Administrative Record Linkage as a Tool for Public Health Research

Douglas P. Jutte,[1] Leslie L. Roos,[2] and Marni D. Brownell[2]

[1]School of Public Health, University of California, Berkeley, California 94720-1190; email: dpjutte@berkeley.edu

[2]Department of Community Health Sciences, Faculty of Medicine, University of Manitoba, Winnipeg, Manitoba R3E 3P5, Canada; email: Leslie_Roos@cpe.umanitoba.ca, Marni_Brownell@cpe.umanitoba.ca

Annu. Rev. Public Health 2011. 32:91–108

First published online as a Review in Advance on January 3, 2011

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

This article's doi: 10.1146/annurev-publhealth-031210-100700

0163-7525/11/0421-0091$20.00

Keywords

population health, social determinants of health, life course, socioeconomic status, longitudinal data

Abstract

Linked administrative databases offer a powerful resource for studying important public health issues. Methods developed and implemented in several jurisdictions across the globe have achieved high-quality linkages for conducting health and social research without compromising confidentiality. Key data available for linkage include health services utilization, population registries, place of residence, family ties, educational outcomes, and use of social services. Linking events for large populations of individuals across disparate sources and over time permits a range of research possibilities, including the capacity to study low-prevalence exposure-disease associations, multiple outcome domains within the same cohort of individuals, service utilization and chronic disease patterns, and life course and transgenerational transmission of health. Limited information on variables such as individual-level socioeconomic status (SES) and social supports is outweighed by strengths that include comprehensive follow-up, continuous data collection, objective measures, and relatively low expense. Ever advancing methodologies and data holdings guarantee that research using linked administrative databases will make increasingly important contributions to public health research.

INTRODUCTION

Insights into the complex web of factors influential in health (12), the importance of determining pathways and mechanisms of causation (48), and the role of life course health trajectories (5, 43) highlight the need for data capable of supporting the next stages of public health and population health research. Traditionally, the most powerful tools for unraveling these complicated questions have been large-scale, longitudinal studies. However, longitudinal collection of data is difficult to coordinate, time-consuming, and expensive. These studies are complicated by issues of loss-to-follow-up, underrepresentation of ethnic and racial minorities, reduced participation by individuals in the tails of the socioeconomic distribution, and frequent use of self-report for disease and risk exposure, as well as limitations resulting from exclusion criteria and patient consent requirements that reduce generalizabilty to the overall population (4, 91). We explore a different approach to data collection and research that could circumvent some of the thorniest of these problems and may shed new light on the complex interplay of risk and protective factors on health over time.

Linked administrative data comprise information already widely and diligently collected on large populations for other purposes, then merged at the individual level using unique, anonymized identifiers and made available for academic research. More extensive than routine vital statistics collected on birth, death, or disease and more inclusive than disease-specific registries, administrative health databases are increasingly linked to population-wide institutional data from social service agencies, educational institutions and census reports and utilized to identify interpersonal family connections within and across generations. These information-rich environments can, relatively inexpensively, supply the very large samples and long-term observations that primary data collection often cannot. And critically for epidemiologic research, in locations where population registries have been created, it is possible to determine the disposition of an entire population (the denominator) rather than that of only those interacting with a health or educational institution (73, 75). Although important issues remain regarding database governance, privacy protection, research access, and distribution of findings, these obstacles are not insurmountable (82).

Some researchers have described linked administrative information as an opportunity to create new data from existing sources. The merging of data derived from disparate authorities to include events occuring to individuals and families over time and intergenerationally, combined with the ability to do this for large populations, allows for a wide range of important and often unique public health investigations. It also provides a relatively low-cost supplement to the too few longitudinal studies critical for future public health research.

LOCALES OF MAJOR DATA LINKAGE EFFORTS

The published literature has identified several centers noted for highly productive research utilizing linked administrative data while maintaining privacy and confidentiality (36, 72, 73). These include the Oxford Record Linkage Study (ORLS), the Scottish Record Linkage System, the MigMed2 Database (Sweden), Statistics Norway, the Rochester Epidemiology Project (Minnesota), the Western Australia Data Linkage System (WADLS), and several Canadian centers, including the Manitoba Center for Health Policy (MCHP), the Center for Health Services and Policy Research (CHSPR) in British Columbia, and the Institute for Clinical and Evaluative Sciences (ICES) in Ontario. Detailed descriptions of these research centers and their investigator-initiated research and government-funded policy work have been published elsewhere (10, 29, 35, 36, 40, 45, 61, 72, 86).

More recently, New Zealand and Wales have begun data linkage operations, and a nationwide Australian effort, the Public Health

Research Network (PHRN), will soon provide that country of 21 million the world's largest population health database (24, 35). The International Health Data Linkage Network (IHDLN), comprising a substantial subset of existing health data linkage centers, was formed in 2008 to foster collaboration and exchange, describe best practices, and record output from research and programs based at its participating centers.

DATA LINKAGE METHODOLOGY

Data linkage is defined as "the bringing together of information from two records that are believed to relate to the same individual or family" (9). Linkage is achieved by using a limited set of basic sociodemographic factors, "linkage variables," to identify uniquely and reliably an individual across two or more datasets (35, 72).

Privacy and the maintenance of secure information are always of utmost concern. To achieve high-quality linkages without compromising confidentiality, multistage deidentification processes have been developed by the different research centers to address legal and political concerns without reducing the value of the data (44, 72, 93). Although the details vary across sites, to enhance privacy in most cases an outside agency serves as a linkage center. The merging of information is completed through a series of steps. In the first step, the linkage center receives files from data trustees (e.g., health or educational agencies) containing only linkage variables but no program data. In the second step, these linkage variables are used to identify individuals with information present across agency datasets. Third, a unique record number is assigned to the identified individuals. Finally, this unique record number is sent back to the data trustees, who use it to provide the data repository or researcher with the requested administrative content data but without the associated identifying information. Thus, confidentiality is maintained because neither the linkage center, researchers, or repository are ever simulaneously in possession of both the individually identifiable linkage data and programmatic content associated with that individual. Only the original agency maintains those complete records (13, 36, 45, 72).

