MEASURING POPULATION HEALTH: A Review of Indicators

Vera Etches
Sudbury & District Health Unit, Sudbury, Ontario, Canada P3E 3A3; email: etchesv@sdhu.com

John Frank
Canadian Institutes of Health Research, Institute of Population and Public Health, Toronto, Ontario, Canada M5G 1L5; email: John.frank@utoronto.ca

Erica Di Ruggiero
Canadian Institutes of Health Research, Institute of Population and Public Health, Toronto, Ontario, Canada M5G 1L5; email: e.diruggiero@utoronto.ca

Doug Manuel
Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada M4N M5; email: Doug.manuel@ices.on.ca

Key Words  history, public health, epidemiological methods, framework, ethics

Abstract  This article reviews the historical development of population health indicators. We have long known that environmental, socioeconomic, early life conditions, individual actions, and medical care all interact to affect health. Present quantitative reporting on the impact of these factors on population health grew out of Bills of Mortality published in the 1500s. Since then, regular censuses, civil registration of vital statistics, and international classification systems have improved data quality and comparability. Regular national health interview surveys and application of administrative data contributed information on morbidity, health services use, and some social determinants of health. More recently, traditional health databases and datasets on “nonhealth” sector determinants have been linked. Statistical methods for map-making, risk adjustment, multilevel analysis, calculating population-attributable risks, and summary measures of population health have further helped to integrate information. Reports on the health of populations remain largely confined to focused areas. This paper suggests a conceptual framework for using indicators to report on all the domains of population health. Future ethical development of indicators will incorporate principles of justice, transparency, and effectiveness.
INTRODUCTION

Whether to ensure fit soldiers and laborers, to improve the plight of the poor, for normative or other purposes, interest in measuring the health of populations has a long history. This chapter traces the historical development and use of population health indicators to the present. To improve the health of populations and increase opportunities for comparability, more valid, comprehensive, transparent, and standardized ways of measuring and reporting on population health are needed. Such metrics will need to be connected to a clear conceptual framework, integrating relationships between the different elements of population health. This paper proposes such a framework.

Uses of Health Indicators

The uses of health indicators should contribute to overall population health goals, namely improving the health of populations and reducing health inequalities. Health indicators support this goal through the following key applications: advocacy, accountability, system management, quality improvement, and research (3). Although users generally accept that indicators have been helpful with respect to these approaches, evidence or well-developed evaluation of indicators is often lacking. The section on evaluation of indicators will further this point below.

Advocacy, one of the first applications of indicators of population health, remains one of the most important for improving population health. For instance, baptism and death registration led to a comparison of the indicator “infant mortality” in Halifax and Saint John, and in Massachusetts and England in the mid-1800s, to shame municipalities into taking measures to improve sanitation (26).

Indicators can be used, or not used, to advocate for different ideologies. Proponents of sanitary measures, again in the mid-nineteenth-century, demonstrate this point. Some used the discrepancy in mortality rates between rich and poor areas to argue that a more egalitarian society is healthier. Others presented mortality information to advocate for healthier laborers and soldiers and to maintain social stability, through the control of contagious diseases (74). Later, “eugenics societies” developed social Darwinist ideas and blamed the rise of diseases among the poor on people with poor heredity (83). Nazi ideology based its program of race hygiene on ideas of sociopathology partly by drawing on indicators of population health (83, p. 595).

Indicators are therefore not often employed neutrally; they can classify situations as “unfavorable” or “favorable,” to show “improvement” or “deterioration” over time (81). Different indicators are chosen not only for their different uses, but also because different societies—and authorities—will value/acknowledge some aspects of health or determinants of health more than others (75).

The next major use of indicators of population health is to achieve accountability. Reporting on population health can satisfy the needs of government, health professionals, voluntary agencies, and the public for information on risks, patterns,
and trends related to health and whether or not expectations for performance are met (29). In a related use, population-level indicators of health may inform system management to improve health. For example, in seventeenth-century London, England, parish clerks made regular weekly reports of the number of burials and the causes of death. Any indication of spreading plague resulted in systemic actions to limit its spread, and the weekly reports provided feedback on measures taken (74).

Indicators may provide information to improve the quality of initiatives for population health, even to improve the quality of measurement of population health itself. Indicators have often been suggested on a theoretical basis, before data are available to calculate them, to guide development of systems that provide better-quality information. Such was the case with indicators of mortality based on civil registration of vital statistics. Early World Health Organization (WHO) lists of indicators to be obtained from health interview surveys were meant to guide the development of new and improved information systems (98); greater international comparability of data collected did result.

Research may be stimulated by reports of population health indicators. Geographic disparities in health status, apparent from reporting of mortality, prompted research into causal explanations in the eighteenth century (47), as it still does today. Researchers also use indicators to evaluate the effectiveness of interventions.

Finally, indicators of population health can be applicable in other contexts, such as improving professional training (73). Development of indicators of population health became the core training of students of public health in the late-nineteenth century (82). Suitable indicators of population health can also be used more broadly for education in schools, workplaces, and institutions.

HISTORICAL DEVELOPMENT OF POPULATION HEALTH CONCEPTS AND RELATED INDICATORS

Current concepts of population health recognize that many interconnected aspects of society, the environment, and individuals all contribute to health. Several authors have pointed out that this recognition is not new (56, 82, 103). For example, the Chinese have acknowledged and written about multiple influences on health for millennia (99). The Greek distinction between the god of medicine’s two daughters Hygeia—the goddess of prevention and wellness, and Panacea—the goddess of treatment (although our current use of the term is strikingly broader!) suggests that people have long believed that there is more to health than health care (74). And, the imposition of quarantine in Marseilles (in 1377) and Venice (in 1403) demonstrated recognition of the need for population-level actions to improve health, above and beyond individual medical treatments (74). Environmental variables producing endemic disease identified by Hippocrates (500 BC) included climate, soil, and water (82). “Corrupting” the environment/air was the subject of the first English Sanitary Act passed in 1388 to control practices at slaughterhouses (74).
Regarding social variables, Johann Peter Frank (1745–1821) emphasized the need to transform social conditions to render them compatible with health and long life for all (30). Rudolf Virchow (1821–1902) felt the right to work safely, and to responsible government, was central to a healthy society (82). Among the demands for other health-promoting measures, nineteenth-century sanitarians called for greater efforts to promote healthy infancy and childhood development as a foundation for later life, education of women, town planning, and new housing for the poor (102). In summary, the beliefs that environmental, social, early life conditions, individual actions, and medical care all contribute to health long predated the availability of much quantitative data to support these views. Reformers in the mid-eighteenth century, however, began to call for more data collection in order to better promote the health of the population.

