The Role of Geographic Information Systems in Population Health

Russell S Kirby, PhD, MS, FACE (1),* Seth L Foldy, MD (2)

(1) Department of Obstetrics and Gynecology, Milwaukee Clinical Campus, University of Wisconsin-Madison Medical School, Milwaukee, WI; (2) Commissioner, Milwaukee Department of Health and Department of Family and Community Medicine, Medical College of Wisconsin, Milwaukee, WI

Abstract

This paper explores applications of geographic information systems (GIS) in population health and the preconditions for its optimal use. Population health involves the assessment, evaluation, and optimization/improvement of health status and outcomes on a population basis. It is the ultimate pursuit of public health programs, which are more often focused reactively than actively on underserved groups or those with diseases or health needs not adequately treated by the health care delivery system. As public health broadens its focus toward the determinants of population health, GIS can perform several functions in population health informatics. GIS has hardware, software, and staffing requirements; in population health, a more important precondition for their use is a systematic, integrated approach to geocoding all population-based health data systems. With routinely geocoded databases, GIS can fulfill many roles in population health informatics. Functions include an interactive environment for the spatial display of health data; a laboratory for the development and dissemination of neighborhood/community health indicators; a tool for integrating disparate data records by location; a vehicle for displaying results of analyses from databases merged by automated record linkage; a platform for testing hypotheses concerning the epidemiologic determinants of health status, diseases/outcomes, or associations between determinants of population health and utilization of health services; and a vehicle for facilitation of public health program planning, evaluation, and community-based decision-making. As the implementation of health-oriented applications of GIS evolves, with appropriate attention to geographic, epidemiologic, and biostatistical methodological concerns and methods of map presentation, opportunities for extending GIS into population health informatics are almost limitless.

Keywords: population health, informatics, geocoding, record linkage

Introduction

While GIS has a significant role in traditional public health activities, their ability to collate, integrate, and display population-based data concerning health events, exposures, risk factors, and socio-environmental data warrants a broader, more holistic perspective on the health of populations. This brief essay explores the opportunities for population health research and practice and the central role for GIS within the emerging paradigm of population health. We have four objectives in this presentation. First, we will define population health as a distinct field of intellectual inquiry, and compare and

* Russell S Kirby, PhD, Department of Obstetrics and Gynecology, University of Wisconsin Medical School, PO Box 342, Milwaukee, WI 53201-0342 USA; (p) 414-219-5610; (f) 414-219-5201; E-mail: r-kirby@whin.net

contrast population health with the current practice of public health in most states, cities, and communities in the United States. Second, we will provide a framework for population health informatics as an operational environment for the practice of population health. Third, we will explore the potential roles and opportunities for GIS in population health. Finally, we will identify methodological issues, opportunities for multidisciplinary interaction and collaboration, and applied research in GIS-based population health.

Population Health

"Population health" refers to the health, well-being, and functioning of entire populations. It shares with public health an explicit focus on whole populations. However, the scope of population in population health may be defined flexibly—to include, for example, covered lives in a managed-care plan or a corporate workforce, rather than people within a geographic government jurisdiction. In addition, population health examines a broader set of inputs and health outcomes than are traditionally studied in public health.

What are the determinants of population health? Evans and Stoddart provide a conceptual framework for population health (1), reproduced with modifications in Figure 1. Factors like social environment and prosperity appear both as inputs and outputs, affecting each other in a reciprocal series of relationships. These extend beyond traditional host-agent-environment concerns of public health (while still incorporating genetics, behavior, and the physical environment). The model for population health also includes, but radically extends, the medical preoccupation with the relationship between disease and health care. While the traditional medical model focuses on the determinants and treatment of diseases through the provision of health care services, population health focuses on goals including general well-being and functioning, not just disease. Most importantly, outcomes of interest from the perspective of population health form a superset of the traditional public health outcomes. Societal levels of health and functioning, or general well-being, are the outcomes of greatest interest. Population health thus calls attention to the relationships between culture, polity, economy, environment, and health care utilization and quality. The model of the determinants of population health places the medical model of health care in broad societal perspective, and provides a general prevention focus for practitioners of population health.

For example, outcomes typically measured in health care delivery include appointment wait times, rehospitalization after emergency room visits, or five-year survival after cancer treatment. Typical public health outcomes include infant mortality, immunization rates, or the incidence of lead poisoning, selected from literally hundreds of measures listed in *Healthy People 2000: National Health Promotion and Disease Prevention Objectives* (2). Population health extends this view to broader health status measures like premature mortality rates (3), change in quality-of-life years (4), and the rate of limitation in activities of daily living. These are clearly influenced by, but not entirely defined by, traditional medical and public health measures.