Despite the potential for larger scale data sharing, this approach can actually enhance individual confidentiality over the traditional, investigator-initiated use of private health data (73). For example, Trutwein and colleagues (90) found that research projects utilizing name-identified health data dropped dramatically after the introduction of a record linkage system in Western Australia (**Figure 1**).

Linkage variables: information such as name, birth date, sex, address, or identification numbers that can be matched across unrelated datasets

DATA CONTAINED WITHIN LINKED ADMINISTRATIVE RESOURCES

The information available for linkage across databases varies by locale. Variation results from the length of time a center has been in existence, the type of health care financing (e.g., single- versus multipayer systems), local politics, financial resources, and often the leadership of participating agencies and data linkage centers. Using Manitoba Centre for Health Policy as an example, **Figure 2** demonstrates the range of data sources that can be linked at the individual level within a population.

Health Data

The backbone for linked administrative health systems is hospital discharge diagnoses and vital statistics on birth, death, and disease. However, a wide range of additional health information has been linked at the individual level, including pharmaceutical and immunization records, outpatient clinical diagnoses, emergency room treatment, home care, nursing records, maternal care and birth records, and cancer or other disease registries. Existing administrative data have also been successfully linked to case-control survey populations with collected biological specimens (46).

Residential Information

A second major category of data is place of residence, which serves several important purposes.

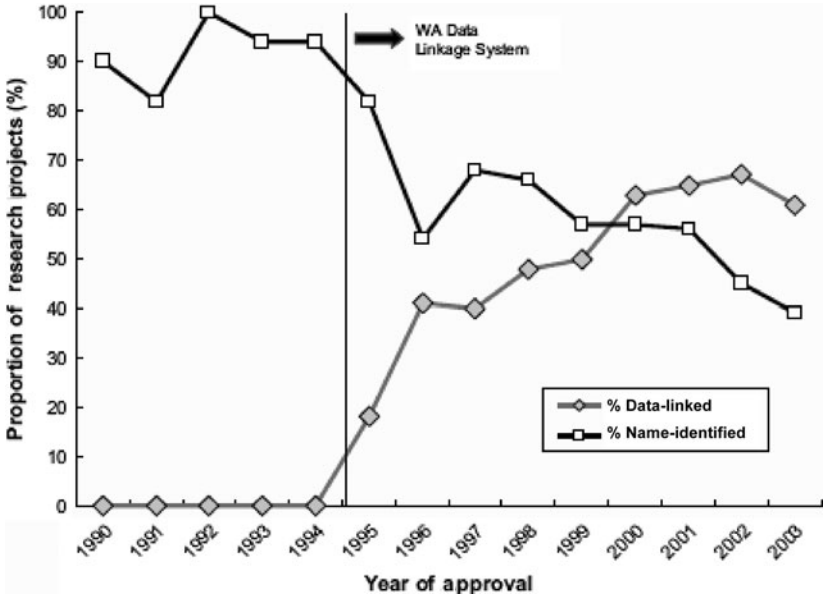

Figure 1

Proportion of ethics-approved research projects using name-identified and anonymized linked administrative data before and after the formation of the Western Australia Data Linkage System. Used with permission from Reference 90.

Mobility: a measure of the number of moves (changes in recorded address) a family or individual makes over a period of time

Length of time spent in high- or low-income areas provides valuable contextual information. Census-based information tied to region, neighborhood, or urban block face can supplement, or even replace, often limited individual-level socioeconomic data (47, 65, 84). Residential information over time provides evidence of mobility, a variable known to influence health outcomes for both children and adults (22, 53). Such mobility may be a disruptive life event if occurring during childhood (2, 33, 70). Change in residence across neighborhoods of differing socioeconomic status (SES) over time also provides information on upward or downward socioeconomic trajectories for individuals over the life course (31).

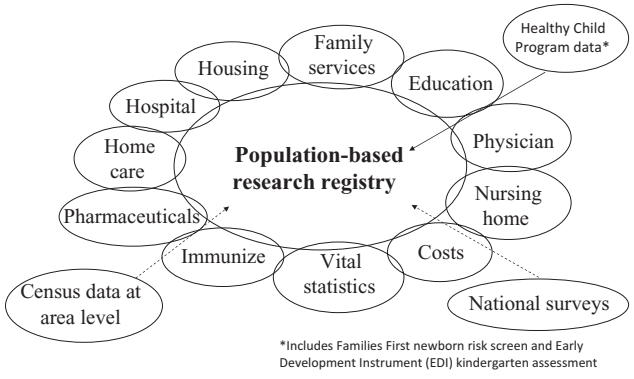

Figure 2

Example of administrative data sources available for linkage in the Manitoba Population Health Research Data Repository. Used with permission from the Manitoba Centre for Health Policy.

Family-Related Variables

Number of children in the family, birth order, parental age and marital status, and number of years living in a single-parent household have all been identified as factors important in health (10, 14, 27, 32). Changes in family structure over time can also be ascertained and are predictive of child well-being (83). Although methods for identifying family relationships vary across centers, near-universal mother-child specification is possible (enabling sibling links) and fathers have been identified in upwards of 85%

of cases (72). In some settings, more extended family links can be made to first-degree cousins and grandchildren as well, the latter allowing for transgenerational inferences (7, 10, 41).

Educational and Social Variables

The most recent categories of data available for linkage, and potentially the most important for future public health research (60, 79, 88), are markers for social and educational predictors and outcomes. Detailed, population-wide information on educational outcomes is available in a limited number of locations for children (15, 56, 58) and in even fewer locations for adults (10, 87). Educational data provide, for health studies in children, the opportunity to examine more than one domain of well-being simultaneously, e.g., Jutte et al. 2010 (41) and, for adults, the ability to account for differences in prior educational success when examining health outcomes. Similarly, population-wide data on interactions with the justice system, the child protection system, or receipt of government financial assistance can be used to identify high-risk individuals better or serve as additional examples of negative individual outcomes. Data on income assistance have been used to supplement census-based socioeconomic information with individual-level information on low-income status (41, 76). Linkages to routine screening data can provide detailed social information at the individual level. For example, the Families First program in Manitoba, Canada, screens all newborns for risks such as maternal history of prenatal smoking, alcohol use, and family stress (17). Population-wide screening of child development at kindergarten entry using the Early Development Instrument (EDI) has been incorporated into administrative data in several jurisdictions (39, 56).