The debate between Farr and Chadwick, discussed in Hamlin’s insightful article, demonstrates the conflict over the meaning imputed to the numbers generated by analyses of routinely collected mortality data (46). Farr listed “poverty” as the cause of death for 63 people and commented, “hunger destroys a much higher proportion than is indicated by the registers in this and in every other country, but its effects, like the effects of excess, are generally manifested indirectly, in the production of diseases of various kinds” (46, pp. 857–58). Chadwick, on the other hand, believed the poor needed to behave more morally, and that environments required sanitation, to reduce mortality. The collection of data and use of indicators did not in itself lead to agreement over solutions; similarly, despite the proliferation of indicators of population health today, debate continues over the strategies to improve health among groups with lower incomes.

Social scientists and medical and other reformers in the nineteenth century both applied and collected new data on populations. By the 1870s, “it is clear that the social body had acquired a firmer existence and that it was increasingly understood to be a medical object” (26, p. 97). The influence of the population-level analyses of Marx, Engels, and Durkheim and the application of quantitative measures of determinants of health, in the late 1800s and early 1900s, continue. Porter (82) gives credit to the rise of the social sciences’ application to health and disease to the rise of “social medicine,” on both sides of the Atlantic, in the 1930s and 1940s.

Monitoring of the causes of death quantified populations’ experiences of the decline of air and waterborne “classical” diseases by the middle of the twentieth century, and their replacement with noncommunicable diseases (e.g., cardiovascular, diabetes, arthritis, depression). Researchers proposed theoretical frameworks for the shift that alluded to broad determinants of health-impacting physiological processes in individuals’ bodies (51, 78, 82). In this context, in 1974, Canada’s Lalonde Report revitalized the concept that multiple underlying causes of mortality and morbidity operate completely outside of the health care system per se, but contributed to public health practice that focused on changing individuals’ lifestyles (59, 102). In 1993, Porter rather ironically suggested that public health practice had “come full circle” to apply the same strategies used by people, like Chadwick in the nineteenth century, who urged the poor to “behave” more healthily (82).
Nevertheless, research on the impacts of broader and more “upstream” (69) determinants of health continued to grow. The WHO’s *Ottawa Charter for Health Promotion* (109) and its accompanying Canadian document *Achieving Health for All* led to a framework for the multifactoral determinants of human health. The conceptual framework developed by the Canadian Institute for Advanced Research’s (CIAR) Population Health Program also received wide acceptance as an approach to research and reporting on population health (35). Critics pointed to a lack of theoretical understanding of the impacts of global political regimes and social structures on health (24), as well as a lack of recognition of the agency of individuals and populations (85). A forthcoming and final book by the CIAR Population Health Program addresses some of these critics’ claims (49). Additional conceptual frameworks for population health, highlighted by Evans & Stoddart’s 2003 review (34), suggest a trend towards multiple “subframeworks” focusing on specific topics (e.g., occupational health).

Indicator development based on conceptual models have also tended to take a “focused area” approach. For example, the dimension of human developmental stages throughout life was one key factor lacking in the early CIAR framework, though early studies by Barker & Osmond (6) clearly showed associations between characteristics of infancy and childhood, and adult chronic disease. Hertzman & Wiens (48) later promoted the inclusion of what has come to be called the “life course” dimension, more explicitly to models of population health. A recently published glossary for life course epidemiology demonstrates the proliferation of measures in this area of focus (58).

Strongly influenced by CIAR’s work, Figure 1 reflects a current synthesis of the broad determinants of population and individual health. The return to a focus on poverty, inequalities in the distribution of resources, and other societal and environmental determinants of health may be seen as another turn of the full-circle wheel to which Porter (82) referred. This suggests that concepts of what makes a population healthy will develop, recede, and re-emerge in the future. Nonetheless, the ideas that health should be measured in multiple ways, at the population level, and that multiple nonmedical factors influence health are likely to remain. In contrast to the fairly stable uses of broad concepts of population health, data sources for constructing indicators have steadily improved over time, and are the subject of the next section.

### TYPES OF DATA

**Censuses and Vital Statistics**

Counting people dates back to ancient times and is central to a population’s governance. The United States was the first country to pass an act for regular, periodical enumeration in 1790 (114), yet such efforts were not true censuses in the modern sense (4). Initially, censuses simply provided population counts, including
numbers of people dying in the census year. Wright (114) notes that in the United States, early enumeration was used to ascertain the military and industrial strength of the country, since free males (16 years of age or older) were enumerated separately. Higgs (50) demonstrates how the censuses in England were constructed not only for social and economic purposes, but also to study the health effects of work in various settings, through the occupational classification system chosen. Difficulties of relying on censuses for mortality data included a large proportion of missing data; the migration of people; low-quality records for cause of death; a lack of knowledge and slowed recall of people supplying the data; and inadequately trained or reimbursed enumerators (36). Present-day censuses collect much other data on basic determinants of health useful for calculating indicators of population health: income, work, housing, immigration/language, and family/household relations.

Though some censuses predate the registration of vital statistics (births and deaths), vital statistics are the oldest continuous and most established type of population health data. That stated, many developing countries have yet to establish reliable systems for recording vital statistics (28).