Population health derives from a variety of intellectual traditions. These include public health (especially population-based data systems), demography, social and behavioral sciences (especially representative sample surveys of health, health status, and

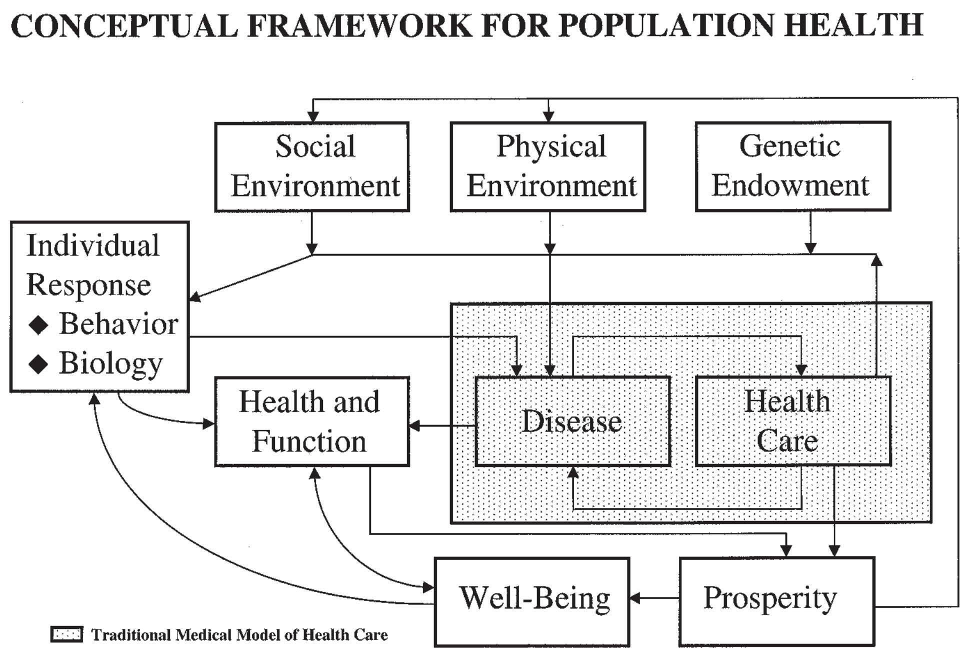

Figure 1 Conceptual framework for population health. Reprinted, with modifications, with permission from the authors, in Evans RG, Barer ML, Marmor TR, Why Are Some People Healthy and Others Not? The Determinants of Health of Populations. New York: Aldine de Gruyter, 1994. Copyright ©1994 Walter de Gruyter, Inc., New York.

well-being), environmental health, epidemiology, and health services research. This emerging discipline gains strength from the diversity of perspectives brought to bear on the issues at hand, and from the necessity for integration of staff and resources from the public, private, and academic sectors. (See Young's *Population Health: Concepts and Methods* [5] for a more thorough introduction to this field and its practice.)

Population Health Informatics

"Informatics" is defined in the American Heritage Dictionary as "information science." Recently, Friede et al. (6) described public health informatics as an emerging field that encompasses public health surveillance methodologies, data and databases, and information systems, used collectively to merge, manage, analyze, and interpret public health data. The practice of public health informatics implies a paradigm shift in the institutional arrangement, management, operationalization, and utilization of databases and information services within the public health sector. In most states and municipalities, integration of public health information services across the spectrum of programmatic activities has yet to be realized (7). The core public health functions of assessment, policy development, and assurance (8) would all be enhanced through the expansion and integration of the current health statistics/epidemiology units at the state, municipal, and local levels into seamless public health system-wide informatics environments.