Population Registry

Another important variation across centers is the existence, or not, of a population registry. Jurisdictions with identification numbers provided for every resident, such as the provincial single-payer health care systems in Canada, allow for creation of a population registry. For example, in Manitoba each resident is provided with a personal health identification number (PHIN). Because a high proportion of the residents interact with the health system annually [e.g., in childhood, greater than 98% over any four-year age period (20)], Manitoba Health is able to maintain a registry of active PHINs covering the entire population. This population registry is ideal for providing a denominator for all population-level analyses. Jurisdictions without an individually linked population registry must rely on less precise regional census information or similar data.

RESEARCH UTILIZING LINKED ADMINISTRATIVE DATA

To illustrate the potential for public health investigations, we explore some examples of research using linked administrative data.

Population-Based Cohort and Case-Control Studies

One strength of population-wide administrative data is the capacity to study low prevalence disease-disease and procedure-disease associations, associations unlikely to be funded for study using randomized controlled trials (30). For example, using a series of nested, case controls in the ORLS, Goldacre and colleagues examined links between gall bladder disease and colon cancer, appendectomy and inflammatory bowel disease, vasectomy and prostate disease, and mumps immunization and aseptic meningitis (30, 49, 62). Similarly, using Swedish MigMed data, Sundquist and colleagues (80) have examined Graves' disease and risks of several cancers.

Such matched set analysis can be done for any comparison of choice (23). For example, to examine the relationship between residential mobility and schizophrenia, Lix and colleagues (53) created three cohorts matched on age, sex, and urban/rural residence. One group

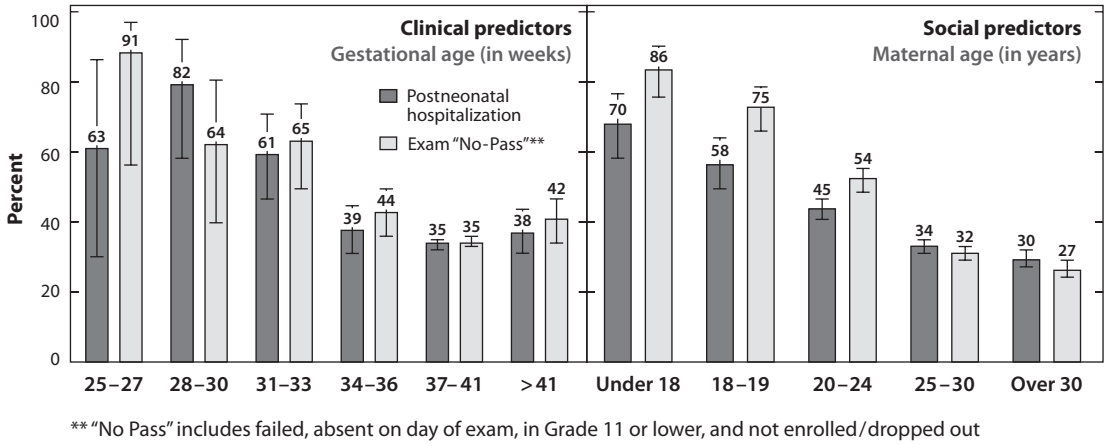

Figure 3

Proportion of poor health and educational outcomes associated with early-life clinical and social predictors: gestational age and maternal age. Used with permission from Reference 41 (complete study comparison results available as **Supplemental Figure 1**; follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**).

had diagnosed schizophrenia. The other two served as comparisons with severe physical illness (inflammatory bowel disease) or no substantial mental/physical problems. With this approach, they showed consistent determinants of residential mobility across all groups (marital status, income, use of physician services) and demonstrated increased residential mobility for patients with schizophrenia.

Multiple and Overlapping Outcome Domains

Data linked from disparate sources provide the capacity to examine outcomes from different domains within the same cohort of individuals. For example, using a regional birth cohort of children, Jutte et al. (41) examined both medical and educational outcomes in relation to several biological and social risk predictors present at birth. **Figure 3** shows rates of hospitalization and scholastic failure by gestational age and maternal age (complete data online, **Supplemental Figure 1**. Follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**). Similarly, in a study examining risks to children of adolescent mothers, in a population cohort of 32,000 children, researchers evaluated concurrent medical (hospitalization, mortality, high hospital use), educational (mean ninth-grade scores, twelfth-grade exam passage), and social outcomes [need for foster care, income assistance as a young adult, teen pregnancy (42)]. Such an approach allows for an enhanced assessment of overall well-being.

Supplemental Material

Data from multiple sources can also be linked to assess overlap in risk factors or service utilization in a population. Brownell and colleagues (16) examined nearly 12,000 children to determine the degree to which families were receiving services from multiple agencies. **Figure 4** demonstrates surprisingly little overlap among the three risk groups considered: family receipt of income assistance, protective care or social services assistance, or teen mother. Whereas only 4% of the children fell into all three risk groups, fully 31% received services from one or more agencies.

Population-Based Prediction of Disease

Using health survey data linked to administrative health services data, ICES researchers in Ontario, Canada, developed and validated an

algorithm for population-based prediction of diabetes (78). The Diabetes Population Risk Tool (DPoRT) accurately predicts diabetes risk in a population using self-reported measures available in routine population health surveys rather than collating only the detailed clinical data used for individual diabetes risk assessment. This population-level approach to estimating disease incidence allows for improved, lower-cost population health planning and enhanced assessment of the impact of illness prevention strategies.