Over the centuries, there have been many ways of, and reasons for, collecting vital statistics data. The first London Bills of Mortality, prepared in 1532, relied on data collected by churches (29). Parish records remained the main system of registration of baptisms, marriages, and deaths until the mid-nineteenth to early twentieth century in Europe and its colonies. However, baptisms served as a poor
proxy for births, since early infant mortality was high and many children died before baptism; others were not baptized through the church or were not baptized at all (89). Civil registration of births and deaths was advanced for reasons beyond merely obtaining more accurate health data: to calculate annuities on lives more accurately (32); for political and moral uses (10); for more complete and accurate information on the population for setting electoral boundaries and direct taxation (36); and to address concerns about inherited property rights (26). Clinicians in the mid-1800s did not generally acknowledge the importance of understanding the drivers of population “average” illness or death rates. Curtis (26) and Emery (33a) concur that the initiation of registrations of deaths and births, and censuses in a country, cannot be regarded as the culmination of “progress in relation to projects of medical mastery.”

Practical and jurisdictional barriers also had to be overcome before vital statistics registries could be applied to measuring the health of the population. The need for a uniform classification of causes of death became apparent. In 1768, John Fothergill declared the Bills of Mortality for London as inaccurate because, “the list of diseases was a very injudicious one,” and, “present labours, therefore, answered very little useful purpose, either to themselves or the community...” (42, p. 293). He blamed the “ignorant poor women,” who were the “common searchers” appointed to view dead bodies in order to prevent the concealment of violence, for making too many diagnoses of consumption based on emaciation alone. In 1855, the International Statistics Congress chose to follow Farr’s classification based on anatomical site. The first conference for the revision of the International Classification of the Causes of Death (ICD) was held in 1900 (36). Since 1948, the WHO has been entrusted with updating the International Statistical Classification of Diseases and Related Health Problems.

Surveys

Early health surveys were conducted of restricted populations to gather information on specific topics. For example, John Haygarth (1740–1827) conducted pioneer surveys in the 1770s, believing that greater knowledge of “natural and artificial variables” would suggest practical policies for disease containment and control (83). Samuel Hitch carried out one of the earliest surveys in the area of mental health, of the “pauper insane” in 1844 in Leicester (20). By 1917, a survey authorized by the U.S. federal government examined households’ use of physicians, medicines, nurses and hospitals (52). However, not until after World War II were surveys conducted on a large scale, representative of national populations. The United States performed the first National Health Survey in 1935 and has implemented health interview surveys regularly since 1957 (29). Japan started regular surveys even earlier, in 1948 (61).

Design and analysis of survey results were aided by developments in mathematics and statistics. Thomas Bayes (1706–61) was one of the first to perform statistical inference. Later, Adolphe Quetelet (1796–1874) and Francis Galton
(1822–1911) applied statistics to human populations and developed correlation and simple regression techniques. As more complex data on populations became available, statistical methods to infer population characteristics, from a subsample, began to be developed in the late-nineteenth century (80). Karl Pearson (1857–1936) argued that statistical tools allowed clinical and preventive medicine practitioners to “conduct experiments” with records already available, especially to provide insights into epidemic disease. Sir Bradford Hill (1897–1991) was most successful in getting medicine to adopt statistics as part of its regular armamentarium.

By 1960, enough countries were conducting national health surveys that the WHO commissioned a study that proposed a template of “factors bearing on the design of health interview surveys” for the benefit of other countries (61). By 1990, the WHO was seeking data for indicators of progress toward Health for All goals and pressed for greater comparability of survey data between countries (27). Differences existed between nations in terms of which subpopulations were excluded (e.g., nursing homes, prisons, military), the method of data collection, and the wording of questions that might otherwise provide data for the same indicator.

As was previously the case for mortality, people began to seek common definitions and measures of morbidity used in surveys. The WHO developed an International Classification of Impairment, Disabilities and Handicaps in 1980. The International Classification of Functioning, Disability and Health (ICF), published in 2001, now supersedes the former. The new emphasis on capacity and social participation in 2001, rather than on disability and handicap in 1980, demonstrates how classification systems evolve according to new needs, understandings, and values.

As another example, early censuses recorded information on “idiocy” or “insanity,” and later differentiated specific types of mental illness. The Diagnostic and Statistical Manual (DSM) has promoted much greater consistency in diagnosis and reporting on mental health and illness since it was first published in 1952 (92). Yet, the DSM-IV (1994) still faces questions about the validity and reliability of diagnoses. Noting the socially constructed nature of disease classification systems, Lee (62) discusses the classification of mental disorders in China and argues that international systems must not displace regional systems, which serve valuable functions locally. The advent of international quality-of-life surveys serves as the latest example of the trend toward seeking comparable data on morbidity. A recent review examined application of quality-of-life measures in developing countries, highlighting the lack of effective assessments of equivalence across countries (13).

Limits to international classifications exist. The desire to continue collecting data that are strictly comparable with past survey information leads to reluctance to adopt new classifications and definitions. Buratta & Egdili (17) consider the necessity of achieving universally acceptable definitions and questions on health interview surveys. Instead, they advocate constructing instruments that are clearly directed toward specific aims, for specific users of the data.
Administrative Data

Although administrative data are generally collected for reasons other than generating health statistics to reflect population health status, the data can be applied to the latter. Some of the oldest health administrative data come from hospitals. Johann Peter Frank (1745–1821) records that numbers of people treated in hospitals, under various diagnoses, were first incorporated to try and assess population health status in continental Europe around the turn of the nineteenth century. Florence Nightingale (1820–1910) also vigorously promoted the use of hospital statistics (65).

Interest in gathering data on entire health care systems grew with pressures from social democratic and labor parties to establish national health insurance systems in the early twentieth century, though many countries did not implement such systems until after World War II (105). The United Kingdom made use of National Health Insurance data for surveillance by 1911 (29).

