Service MW. Rice, a challenge to health. Parasitol Today 1989;5:162-5.
8. Ault SK. Environmental management: a re-emerging vector control strategy. Am J Trop Med Hyg 1994;50(Suppl):35-49.
9. Silver GA. 1995. International Health Organization Policy Watch. The Federation of American Scientists. (http://www.clark.net/pub/gen/fas/ihm).
10. Arata AA. Difficulties facing vector control in the 1990s. Am J Trop Med Hyg. 1994;50(Suppl):6-10.
11. Longstreth J, Wiseman J. Human health. In: Smith JB, Tirpak DA, editors. The potential effects of global climate change on the United States: Appendix G, Health. Washington, DC: U.S. Environmental Protection Agency, 1989.
12. Beck LR, Rodriguez MH, Dister SW, Rodriguez AD, Rejmankova E, Ulloa, et al. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. Am J Trop Med Hyg 1994;51:271-80.

13. Knudsen AB, Sloof R. Vector-borne disease problems in rapid urbanization: new approaches to vector control. Bull WHO 1992;70:1-6.
14. Consortium for International Earth Sciences Information Network (CIESIN) Thematic Guides. Provisional Release. 1995. Programs for surveillance, treatment, and control of vector-borne diseases.
15. Haines A, Epstein PR, McMichael AJ. Global health watch: monitoring impacts of environmental change. Lancet 1993;342:1464-9.
16. Washino RK, Wood RL. Application of remote sensing to arthropod vector surveillance and control. Am J Trop Med Hyg. 1994;50Suppl:134-44.
17. Rogers DJ, Williams BG. Monitoring trypanosomiasis in space and time. Parasitology 1993;106:S77-S92.
18. Barinaga M. Satellite data rocket disease control efforts into orbit. Science 1993; 261:31-2.
19. World Health Organization. WHO establishes new rapid-response unit to combat growing world-wide threat of emerging diseases. WHO/75. Press release, 17 October 1995.
20. Plianbangchang S. Southeast Asia intercountry consultative meeting. Emerging Infectious Diseases 1995;1:158.
21. Environmental Protection Agency. The common sense initiative. Pub. No. EPA100F94004. Washington, DC: Environmental Protection Agency.