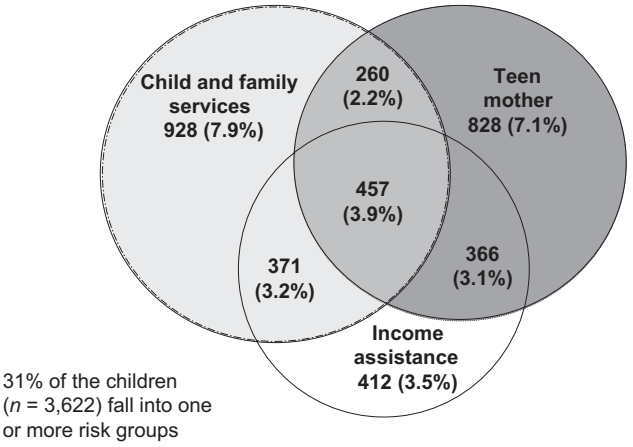

Figure 4

Overlapping risks for a population of 11,703 children; 1984 and 1985 birth cohorts, Winnipeg, Canada (16).

The Inclusion of Individuals Not Receiving Services

In public health research, knowledge of the denominator—the total population—can be critical. In locales with a population registry (e.g., Sweden, Norway, and Manitoba, Canada) or locations utilizing multiple data sources capable of capturing nearly the entire populace (e.g., Western Australia), the population denominator is available. A study of school children completed by MCHP researchers vividly illustrates the importance of the denominator (15). In **Figure 5**, the left graph illustrates the passing rate for all twelfth-grade students who took the compulsory language arts exam. A clear gradient is present across neighborhood SES quartiles; 75% of the lowest-SES children passed versus 92% of highest-SES children. In contrast, the right set of columns shows the disposition of all 18-year-olds from the 1984 birth cohort who should have taken the exam, not just those in school the day of the exam. When accounting for the children who had missed the exam, or had fallen behind or withdrawn from school, the actual passing rate among the lowest-SES group dropped to 27%, just one-third of the 77% passing rate in the highest-SES neighborhoods.

Likewise, linked data can help investigators determine the extent to which a program or screening tool reaches the entire target population. For example, the province of Manitoba screens all newborn infants for health and social risk using the Families First screen (originally called BabyFirst) to stratify children into no-risk and at-risk categories. However, the population registry also identifies a critical third category of children: those born but not screened. Brownell et al. (17) showed that although there existed a threefold variation across health districts in percentage of infants (born 2002) meeting at-risk criteria (8%–24%), in some jurisdictions nearly three-quarters of infants had never been screened (see **Supplemental Figure 2** for details). Subsequent analysis, again using data linkage, found that among children requiring foster care by age two, 42% had been among those not screened. Combined, these findings provided powerful evidence for policy makers that screening efforts were, in some cases, inadequate and that these missed children represented a particularly high-risk group meriting additional resources.

Supplemental Material

Chronic Disease Surveillance and Improved Prevalence Estimates

Case definition using a single data source alone (e.g., physician office claims or hospital discharge diagnoses) is generally inadequate for surveillance purposes (55). However, using data from more than a single source improves the necessary sensitivity and specificity for cost-effective chronic disease surveillance

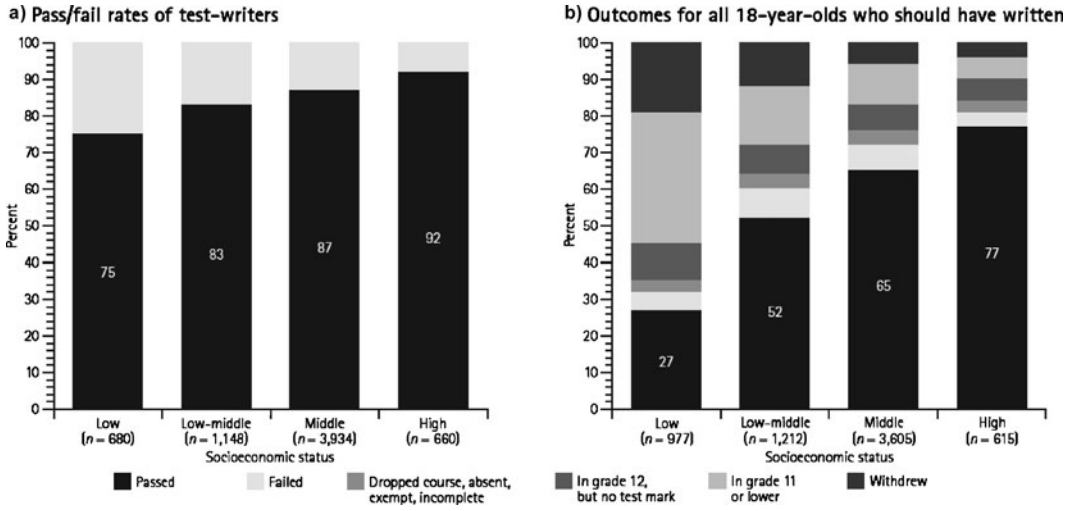

Figure 5

Grade 12 performance on language arts test by socioeconomic status, Winnipeg, Manitoba, 2001–2002. Used with permission from Reference 15.

(6, 54). For example, links to prescription data can improve the capture of conditions with disease-specific medications such as diabetes and asthma. The eventual linkage of gold-standard clinical data or laboratory evidence will further improve case identification at the population level (54).

Longitudinal administrative data linkage can also improve estimates of disease prevalence and health service utilization. For example, in any given year ~7% of children in Manitoba are involved with Child and Family Services. However, over a 10-year period, more than 17% of children required intervention (77). Information such as this has been used successfully to create a sense of policy urgency and to provide a perspective on the true proportion of health disparities and population health risks (77).

Life-Course and Transgenerational Investigations

Another feature of administrative data that is difficult to achieve in other settings is the enhanced capability to evaluate exposure effects over the life span because data are collected continuously over time. For example, to assess the impact of low-income housing on later life chances, Oreopoulos (67) linked home address data identifying children in public housing to information on long-run labor market outcomes (eventual earnings, unemployment likelihood, and welfare participation) extending into their thirties. Similarly, Jutte et al. (42) used existing data to follow children of teen mothers into early adulthood and assessed outcomes at different points over the life course.