With the rise of national health insurance plans in many countries and the computerization of databases, an extensive body of administrative data has become available for analysis, but many developing countries continue without publicly administered health insurance systems (29). Created in developed countries in the 1950s and 1960s, applications of computerized health information systems (HIS) in developing countries did not take off until the 1970s. A 1982 conference asked whether developing countries could afford to spend scarce resources on informatics when many millions of people did not have access to basic primary health care (38). Developers of HIS realized that “the dream of one’s own computer will not automatically provide the answer to all problems and may even give rise to new ones” (38, p. 8). The goal remains the accurate and efficient assessment of the health of populations, not computerized systems as an end in themselves.

Elements of the health care system based outside of hospitals have also developed electronic databases that aid indicator development. Following the example of the first Cancer Registry, established in 1943 in Holland, many jurisdictions now collate diverse sources of cancer records centrally (29). Other disease-specific registries have been created but there is little international consistency. Medical clinics are now able to perform more complex analysis of their records electronically, including linking patients’ primary care information with socioeconomic data (21).

The rising costs associated with publicly funded health care systems provided the largest stimulus for creating data banks of health system information to compare spending across countries (84). Poullier (1987) notes that the rising costs were likely primarily due to behavioral, cultural, and environmental factors leading to increased chronic diseases, but that a paucity of readily available data led to neglecting these “upstream” factors in formal statistical analyses (84, p. 36).

Interest in the quality and cost-effectiveness of interventions also grew. In addition to the money spent on interventions, their expected benefits in the population covered had to be estimated. Cost-benefit analyses in the health field were first crucially performed by social reformers in the 1850s, in an attempt to demonstrate that sanitary measures would save lives and thus money, through decreased loss of fit and productive laborers (26).
In the context of evaluating health care, Donebedian (33) made a fundamental distinction between health demands, health needs, and the use of health services, which should inform the design of indicators of health care system functioning. Perceptions of health needs and of the ability of interventions to improve them, as well as observations regarding the care experiences of other people, lead to individuals’ health demands. The use of health services may not meet all demands and does not (often) meet the actual health needs of patients or populations. Simply reporting the number of procedures that different populations receive may not correlate with the health outcomes achieved. For instance, the OECD’s 2003 health indicators report presents the high and escalating coronary-artery bypass and angioplasty rates in the United States with no evidence of an associated marginal decrease in mortality from coronary artery disease (77). Finally, those with greatest health needs often have the least access to care or health care coverage. Therefore, data on health care services alone cannot provide adequate indicators of population health. The conceptual framework for population health proposed in this paper (see below) points to the need to put cost, quality, and coverage of health care services in context: They are useful but inherently very “downstream” in nature.

Looking beyond health care systems, “public health systems” are intended to protect and promote health at the population level, a mandate that dates back to at least the nineteenth century. For over a century, public health administrative data have been used to calculate indicators of population health, such as vaccination coverage and incidence of certain reportable infectious diseases. However, measuring the contribution of public health systems to population health requires a broader set of indicators. Examples of indicators used by public health systems in eight different countries can be found in a recent review (1).

Woodward et al. (111) suggested a “balanced scorecard” with four domains for measuring the functioning of a public health system, in relation to the system’s overall goal to improve population health. First, “health determinants and status” would include intervention impact measures. Second, “community engagement” would measure perceptions and degree of involvement of residents, “high-risk” groups, health care providers, policy makers, and staff. Third, “resources and services” would include per capita measures, human resources data, and the proportion spent on administration. And fourth, “integration and responsiveness” would include measures of partnership, collaboration, and coordination. The National Advisory Committee on SARS pointed out that Canada is among the countries missing national public health goals, priorities, strategies, and reporting mechanisms (74), although plans are underway to remedy these gaps. In fact, administrative data for all four quadrants do not currently exist in most Canadian jurisdictions.

Non-Health (Sector) Data

Meterological data are one type of non-health sector data with a long history of measurement, at times for health purposes. Galen’s medical writings from around 170 AD included attempts to devise a standard temperature scale (84). Environmental monitoring and regulatory data continue to be related to health indicators, such as that of air quality, for population health (101).
Historically, data on income was sought, in some cases, to study its relationship with health; housing quality was sometimes used as a surrogate (82). With re-emerging interest in the societal determinants of health, researchers are now examining data collected from economic, social service, and education sectors, most notably to ascertain socioeconomic status. However, as Lynch & Kaplan report, the use of occupation, income, education, or wealth as indicators of socioeconomic status poses methodological challenges (64). For example, data in an education database in Manitoba showed small differences in success on exams at the end of secondary school between the highest and lowest socioeconomic groups. Linkage with a health database allowed researchers to follow the entire birth cohort. The difference in successful completion of grade 12 tests between highest and lowest socioeconomic groups was much more striking (77% pass rate compared with 27%) using the linked databases (16).

Both non-health and health data are used together to produce comprehensive indicators of population health. For instance, the OECD 2005 report on social indicators presents indicators to reflect self-sufficiency, equity, and social cohesion, involving financial, employment, social spending, and well-being measures, permitting comparisons among the most economically developed nations (76). Sweden provides another example. The overall national public health aim is to "create social conditions that will ensure good health for the entire population" (63) (see Box 1). The new model for monitoring the impact of policy makes use of non-health sector indicators that are correlated with health outcomes. Examples of these indicators are

- election turnout in municipal elections, as a measure of the population’s participation and influence in society; and
- how pupils are treated by teachers, other grown-ups, and fellow pupils (measured at the national level) as an indicator of secure and favorable conditions in childhood and adolescence.