Were we to accomplish this task, however, we would still fall short of the optimal information basis for population health. We propose a new field of intellectual endeavor, population health informatics, that builds upon public health informatics but includes the following additional features:

1. *Information on entire populations (not just service users).* Information on entire populations includes public health surveillance data (vital records, reportable diseases) and other government data. It can also be gathered by assembling data from the universe of overlapping organizations that serve an entire population (for example, immunization registries or cancer registries using data assembled from multiple health care providers).
2. *Integration of databases linking public health information, environmental information, health services information, and socioeconomic information to health outcomes.* Because both the social and physical environment (factors like income, education, housing quality, and air quality) and health services (accessibility, utilization, effectiveness, efficacy) have considerable impact on health, data reflecting these conditions should be linked to health outcome data at a reasonably discrete level.
3. *Focus on broader health outcome measures including functional status, disability, and quality of life, assessed across populations.* An added goal is to associate the above information with population-level indicators of health that go beyond the traditional (mortality, prevalence) to include meaningful indicators of quality outcomes for large portions of the population (well-being, function, quality of life).
4. *The ability to define sub-populations flexibly (not only across administrative geographic units, but also by other characteristics).* Because services and policies typically affect sub-populations (e.g., members of a neighborhood or a managed-care group), a population health information system should allow the creation and analysis of sub-population information sets. Health planning is facilitated through such networks. Population health informatics can support a broader perspective on the determinants of population health; rarely does the zip code or municipality of residence correlate directly at the individual unit of analysis with likelihood of exposure to environmental hazards or utilization of health services.

Population health informatics establishes an information environment for assessing and monitoring the health, functioning, and well-being of entire populations, consistent with the Evans and Stoddart model in Figure 1. The integration of such population-based information from the many, unconnected systems that already exist would profoundly improve health planning, public health surveillance, and health services research. Integrating this information would also provide a powerful platform from which to study the impact of health and social policies on population health. Like informatics in general, population health informatics represents a fundamental transformation of the manner in which population-based health data are collected, managed, and used to support the core functions of public health—assessment, assurance, and policy development—as applied to entire populations. Although there are many barriers to the realization of such an information environment, it is not too soon to consider the technical preconditions for such a vision. In the following section, we explain why geographic information systems (GIS) are central to the conceptualization and practice of population health informatics. The balance of this paper describes some of the potential roles for GIS in population health, as well as related issues.

Roles for GIS in Population Health Informatics

One of the central requirements for population health informatics is the integration of information systems that contain the broad range of data of interest. The integrative nature of population health requires data from systems containing health information, socioeconomic information, environmental information, and subjective or solicited information on outcomes like well-being or functioning. Existing sources include administrative databases (hospital billing records, tax files), public health databases (reportable diseases, vital statistics), programmatic databases (lead poisoning surveillance, immunization or cancer registries), census information, governmental housing and environmental databases, and representative sample surveys.

Database integration requires linkages among the various data at a discrete enough level to allow meaningful inferences about relationships, trends, co-factors, and confounding variables. The most discrete and useful form of linkage is through personal identity, represented by name or unique alphanumeric identifier. However, confidentiality and privacy concerns are very real and reasonable (9); thus, using personal identity may not be either ethically or politically feasible for many sources of data using current technology. A common but unsatisfying way to link health information is by broad categories like race (e.g., showing trends of low birth weight over time by race). Such broad linkages provide little insight into the relationships among the many variables that directly affect health outcomes, focusing instead on surrogate variables like skin color that may have little direct relationship to the outcomes of interest (10). Place, however, offers many advantages as a means of linking and then analyzing disparate data sources. Information on place is almost universally collected in health care documents (in the form of address), though it is not always entered into databases. It is often associated with a broad range of both socioeconomic and environmental factors. Using location (which can be manipulated or aggregated in various fashions) may also offer a lesser threat to personal privacy. It can serve as a definitive linkage point between two address-bearing databases, and as a categorical or a continuous two-dimensional variable along which imputation of data is possible (with appropriate care on the part of researchers and end-users).

We propose four levels of database integration in population health informatics. GIS would play important roles at each level.

- *Level 1: Surveillance of indicators.* This could include traditional surveillance, such as the incidence of communicable diseases over time. With the expanded data linkages of population health informatics, it might also include ongoing monitoring of indicators like the adequacy of prenatal care, ambulatory care sensitive hospitalizations, or arrests for illicit substance sales. While this form of unidimensional monitoring of trends can be accomplished without GIS, GIS provides the additional capability to rapidly analyze or display geographic sub-populations and to overlay geographic information over temporal information. An example might be maps to discern that increasing rates of low birth weight are occurring in a specific portion of a city.