The addition of parent-to-child family links also allows for transgenerational analyses. For example, using nationwide population data in Sweden, Sundquist, Li, and colleagues examined the differential impact of maternal and paternal factors on the transmission of coronary heart disease (85), risk of small-for-gestational-age status (52), and risk of schizophrenia and other psychotic disorders (51).

LIMITATIONS AND STRENGTHS OF ADMINISTRATIVE DATA

Roos et al. (72) published a detailed list of the attributes associated with linked administrative data compared with those of a well-respected example of longitudinal primary data

collection, the Panel Study on Income Dynamics (PSID). See **Table 1**. However, a few limitations and strengths of administrative data merit specific attention.

One important limitation of administrative data is the frequent lack of individual-level SES information. Traditional SES measures of income, wealth, achieved educational level, or job description often are not recorded at the population level, Scandinavian databases being the notable exception (8, 11). However, several researchers have shown that census data can provide a fair approximation of family or household levels of income (47, 65, 84). Others have successfully linked survey data on a subset of the population to partially replicate analyses with individual-level data and, thus, correlate the findings (74). Composite indices of census-derived data have also shown promise in approximating household income

Table 1 Comparing longitudinal primary data and population-based administrative data. Used with permission from Reference 72

Characteristics	Longitudinal primary data (Panel Study on Income Dynamics)	Population-based administrative data (Manitoba and other sites)
Number of cases	Several thousand or smaller	Often more than one million
Cost	High on a per-person basis	Very low on a per-person basis
Representativeness	Often national	Often from a province or state
Population studied	Subjects sampled and tracked	Built on a registry of an entire population
Research design	Often complex designs needed to increase power and control costs	Given a population, complex designs can be imposed retrospectively as needed
Record linkage	Useful in some contexts	Critical to check data quality and expand scope of information sources
Individual follow-up	Before and after an event	Before and after an event
Coverage and loss to follow-up	Nonresponse and differential attrition possible	Differential attrition possible
Updating	New data must be collected and merged with existing data	Multifile information must be cleaned and merged with existing data. Cleaning relies on record linkage
Time	Information must be collected (typically annually or at longer intervals)	Information provided at relatively short intervals (from daily to annually)
Place	Information at time of study (historical reconstruction possible)	Detailed information usually provided close to date of move
Longitudinal	Goes back many years	Goes back many years
Neighborhoods	Flexible construction from postal code or census area	Flexible construction from postal code or census area; large *N* may permit flexible assignment to generate nearest neighbors
Life events	Collected as part of design	Possibly available from registry or other sources
Family and intergenerational data	Collected as part of design; sibling and intergenerational studies facilitated	Assessing a family composition at any point in time possible; sibling and twin studies facilitated
Limits	Important information likely to be collectable for entire sample	Important information may be missing or available only for a subpopulation
Variables	Defined by researchers; scaling possible	Defined by others for administrative purposes; creating meaningful variables may be very time-consuming; scaling possible
Intellectual history	Scope of data collection often expanded to provide a rich data set	Scope expanded beyond initial health care data by receiving files from other agencies

levels. For example, the socioeconomic factor index (SEFI) is a validated measure derived from principal components analysis that utilizes neighborhood-level unemployment rate, rate of high-school completion, percentage of lone parent households, and female workforce participation to provide a summary score for any residential address (59). Similar efforts in Western Australia resulted in the design and use of the socio-economic indices for areas (SEIFA) (58, 89). Such indices can also be used as a contextual community measure. For example, Swedish researchers have incorporated a neighborhood deprivation index into multilevel analyses of coronary heart disease (87, 94).

Data on social supports and interpersonal relationships are another type of information not often available in population-wide administrative data. Studies have shown these to be important components of health and well-being (19, 37). Race/ethnicity is also often unavailable in administrative data. However, at least in children, published epidemiologic work has shown that minority status and ethnicity are inconsistently related to poor health outcomes, particularly after accounting for SES factors (66, 69). Combined, these limitations remain important issues to overcome for research attempting to to disentangle the roles of SES, discrimination, social supports, and neighborhood effects on population health.

Administrative data do have a number of important strengths, however. For example, such data are free from many of the measurement and loss-to-follow-up problems associated with longitudinal surveys (4, 72). The flexibility of continuously collected administrative data presents some advantages over surveys dependent on expensive, repeated interviews of the same respondents at fixed intervals (72). Researchers are able to study the health or health care utilization of a population for any length of time and for any year or set of years (72). This flexibility can be useful for natural experiments such as abrupt economic changes or the implementation of new policies. Roos et al. (73) highlight a number of capabilities for administrative data pertinent to health policy analysis. These include the ability to compare regions, areas, and hospitals; study policy interventions longitudinally; combine information related to physicians and their patients; sum expenditures across different services within the health care system; and examine social determinants of health using education and family services data in conjunction with health-related information. The latter was demonstrated in several of the examples provided above.

IMPORTANT CHALLENGES FACING ADMINISTRATIVE DATA LINKAGE

Issues of privacy and confidentiality continue to be among the most prominent concerns facing the collection and use of administrative health data. In a 2006 issue of *BMJ* dedicated to this topic, Davies & Collins (21) noted that "interpretations of legislation [related to topics of health data use] seem to have been driven less by careful consideration of the likelihood of real harm for individuals than by the desire to minimize the risk of criticism for organisations." They argued that, in fact, the overinterpretation of protective legislation could represent a risk to the public in its own right. The British Academy of Medical Sciences (1) noted there is little evidence that the use of confidential records in medical research has caused serious harm, and they further stated that advances in the fields of public health and population health have been increasingly inhibited by inappropriate constraints laid down by confusing legislation as a result of an undue emphasis on privacy and autonomy.