Box 1. Sweden’s domains of objectives for population health (63)

1. Participation and influence in society
2. Economic and social security
3. Secure and favorable conditions during childhood and adolescence
4. Healthier working life
5. Healthy and safe environments and products
6. Health and medical care that more actively promotes good health
7. Effective prevention against communicable diseases
8. Safe sexuality and good reproductive health
9. Increased physical activity
10. Good eating habits and safe food
11. Reduced use of tobacco and alcohol, a society free from illicit drugs and doping, and a reduction in the harmful effects of excessive gambling
The Swedish example demonstrates that high-level acceptance can be achieved on the importance of collecting non-health data to measure a population’s progress toward better health for all. Future applications of a conceptual framework for population health should continue to draw on new “non-health” sector data sources.

Map Data

Maps are “permanent, graphic depictions, which link places in the world to what comes with them” (110). Mapmaking has many historical origins, mainly in societies where people lived in large groups with centralized bureaucracies. Early maps were constructed with large scales but were infrequently made, in part because there was no perceived utility for them: traditional cultures did not lack “spatial perception intelligence” or the ability to find their way (110). In the sixteenth century, maps became strikingly more common as navigation needs increased for global commerce and exploration. Small-scale maps were used as tools of governance in developing nation-states to assess taxes, wage war, facilitate communications, and exploit resources, as well as for more humanistic purposes (110).

Computer technology has prompted more widespread use of mapping (71). Presently, public health practitioners, spatial researchers, and policy analysts are increasingly applying Geographic Information Systems (GIS) and the related sciences (25, 88). Mullner et al. suggest the greatest appeal of GIS is its ability to depict clearly with maps the results of complex analyses (71).

The WHO has developed a computer application called HealthMapper, which has been successfully used to present health indicators in developing countries. Health administrators can obtain a quick, visual representation of the availability and use of specific health and social services (28). Information depicting other health services or risk factors for disease can be overlaid on the maps. Apparently, the maps have “a more visceral and direct impact” on the reader than tables or lists of numbers (28).

Qualitative Data

Although this paper has focused largely on quantitative data for indicators of population health, some indicators do incorporate qualitative information. For instance, neighborhood- or community-level characteristics that may affect health have been collected qualitatively—from the time of the eighteenth century, when they were sometimes assessed by smell, to the present (54, 96). A full discussion of qualitative data for indicators of population health is beyond the scope of this paper. However, significant downsides of quantitative approaches should be noted. Numbers in quantitative health indicators are often presented as “the truth” without acknowledgment of the assumptions, methodological and epistemological, that led to their creation. The loss of contextual information in quantitative data increases the risk of misinterpretation of meanings. In discussing attributions of causes of death in about 1840, Hamlin suggests that, “to require that the narrative history of a patient’s constitution be condensed to a single word (i.e., the “cause of death”)

was to give up the possibility of a medicine that would take an interest in, and see as problematic, the full variety of pathological influences a person encounters” (46, p. 863). Quantitative indicators can also be said to remove the “voice” or a sense of the “lived experience” of the individuals or populations that generated them. Thus, particular populations may be labeled without an opportunity to present their interpretative view. Survey instruments with closed questions and established templates for data collection provide little opportunity for gaining new insights on factors important to population health. For reasons such as these, some research and reports on population health now sometimes present quantitative and qualitative data together (59a).

INTEGRATING INDICATORS

Basket Presentation

Comprehensive consideration of the relationships between indicators can be encouraged by simply presenting different types of indicators, under different categories, in one “basket” or report. For example, the 2003 OECD report on health places indicators into these five categories: (a) health status, (b) health care resources and utilization, (c) health expenditure and financing, (d) nonmedical determinants of health, and (e) demographic and economic context. The categories, and indicators selected, reveal what the authors value as well as their underlying concepts of population health. For instance, the “nonmedical determinants of health” section consists entirely of indicators of “lifestyle” behaviors—for tobacco and alcohol consumption and body weight. Indicators of the state of the environment or social context are notably absent.

Placing indicators for different subpopulations in the same basket can highlight inequities [see Braverman’s review of measuring health disparities in this issue (8)]. However, Szreter (103) warns that data can be collected in such a way that information on subpopulations cannot be properly captured or disaggregated. WHO, OECD, and national reports on health generally present information on subpopulations by sex, age group, or health region, but not often by socioeconomic status, or ethnic group, or other category related to the social determinants of health.

Advantages of the basket approach include the straightforward use of readily available data and the ability to prompt testable hypotheses concerning relationships between indicators. However, the precise relationships between indicators are virtually always left unspecified, preventing policy makers from determining how best to maximize contributions to population health. Another downside of an unstandardized basket approach may be a tendency to choose indicators based on what will appear favorable in a report rather than what will more fully depict actual population health status. Next, linked databases are discussed as a way to quantify relationships between diverse sectors and their impacts on health.
Linked Databases

Linking health status databases with databases on other determinants of health can help increase understanding of causal pathways in both directions. Roos et al.'s (94) review suggests that record linkage promotes policy applications, when databases allow researchers to (a) study interventions longitudinally; (b) compare regions, areas, and hospitals; (c) combine information on patients and physicians; (d) add up expenditures for different services within the health-care system; and (e) examine the determinants of health using education and family services data in conjunction with health-related information. Upon reviewing possible “non-medical” datasets to monitor community well-being, Saunders et al. (97) concluded, however, that no one dataset will contain information on all the determinants of health. Furthermore, many populations will not have as complete datasets for linkage as the “information-rich environments” highlighted by Roos et al.

Multilevel Analysis

Linking data from different domains may involve individual- and-community-level information. In her 1998 landmark article on the value and methodological challenges of multilevel analysis, Diez-Roux noted that following the vague holistic notions of the causes of ill health in the early nineteenth century, and then the rise of the unicausal model of the germ theory, a “web of causation” approach developed, but with the focus remaining largely on individual-level risk factors (31). Krieger (57) asked, “Where is the spider?” implying that simply collecting information on multiple individual-level causes of ill health is not enough. Multilevel analysis enables examination of specific processes and pathways linking individual and community characteristics to good or poor health.