- *Level 2: Geographic integration of multiple variables.* This integration can occur by area (data aggregated into administrative areas like census tracts, zip codes, or municipalities) or by discrete-point geocoding that displays and analyzes points or imputed spatial surfaces. To continue our example, geographic patterns of rates of low birth weight might be compared with rates and trends in premature deliveries, maternal smoking rates, prenatal care adequacy, prenatal clinic service areas, Medicaid enrollment rates, substance abuse arrests, and the incidence of sexually transmitted diseases. Because population sets are geocoded to discrete locations or very small areas, it is possible both to analyze and to display small-area information. Some information (like economic status) may be cautiously imputed from small-area census or other data. This may indicate associations between low birth weight and other features (with a cautious respect for potential fallacies of multi-level comparisons).
- *Level 3: Individual-level record linkage.* Automated linkage of individual records from multiple databases is now feasible, using probabilistic or deterministic linkage strategies. Record linkage methodologies have become standardized in recent years, and unique identifiers are not always necessary (11). By these methods, linked datasets are created (and subsequently stripped of personal identifiers). These datasets include health risks and outcomes (e.g., from birth records), participation in service programs, insurance type, and community-level data (such as income or exposure to drug sales) imputed from small-area data. To continue our example at this level, imagine that it can be shown that participants in a comprehensive prenatal care coordination program combining clinic-based and outreach-worker care have higher birth weights than individuals who live in the same area, with similar demographic and perinatal risk factors, but who do not participate in the program.
- *Level 4: Real-time, point-of-service information.* At the highest level of integration and functionality, population health databases accompany patients, health care workers, and public health workers on their daily business. Imagine that a young woman presenting for emergency care is automatically identified as receiving (or not receiving) high-risk pregnancy-related outreach services when she registers for care at the local emergency department. Although using population health databases at this level presents the greatest technical and confidentiality-related challenges, there are some existing applications that demonstrate the usefulness of population health systems for individual health services (12). Within an integrated population health information system linked to service delivery, opportunities abound for tailoring prevention, evaluation, and planning to achieve true continuous improvement in population health.

GIS is crucial at each of these levels. It greatly simplifies data management, display, and calculations for the first two levels. Also, when displaying information using commonly understood geographic boundaries, GIS helps communicate the immediate significance of information to a public who might otherwise fail to comprehend that they are at risk, inviting greater participation in planning and policy. The third and fourth levels use the specificity of point location both for linkage of records and for more discrete display and analysis (spatial representation free of arbitrary administrative polygons like zip codes or census tracts, allowing more natural visual and statistical representations of data). This point-specificity can also facilitate linkage to a greater number of databases, and can do so in a way that may be more respectful of individual confidentiality than would use of names or other personal identifying variables. Address information could link, for example, building age and ownership status (from plat records), median census block income (from the decennial census), housing inspection and lead abatement interventions (from administrative records), and reported blood lead levels (from public health surveillance data). These data could target interventions (service planning), derive predictive models (population-based epidemiology), or evaluate the effectiveness of housing policy changes (outcome effectiveness research). For these reasons, GIS, facilitated by geocoding of health and other data records, becomes the *sine qua non* of population health informatics.

While the first and second levels of information system integration can be accomplished with lower degrees of spatial specificity, creation of spatial surfaces, automated record linkage, and point-of-service integration requires that data records be geocoded. Geocoding of locational data (address of residence, location of injury event or exposure if known, address of place of employment, location of health care service provider) is the essential element of population health informatics. Through geocoding and map generation, GIS provides an essential tool for integration of data records from various sources by location. GIS applications can also serve as laboratories for development, interpretation, and dissemination of neighborhood/community health indicators. GIS also provides an interactive environment for spatial display of health data.

In public health data systems there are numerous perceived barriers to geocoding. These include cost, timeliness and accuracy, staff and equipment needs, and privacy/confidentiality concerns. All of these perceived barriers are smoke screens. There is no valid rationale for not routinely geocoding all records in vital statistics or hospital discharge databases, cancer or birth defects registries, and all other population-based public health information systems. In fact, geocoding can improve the precision of these databases by correctly allocating cases to county, zip code, minor civil division, or census tract, and is extremely efficient when integrated into the routine, day-to-day processing of records accruing to administrative health data systems. Geocoding of population-based public health data system records also facilitates the public health mission of the agencies and programs that sponsor or support the information system, by ensuring that valid, geocoded addresses are available for every record as a precondition for filing. Further, geocoding is easily made routine. Administrators who embrace the approaches of public or population health informatics will achieve maximum value by establishing centralized geocoding centers to process all records for their agencies.

GIS can also play a role in population health as a tool for the generation of research hypotheses concerning the epidemiologic determinants of health status, well-being, health outcomes, and health service utilization. GIS has a more limited role as a platform for hypothesis testing *per se* (13). GIS also provides a supportive environment for population-based public health program planning, program evaluation, and community-based decision-making.