A related controversial issue is whether informed consent is required from patients prior to the use of personal data within administrative databases. Although deliberate care is required to create standards that protect privacy (81, 82), the rigid use of a "consent or anonymize" policy can be detrimental to research in terms of scientific opportunity, time, and financial resources (1). For example, Al Shahi & Warlow (3) argue that an "authorization bias" is introduced by the consent/authorization

process, which limits potential uses of the data and makes generalizabilty uncertain. Tu et al. (91) reported in the *New England Journal of Medicine* that in a clinical stroke registry requiring consent, a substantial bias in costs and death rate (a threefold lower in-hospital death rate for consented patients) occurred, even in a setting with specialized assistance available for the consent process. These authors argued for the need for legislation that would permit waivers of informed consent for minimal-risk observational research. Others believe that in the case of publicly funded health care, a social obligation exists for patients to permit use of their deidentified information without consent to allow the health care system to be monitored and improved for the benefit of all (3, 92).

The form in which administrative data is maintained is another decision that confronts groups creating or managing linked information and researchers planning to use the data. Some locales maintain a data repository of information collected periodically from a number of contributing agencies, e.g., the MigMed2 database in Sweden or the Manitoba Population Health Research Data Repository. However, the degree of data integration maintained on an individual varies. For example, the Manitoba repository does not maintain longitudinal, complete data files on each individual. Rather, the Repository partitions sociodemographic and content data until requested for a specific research project. One impact of the repository approach is the need, in most cases, for the investigator to work with a dedicated and specially trained programmer or to be on-site to access the data.

Alternatively, some centers use a virtual repository, where primary data custodians (government departments or ministries) maintain direct control of their data and do not provide copies to a third party for storage and access. For example, Western Australia and PHRN (35) maintain only demographic data with which they use a linkage algorithm to request specific data from custodians only upon request from researchers with approved projects. Although this approach is seen by some to be at lower risk for breach of confidentiality, it requires greater computing power to recreate the linkages from scratch with each new project.

NEXT STEPS, NEW RESEARCH

Information-rich environments based on longitudinal administrative data provide an opportunity for both methodological improvements and new research directions.

Design and Analysis

Linked data can support better research design for both academic and policy research. Work by Mosteller & Tukey (64) has proven particularly helpful in allowing messy categorical data to be combined so as to generate reasonably normalized distributions. In Manitoba, this has facilitated building and validating several indices of educational performance (based on the type of information in **Figure 5**). Such population-based indices support multivariate analyses of a variety of outcomes.

Multilevel modeling seems especially well-suited for administrative data with different types of variables measured at different levels. For example, individuals can be nested within families nested within particular neighborhoods (26). Such modeling can account for (*a*) the dilemma of unit of analysis, (*b*) the lack of independence among observations (e.g., children in the same family), and (*c*) unobserved, confounding variables (57). Using this approach with administrative data, the influence of social environmental factors can be compared with that of individual-level factors (eg., 58, 87, 94), although the relationships among variables at different levels also suggest comparing sibling with neighbor correlations (67).

Sibling and Twin Analysis

With birth cohorts constructed from multiple years of administrative data, sibling and twin designs are possible (7, 11). Statistical analyses (e.g., family fixed-effects regression and

multilevel modeling) controlling for unobserved factors that lead to sibling correlations are particularly important when widely used measures such as individual household income and parental education are not generally available on a population-wide basis, as is the case with most administrative data (7, 63, 68). Although not true experiments, such analyses of sibling data represent one of the most powerful nonexperimental approaches (63). The future should see more sibling/parent designs linking parental and child histories to assess the effects of such conditions as attention-deficit/hyperactivity disorder on subsequent offspring (71).

Life Course Research

Life course research typically examines aspects of family background that may affect health and other outcomes into adulthood. Strohschein et al. (83) used the longitudinal nature of the Manitoba population registry to characterize family circumstances supportive of (or detrimental to) well-being over time. Building on this study, ongoing work is directed toward construction of a different history for each child using significant events (parental death, parental mental health issues, divorce, residential mobility, change of school) and the age at which each occurred. The ability to link such experiences to subsequent outcomes while accounting for interfamily differences is a powerful resource.

Several Australian states and Canadian provinces are collecting population-wide data using the Early Development Instrument, a holistic measure of children's developmental status at kindergarten (25, 37, 56). An increasing interest in early life suggests the need for empirical estimates of the effects of various factors at different stages in child development (34); as linkages are performed and these younger cohorts age, the analytical possibilities will multiply.

Life course research merges into intergenerational research, raising many questions of interest across disciplines. In Western Australia, the Family Connections Genealogical Project has utilized the linkage of births, midwives, and marriage data to the core system. More than 10 research projects are now under way, analyzing the genealogy of up to three generations (28).

CONCLUSION

Investigators are engaged in many efforts to increase the scale and scope of administrative data links. As mentioned previously, efforts are now under way to expand the extensive work done in Western Australia to incorporate the remaining Australian states into the PHRN, thus encompassing the 21 million inhabitants of the continent. The United Kingdom is moving forward in similar efforts to merge hospital to hospital and hospital to death records for the entire population of 51 million (M.J. Goldacre, personal communication).

In addition to enlarging the scale of administrative data, work is now ongoing in multiple locales to increase the scope of information. In Manitoba, for example, databases on subsidized housing and from the Justice Ministry are being added to the Population Health Research Data Repository. Many sites are interested in links to clinical, radiological, and laboratory results. The incorporation of routine blood screens on newborns, pregnant women, and adults could also present new research opportunities. Genetic data—requiring collection only once over the lifetime—could also theoretically be incorporated, although an even higher level of scrutiny for privacy/confidentiality protections would be required.

The increase in scale and scope has allowed for expanded interdisciplinary research. Population-based studies seem especially suitable for combining the perspectives of economics with its "search for the closest approximation to an experiment" and those of social epidemiology with a focus on the "accretion and consistency of a broad range of research and evidence" (38, p. 12). The potential for increasing links among the work of researchers in public health and population health, education,

clinical medicine, housing, community development, and justice holds great promise.