Risk Adjustment

An international example of risk adjustment comes from a comparative study of 14 OECD countries based on household surveys from 1996. The researchers adjusted the number of physician visits across different income groups for need (based on self-reported health status). The results showed that most countries had achieved equity. Only in Portugal, the United States, Austria, and Greece was significant inequity demonstrated (106a).

Attributable Proportions

The amount of health or ill-health attributable to an intervention or determinant of health is often of interest to decision-makers. A common approach is to calculate “population attributable risk” (62a), which is affected by the population’s prevalence of the causative exposures or determinants being examined, as well as the strength of their association with the health outcome. Such attributable proportions should be interpreted with caution as confounding variables and important more “upstream” determinants may not be measured. For instance, the recent
calculations by Allison et al. (2), which attributed nearly 300,000 deaths in a year to obesity in the United States, have been criticized for only controlling for age, sex, and smoking status, and assuming all remaining excess mortality associated with elevated BMI was causally related to obesity per se. Other researchers have shown that obesity relates to socioeconomic status (116) and race as well as to community disadvantage in multilevel analyses (90). In short, obesity itself is not an “upstream” determinant of health but rather an intermediary physiological outcome affected by the whole spectrum of cultural, economic, social, and other forces, as well as genes, that jointly influence human caloric intake and energy expenditure.

Modeling

Statistical models are increasingly used to attribute health outcomes to multiple risk factors. For example, in the INTERHEART study authors recently assessed nine variables related to the risk of acute myocardial infarction in 52 countries (95, 115). From data collected using a case-control design, incremental population-attributable risks (PAR) were calculated using logistic regression models. After adjustment for confounders, the top three population-level risk factors in all the regions of the world were (a) the ratio of blood apolipoproteins, with a PAR of 49.2% (95% confidence interval: 43.8%–54.5%); (b) current or former smoking status, 35.7% (32.5%–39.1%); followed by (c) psychosocial factors, 32.5% (25.1%–40.8%). Altogether, the nine risk factors were associated with 90.4% (88.1%–92.4%) of the risk for myocardial infarction.

However, difficulties arise in interpreting these PARs. First, the total of all the attributable fractions, as is usual in multifactorial disease, adds to more than 100%, due to complex interactions between factors, presenting a credibility hurdle to policy makers.

Second, the risk factors modeled were at various levels in the “web” of multistep causal pathways that lead from truly exogenous forces acting in society—such as cultural and economic drivers of smoking, eating and physical activity patterns, and individual, “intraorganismal”-level factors, such as blood pressure and lipid levels. Furthermore, these latter sorts of risk factors are in turn themselves a homeostatically controlled result of many different genes and aspects of environmental factors, many of which operate differentially over the life-course and are poorly understood. To their great credit, however, the investigators do compare their 52 study settings around the world for their relative degree of urbanization and hypothesize that this is a key “upstream driver” of the new coronary heart disease pandemic engulfing low- and middle-income countries.

Summary Measures

Building on data in life tables, Summary Measures of Population Health (SMPH) include information on mortality and morbidity. Joshua Milne presented the first life table with accurate data in 1815 (70). Life table methods were further developed
in the 1930s to explore how long individuals could be expected to work in their lifetimes (58a). Multistate life tables in the 1970s integrated information on states of health lying in between the pure presence and absence of a health condition, as well as the probability of moving between these states. The data for multistate life tables ideally arises from longitudinal surveys, or may be calculated from cross-sectional data using certain assumptions.

SMPH fall into two general categories: (a) health expectancies (HE) and (b) health gaps; many specific types have been summarized (45, 92). Health expectancies go beyond purely mortality-based calculations of life expectancy, to determine how long people can expect to live without certain diseases or limitations on their normal activities. Health gaps choose a norm or goal for the population (such as an average age at death of 75), and quantify the difference between the actual health of the population and that goal. Disability-adjusted life years (DALYs) remain the most commonly used health gap measure since their first publication in the World Bank’s 1993 World Development Report: Investing in Health (73, 107).

In general, methodological issues for SMPH arise from conceptual differences in approach and variations in data sources, collection, and analysis (17). Methodological challenges related to vital statistics and survey and administrative system data also apply here, since summary measures are built up from these databases.

Theoretically, SMPH can facilitate integration of the dimensions of population health, from social determinants, through proximal determinants such as physiological risk factors, because they can be adjusted to show the impact of various risk factor modifications, or hypothesized interventions, at different levels of intensity. Furthermore, impacts at different ages may be derived from the calculations in the life tables. However, there is often difficulty finding enough high-quality epidemiological data for individuals in various subgroups for all required domains (7). Currently, most calculations that break down summary measures of population health status into their contributing causes are done for rates of medically diagnosed diseases, such as coronary heart disease mortality. For this reason, McDowell et al. (68) correctly assert that SMPH describe population health status but have limitations in analytic studies examining the “whys” and the “hows”/processes that create population health. They propose a classification of indicators of population health based on four applications: description, prediction, explanation, and evaluation, which should all be reflected in indicators of population health.

A Unifying Conceptual Framework

A framework for measuring population health assembles different types of indicators together, provides an integrated picture of health, and clarifies at what level different indicators are measuring population health. As Krieger (55, 56) suggests, a conceptual framework as part of a theory helps structure ideas to explain causal connections between, within, and across specified domains. A clear conceptual framework for the selection and use of indicators may help point to the dimensions of population health of import, and lead to more balanced discussions about
what indicators should be targeted by interventions in order to impact population health. The authors offer the Canadian Institutes of Health Research—Institute of Population and Public Health (CIHR-IPPH) conceptual framework for determinants of individual and population health (Figure 1) as a basis for organizing indicators of population health. Ideally, reports on population health would present and integrate indicators from all seven key domains depicted in Figure 1.