Methodological Issues, Concerns, and Opportunities

The application of GIS in population health provides numerous methodological opportunities, but raises some significant issues and concerns. These have been discussed at greater length elsewhere (14–18), but include the following:

- Scale and aggregation in measuring contextual variables (i.e., the role of individual characteristics versus neighborhood variables or ecological correlates).
- Points versus areas, and rates versus numerators and denominators.
- Theoretical conceptions of space and analytical applications.
- Integration of spatial modeling and biostatistical methods with social and epidemiological theory.
- Methods of map presentation and interpretation.
- Methodological issues surrounding the quality of matching and record linkage (including geocoding).

Space does not permit a lengthy discussion of these issues here. While none of these issues has a simple solution, identification and development of a multidisciplinary working group to devise and implement GIS-based population health applications will prove beneficial in most settings.

Summary

Population health is an emerging framework for assessing and evaluating the health status and health outcomes of defined populations. It is in many ways a superset of traditional public health functions and goals. Population health informatics is the operationalization of an integrated information systems environment for the practice of population health. GIS is an integral and in many ways essential component of a comprehensive population health informatics system. GIS is, however, only a tool, not an end unto itself in the practice of population health.

References

1. Evans RG, Stoddart GL. 1994. Producing health, consuming health care. In: *Why are some people healthy and others not? The determinants of health of populations.* Ed. RG Evans, M Barer, TR Marmor. New York: Aldine de Gruyter. 27–64.
2. US Public Health Service. 1991. *Healthy people 2000: National health promotion and disease prevention objectives—full report, with commentary.* Washington, DC: Department of Health and Human Services.
3. Cohen MM, MacWilliam L. 1995. Measuring the health of the population. *Medical Care* 33:DS21–42.
4. Murray CJ. 1994. Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bulletin of the World Health Organization* 72:429–45.
5. Young TK. 1998. *Population health: Concepts and methods.* New York: Oxford University Press.
6. Friede A, Blum HL, McDonald M. 1995. Public health informatics: How information-age technology can strengthen public health. *Annual Review of Public Health* 16:239–52.
7. Lumpkin J, Atkinson D, Biery R, Cundiff D, McGlothlin M, Novick LF. 1995. The development of integrated public health information systems: A statement by the Joint Council of Governmental Public Health Agencies. *Journal of Public Health Management and Practice* 1:55–9.

8. Institute of Medicine Committee for the Study of the Future of Public Health. 1988. *The future of public health.* Washington, DC: National Academy Press.
9. Donaldson MS, Lohr KN, Eds. 1994. *Health data in the information age: Use, disclosure, and privacy.* Washington, DC: National Academy Press.
10. Krieger N. 1992. The making of public health data: Paradigms, politics, and policy. *Journal of Public Health Policy* 13:412–27.
11. Kirby RS. Manuscript in preparation. *Controlling the "urge to merge": Diagnosis and treatment of a new clinical psychosis affecting public health workers and researchers.* Unpublished paper presented at the Centers for Disease Control and Prevention 1996 Maternal, Infant, and Child Health Epidemiology Workshop: Data and Information for Planning, Prevention and Evaluation. Atlanta, GA. December 3–4, 1996.
12. Hall K, Zimmerman A, Samos J, Simon PR, Hollinshead WH. 1997. Coordinating care for children's health: A public health integrated information systems approach. *American Journal of Preventive Medicine* 13(Suppl 1):S32–6.
13. Jacquez GM. 1998. GIS as an enabling technology. In: *GIS and health.* Ed. T Gatrell, M Loytonen. London: Taylor and Francis. 17–28.
14. Kirby RS. 1996. Toward congruence between theory and practice in small area analysis and local public health data. *Statistics in Medicine* 15:1859–66.
15. Clarke KC, McLafferty SL, Tempalski BJ. 1996. On epidemiology and geographic information systems: A review and discussion of future directions. *Emerging Infectious Diseases* 2:85–92.
16. Croner CM, Sperling J, Broome FR. 1996. Geographic information systems (GIS): New perspectives in understanding human health and environmental relationships. *Statistics in Medicine* 15:1961–77.
17. Briggs DJ, Elliott P. 1995. The use of geographical information systems in studies on environment and health. *World Health Statistics Quarterly* 48:85–94.
18. Waller LA, McMaster RB. 1997. Incorporating indirect standardization in tests for disease clustering in a GIS environment. *Geographical Systems* 4:327–342.