The significance of these new insights has also influenced policy development. A review of projects using Western Australia's linked data showed consequential policy and clinical practice reforms (13). An assessment of the impact of the Manitoba Center for Health Policy led to similar conclusions (50). Influences on projects as diverse as scaling back rural hospital construction plans, moving to a policy on the generic substitution of drugs, and focusing health planners on the mismatch between population need and services delivered were all recognized as positive outcomes from research utilizing the population data repository. Advancing research methodologies and increasing holdings of linked health and social data in ever more locales guarantee continued important contributions to investigations in public health and the social determinants of health and disease.

SUMMARY POINTS

1. Linked administrative databases are powerful resources that provide longitudinal health and social data on large populations for flexible and relatively low-cost investigation of pressing public health concerns.
2. Well-established research centers in several jurisdictions around the world have developed and implemented methods that achieve high-quality data linkages for conducting health and social research while maintaining individual privacy and confidentiality.
3. Information-rich repositories link individual-level longitudinal data from a variety of sources, including data on health services, population and disease registries, place of residence information, family-related social and structural variables, educational outcomes, and use of social services.
4. Linking events for large populations of individuals across disparate sources and over time permits a range of research possibilities including the capacity to study low-prevalence exposure-disease associations, different domains within the same cohort of individuals, individuals within a population who do not use services, chronic disease surveillance, population-based disease prediction, and life course and transgenerational patterns in health.
5. Common limitations of administrative data include a lack of individual-level measures of socioeconomic status, social supports, and nonfamilial interpersonal relationships.
6. Strengths of administrative data include comprehensive follow-up, inclusion of underrepresented racial/ethnic and socioeconomic groups, reduced reliance on self-report, flexibility of study period due to continuously collected data, and relatively low expense because information is routinely collected for other purposes.
7. Concerns regarding privacy, confidentiality, and informed consent continue to challenge the use of linked administrative data despite rigorous security procedures and demonstrated public benefit.
8. Emerging methodologies and expanding data holdings guarantee that research using linked administrative databases will continue to make important contributions to the public health arena.

FUTURE ISSUES

1. The increasing use of sibling and twin analyses will allow for greater accounting of contextual variables unmeasured in population-wide data.
2. Multilevel modeling merits increased use in future studies to better utilize the strengths of administrative data derived from different levels (individual, family, school, community).
3. Sibling/parent research designs should be expanded to support the incorporation of multigenerational data into outcome analyses.
4. Further work is needed to improve the statistical manipulation of multisource, categorical data to generate population distributions more amenable to analysis and sensitivity testing.
5. Future efforts must continue toward creating links to exisiting population-wide clinical data such as radiological findings, laboratory results, and routine screening tests (e.g., newborn and prenatal blood screens).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Acad. Med. Sci. 2006. *Personal Data for Public Good: Using Health Information in Medical Research*. London: Acad. Med. Sci.
2. Adam EK, Chase-Lansdale PL. 2002. Home sweet home(s): parental separations, residential moves, and adjustment problems in low-income adolescent girls. *Dev. Psychol.* 38(5):792–805
3. Al Shahi R, Warlow C. 2000. Using patient identifiable data for observational research and audit. *BMJ* 321:1031–32
4. Atherton K, Fuller E, Shepherd P, Strachan DP, Power C. 2008. Loss and representativeness in a biomedical survey at age 45 years, 1958 British birth cohort. *J. Epidemiol. Community Health* 62(3):216–23
5. Ben-Shlomo Y, Kuh D. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int. J. Epidemiol.* 31:285–93
6. Bernatsky S, Joseph L, Pineau CA, Belle P, Hudson M, Clarke AE. 2009. Scleroderma prevalence: demographic variations in a population-based sample. *Arthritis Rheum.* 61(3):400–4
7. Bjorklund A, Jantti M, Solon G. 2005. Influences of nature and nurture on earnings variation: a report on a study of various sibling types in Sweden. In *Unequal Chances: Family Background and Economic Success*, ed. S Bowles, H Gintis, M Osborne Groves, pp. 145–64. Princeton, NJ: Princeton Univ. Press
8. Bjorkland A, Lindahl M, Plug E. 2006. The origins of intergenerational associations: lessons from Swedish adoption data. *Q. J. Econ.* 121(3):999–1028
9. Black C, Roos LL. 2005. Linking and combining data to develop statistics for understanding the population's health. In *Health Statistics: Shaping Policy and Practice to Improve the Population's Health*, ed. DJ Friedman, EL Hunter, RG Parrish II, pp. 214–40. New York: Oxford Univ. Press
10. Black SE, Devereux PJ, Salvanes KG. 2005. Why the apple doesn't fall far: understanding intergenerational transmission of human capital. *Am. Econ. Rev.* 95(1):437–49
11. Black SE, Devereux PJ, Salvanes KG. 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. *Q. J. Econ.* 122(1):409–39