PERSISTENT ETHICAL ISSUES

Early quantitative indicators of population health raised ethical issues of individuals’ privacy and confidentiality, and the just distribution of resources, which persist today. Concern over ethical conduct by medical practitioners dates back to at least 1700 BC, in the Code of Hammurabi, and the Hippocratic Oath in 460 BC (29). Early census enumerators were required to take an oath of secrecy (36). Legislation to protect individuals’ privacy now extends to surveys, administrative health data, and linkable databases used to construct indicators of population health, and can vary across jurisdictions. Other ethical questions, such as how to place appropriate value on different populations and health states, often become evident in the development of newer indicators (73). Recent papers examine the difficulties of extending bioethical principles to the field of population and public health (14, 53, 79). However, the developing literature on population and public health ethics points to some consensus on ethical principles that may guide the construction and use of indicators of population health: justice, transparency, and effectiveness.

First, the principle of seeking justice must guide decisions about what to measure, which is related to decisions about health goals and uses of indicators. Justice looks different, and will be reflected in different types of indicators, under different philosophical traditions (41, 91). For example, measures that focus on health outcomes, such as quality-adjusted life years (QALYs), lend themselves to utilitarian analysis. Indicators of the equity of distribution of resources are necessary for a liberal rights assessment of justice. Indicators developed based on a community’s ideas of “the good” would be supported by people with a communitarian approach. Some feminists would seek indicators of preferential support for those who provide care (91). Ethicists are still working to articulate the underlying philosophical basis for a population health approach that balances resource allocation to address the worst off, with expenditures on the rest of the population (11, 14, 106b). A human rights basis for a bioethics of population health is one option (12, 106b), though not without problems (79).

Bayer & Fairchild declare that a moral obligation exists to prevent avoidable suffering and death, a proposition with which most people pursuing population health would agree. Thus, indicators that present “preventable” fractions of disease highlight the ethical work to be done. Ethicists argue that for people to have equitable health opportunities, important social determinants of health should be as much our concern as the differences in health arising from inequities in access to health care (12, 14, 106b). Starting from agreed-upon obligations of certain
groups or authorities, to provide specific products or services may be less contested (79). Ethical measures of population health would then pointedly indicate what obligations of whom remain unfulfilled. The literature on ethics and population health also demands looking beyond state borders and creating internationally comparable indicators (79, 106b).

Second, the principle of transparency must guide the development and use of indicators of population health (22, 106). Insistence on transparency underlines the fundamental belief that people have a right to be involved in decisions affecting them. Advocates for the first systems of civil registration stipulated that the data should be available to the public—open to consultation by all citizens (26). Public communication strategies must be considered. Early Bills of Mortality were published with poetry, such as a sonnet on the subject of death by William Wordsworth, which perhaps gave the publication more popular appeal (70). Once the General Register Office was established in London, weekly and quarterly bulletins were used to heighten awareness of preventable deaths, while annual and decennial reports presented more rigorous scientific analysis (104).

Ideally, indicators promote information sharing in multiple directions. Advocates for knowledge transfer and exchange seek engagement of members of the public, policy-makers, and professionals, together with researchers, but barriers remain (23). Attention must be paid to what public participates, with those least well off included and the interests of powerful groups contained. A decision-making process open to input from different affected subpopulations can recognize the scientific expertise required for developing population health indicators. Furthermore, public participation to define public health problems and goals can increase awareness of the social determinants of health and empower people to pressure authorities to address them (18; http://www.healthy-canadians.ca). The ethical development of population health indicators in the future will involve greater dialogue with a broad range of stakeholders to reflect social values, as well as moral questions of justice.

Finally, the principle of effectiveness suggested by Childress et al. (22) highlights the ethical need to evaluate indicators (also see 53). Resources spent on developing indicators of population health must be in proportion to the evidence of expected benefits. Potential harms, such as infringement on an individual’s privacy, are also weighed against the benefits. Evaluations of the development and use of indicators may assess whether a consultative, ethical approach itself promotes health (11, 53).

EVALUATION OF INDICATORS

Have indicators of population health, and their use, improved over time? How should this be assessed? One way to determine the quality of indicators may be to compare them to a set of ideal characteristics, such as those in Box 2. The Working Party on Community Health Data and Indicators in Denmark produced
a template for evaluating indicators along these lines (112, 113). Few studies have quantified the intelligibility of different indicators for different stakeholders. Since methods exist to assess routinely the “readability” of documents (67a), standard tests of “indicator intelligibility” could be developed for routine use. The availability of data and local capacity to perform data analysis—the feasibility of constructing an indicator—often take precedence over other qualities (63, 111a). However, indicators that are merely “feasibility-driven” or “available-data-driven” are less preferable than those with an underlying clear and widely accepted conceptual framework (86). A concept-driven selection process should result in more methodologically sound indicators.

Box 2. Ideal indicators are

- Built on consensus
- Based on a conceptual framework
- Valid
- Sensitive
- Specific
- Feasible
- Reliable and sustainable
- Understandable
- Timely
- Comparable
- Flexible for use at different organizational levels

and include measures of all these aspects

- Incidence and prevalence
- Central tendency (e.g., mean, median, etc.) and distribution
- Stratification by subpopulations

Another way to assess whether indicators have improved over time is to briefly review indicators actually used by, for example, societal decision-makers. Manuel & Goel (66) first suggested the following framework when summarizing the development of indicators in the health care sector. Also, see Table 1 for a summary of the links between population health concepts, types of data, and indicators used.

The first indicators of population health were simple reports of information at the population level. Reporting of infant mortality rates (IMR) and life expectancies from the 1800s are relevant examples. Improvements after some time included the calculation of age-standardized mortality rates. Currently, many countries calculate an extension of mortality rates, potential years of life lost (PYLL), by individual causes, first suggested in the literature as an indicator to guide health planning by Romeder & McWhinnie in 1977 (93). Differences in registering practices of premature and especially very low-birth-weight infants (i.e., whether they are reported as live births or not) still affect international comparability of IMRs (77).
## TABLE 1  Key developments of population health indicators

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Types of data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reports of baptisms, marriages and deaths often from church records; weather data</td>
<td>First regular censuses, and mandatory civil registration of vital statistics</td>
<td>First international disease classification system for causes of mortality (ICD)</td>
<td>First national representative health surveys and cancer registries; national health insurance data</td>
<td>Computerized health system and other administrative data; first DSM classification</td>
<td>Linked databases, more longitudinal surveys for incidence; maps; ICF classification</td>
<td></td>
</tr>
<tr>
<td><strong>Types of indicators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality reports, by climate, area</td>
<td>Mortality rates, life expectancies</td>
<td>Standardized comparisons between populations</td>
<td>Prevalence of diseases and use of health services</td>
<td>Prevalence of risk behaviors; quality of life; health expectancies and health gaps</td>
<td>Risk-adjusted, multilevel analysis, SMPH, PARs, qualitative and quantitative indicators combined</td>
<td></td>
</tr>
</tbody>
</table>
Second, indicators of population health began to be mapped to frameworks in focused areas. For example, one can present a comprehensive picture of the prevalence of coronary heart disease (CHD), CHD mortality, DALYs lost due to CHD (81a), CHD risk factors, prevalence and CHD-related health services use, for males and females, rural and urban populations (5).

Third, indicators began to reflect the interwoven nature of population health domains (Figure 1) by more clearly creating connections between those domains. Examples of this approach include stratified reporting by subpopulations, such as cardiovascular mortality by socioeconomic status, after adjusting for known behavioral risk factors (67). Current reporting on prevalence of diseases and risk factors by “peer group” recognizes that multiple common community factors affect the health of populations in similar ways (102a).

Finally, indicators are beginning to be used to explicitly address population health goals. WHO’s 2000 World Health Report stands out in making this attempt using SMPH, though data were lacking to properly execute the task worldwide (108). Even critics agree that the addition of health outcome data to the previously predominant focus on process indicators of health care systems is a desirable step (75). Because improving population health is the ultimate goal/use of indicators, this approach to the use of indicators does seem to be an improvement over earlier approaches, although more rigorous methods are still required.

If successful use of indicators directs resources to public policies that improve population health, then indicator evaluations logically need to examine policy implementation and their effects. Evans & Stoddart reviewed the uptake of population health concepts in Canada in 2003. Lavis found that economists in particular, especially those at high levels in public policy making, have not assimilated population health concepts (60). In addition, a recent evaluation revealed that most Canadians are largely unaware of population health concepts, such as the importance of nonbiomedical determinants of health (19). The question remains, however, “Has the use of population health indicators improved the health of populations?” Quantifying the differential impact of the use of various specific indicators of population health may be a fruitful area for further research.

CONCLUSION

Population health indicators in common use today are built on mortality measures initiated in the 1500s, regular censuses started in the 1800s, civil registration of vital statistics commencing in the 1850s, regular national surveys first initiated in the 1950s, and health system and other administrative databases facilitated by computer applications used on a wide scale since the 1960s. An explicit and comprehensive conceptual framework, such as the one outlined in this chapter, will acknowledge the complexity of the determinants of health and clarify what is being measured, thus promoting comparability of indicators and the use of appropriate statistical models to analyze them. Finally, to improve indicators of population...
health, both indicators and the frameworks from which they stem must be better evaluated, to determine whether they effectively and efficiently serve their intended purposes.

The Annual Review of Public Health is online at http://publhealth.annualreviews.org

LITERATURE CITED

9. Deleted in proof
10. Burrows GM. 1818. Strictures on the Uses and Defects of Parish Registers, and Bills of Mortality. See Ref. 44
15. Deleted in proof


32. Dodson J. 1752. A letter from Mr. James Dodson to Mr. John Robertson, F.R.S. concerning an improvement of the Bills of Mortality. See Ref. 44


37. Deleted in proof


39. Deleted in proof

40. Deleted in proof


42. Fothergill J. 1768. Some remarks on the Bills of Mortality in London. See Ref. 44

43. Deleted in proof

52 ETCHES ET AL.

Bill for this Nation. Farnborough, Germany: Gregg Int. 300 pp.


70. Milne J. 1815. A Treatise on the Valuation of Annuities and Assurances on Lives and Survivorships; on the Construction of Tables of Mortality; and on the Probabilities and Expectations of Life, Vol 1. See Ref. 44
72. Deleted in proof
84. Poullier JP. 1987. OECD experiences with the initiation and coordination of health indicator systems, with special emphasis on interinstitutional coordination and comparability. See Ref. 98, pp. 23–36
87. Deleted in proof
89. Rickman J. 1802. Observations on the Results of the Population Act, 41 Geo.III. See Ref. 44
100. Deleted in proof
CONTENTS

Epidemiology and Biostatistics

Effective Recruitment and Retention of Minority Research Participants, Antronette K. Yancey, Alexander N. Ortega, and Shiriki K. Kumanyika

Measuring Population Health: A Review of Indicators, Vera Etches, John Frank, Erica Di Ruggiero, and Doug Manuel

On Time Series Analysis of Public Health and Biomedical Data, Scott L. Zeger, Rafael Irizarry, and Roger D. Peng

The Promise and Pitfalls of Systematic Reviews, Patricia Dolan Mullen and Gilbert Ramírez


Environmental and Occupational Health

Environmental Justice: Human Health and Environmental Inequalities, Robert J. Brulle and David N. Pellow

Speed, Road Injury, and Public Health, Elihu D. Richter, Tamar Berman, Lee Friedman, and Gerald Ben-David

The Big Bang? An Eventful Year in Workers’ Compensation, Tee L. Guidotti


Public Health Practice

Health Disparities and Health Equity: Concepts and Measurement, Paula Braveman

The Politics of Public Health Policy, Thomas R. Oliver


What Works, and What Remains to Be Done, in HIV Prevention in the United States, David R. Holtgrave and James W. Curran

vii