12. Boyce WT. 2006. Symphonic causation and the origins of childhood psychopathology. In *Developmental Psychopathology*, Vol. 2 *: Developmental Neuroscience*, ed. D Cicchetti, DJ Cohen, pp. 797–819. Hoboken, NJ: Wiley. 2nd ed.
13. Brook EL, Rosman DL, Holman CDJ. 2008. Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. *Aust. J. Public Health* 32:19–23
14. Brooks-Gunn J, Duncan GJ, Britto PR. 1999. Are socioeconomic gradients for children similar to those for adults? Achievement and health of children in the United States. In *Developmental Health and the Wealth of Nations: Social, Biological and Educational Dynamics*, ed. DP Keating, C Hertzman, pp. 94–124. New York: Guilford
15. Brownell M, Roos NP, Fransoo R, Roos LL, Guevremont A, et al. 2006. Is the class half-empty? Socioeconomic status and educational achievement from a population-based perspective. *IRPP Choices* 12(5):1–30
16. Brownell M, Roos NP, MacWilliam L, Leclair L, Ekuma O, Fransoo R. 2010. Academic and social outcomes for high-risk youths in Manitoba. *Can. J. Ed.* In press.
17. Brownell M, Santos R, Kozyrskyj A, Roos N, Au W. 2007. *Next Steps in the Provincial Evaluation of the BabyFirst Program: Measuring Early Impacts on Outcomes Associated with Child Maltreatment*. Winnipeg: Manitoba Cent. Health Policy
18. Deleted in proof
19. Cohen S. 1988. Psychosocial models of the role of social support in the etiology of physical disease. *Health Psychol.* 7(3):269–97
20. Currie J, Stabile M, Manivong P, Roos LL. 2010. Child health and young adult outcomes. *J. Hum. Resour.* 43(3):517–48
21. Davies C, Collins R. 2006. Confidentiality and consent in medical research: balancing potential risks and benefits of using confidential data. *BMJ* 333:349–51
22. Dong M, Anda RF, Felitti VJ, Williamson DF, Dube SR, et al. 2005. Childhood residential mobility and multiple health risks during adolescence and adulthood: the hidden role of adverse childhood experiences. *Arch. Pediatr. Adolesc. Med.* 159:1104–10
23. Fedson DS, Wajda A, Nicol JP, Hammond GW, Kaiser DL, Roos LL. 1993. Clinical effectiveness of influenza vaccination in Manitoba. *JAMA* 270(16):1956–61
24. Fuller EL. 2008. *International Health Data Linkage Centres: Findings of a Fay Gale Fellowship. School of Population Health*. Perth, WA: Univ. West. Aust.
25. Furler L, Engelhardt D. 2009. *Department of Education and Children's Services data linkage opportunities in South Australia*. Adelaide, South Australia: Dep. Educ. Child. Serv., Gov. of S. Aust.
26. Gelman A, Hill J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge Univ. Press
27. Ginther D, Pollak RA. 2004. Family structure and children's educational outcomes: blended families, stylized facts, and descriptive regressions. *Demography* 41(4):671–96
28. Glasson EJ, de Klerk NH, Bass AJ, Rosman DL, Palmer LJ, Holman CDJ. 2008. Cohort profile: the Western Australian genealogical project. *Int. J. Epidemiol.* 37:30–35
29. Goldacre MJ, Griffith M, Gill LE, Mackintosh A. 2002. In-hospital deaths as fraction of all deaths within 30 days of hospital admission for surgery: analysis of routine statistics. *BMJ* 324(7345):1069–70
30. Goldacre M, Kurina L, Yeates D, Seagroatt V. 2000. Use of large medical databases to study associations between diseases. *QJM* 93:669–75
31. Hango DW. 2006. The long-term effect of childhood residential mobility on educational attainment. *Sociol. Q.* 47(4):631–64
32. Haveman R, Sandefur G, Wolfe B, Voyer A. 2004. Trends in children's attainments and their determinants as family income has increased. In *Social Inequality*, ed. KM Neckerman, pp. 149–88. New York: Russell Sage Found.
33. Haynie DL, South SJ, Bose S. 2006. Residential mobility and attempted suicide among adolescents: an individual-level analysis. *Sociol. Q.* 47(4):693–721
34. Hertzman C, Boyce T. 2010. How experience gets under the skin to create gradients in developmental health. *Annu. Rev. Public Health* 21:329–47

35. Holman CDJ, Bass AJ, Rosman DL, Smith MB, Semmens JB, et al. 2008. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA Data Linkage System. *Aust. Health Rev.* 32(4):766–67
36. Holman CDJ, Bass AJ, Rouse IL, Hobbs MST. 1999. Population-based linkage of health records in Western Australia: development of the health services research linked database. *Aust. J. Public Health* 23(5):453–59
37. House JS, Landis KR, Umberson D. 1988. Social relationships and health. *Science* 241(2865):540–45
38. House JS, Schoeni RF, Kaplan GA, Pollack H. 2008. The health effects of social and economic policy: the promise and challenge for research and policy. In *Making Americans Healthier: Social and Economic Policy as Health Policy*, ed. RF Schoeni, JS House, GA Kaplan, H Pollack, pp. 3–26. New York: Russell Sage Found.
39. Janus M. 2007. The early development instrument: a tool for monitoring children's development and readiness for school. In *Early Child Development—From Measurement to Action. A Priority for Growth and Equity*, ed. ME Young, LM Richardson, pp. 141–55. Washington, DC: World Bank
40. Ji J, Hemminki K, Sundquist J, Sundquist K. 2010. Ethnic differences in incidence of type 1 diabetes among second-generation immigrants and adoptees from abroad. *J. Clin. Endocrinol. Metab.* 95(2):847–50
41. Jutte DP, Brownell M, Roos NP, Schippers C, Boyce WT, Syme SL. 2010. Rethinking what is important: biologic versus social predictors of child health and educational outcomes. *Epidemiology* 21:314–23
42. Jutte DP, Roos NP, Brownell M, Briggs G, MacWilliam L, Roos LL. 2010. The ripples of adolescent motherhood: social, educational and medical outcomes for children of teen and prior teen moms. *Acad. Pediatr.* 10(5):293–301
43. Keating DP, Hertzman C, eds. 1999. *Developmental Health and the Wealth of Nations: Social, Biological, and Educational Dynamics*. New York: Guilford. 406 pp.
44. Kelman CW, Bass AJ, Holman CDJ. 2002. Research use of linked health data—a 'best practice' protocol. *ANZ J. Public Health* 26:251–55
45. Kendrick SW, Douglas MM, Gardner D, Hucker D. 1998. Best-link matching of Scottish health data sets. *Methods Inf. Med.* 37(1):64–68
46. Kozyrskyj AL, HayGlass KT, Sandford AJ, Paré PD, Chan-Yeung M, et al. 2009. A novel study design to investigate the early-life origins of asthma in children (SAGE study). *Allergy* 64(8):1185–93
47. Krieger N. 1992. Overcoming the absence of socioeconomic data in medical records: validation and application of census-based methodology. *Am. J. Public Health* 82(5):703–10
48. Krieger N. 1994. Epidemiology and the web of causation: Has anyone seen the spider? *Soc. Sci. Med.* 39(7):887–903
49. Kurina LM, Goldacre MJ, Yeates D, Seagroatt V. 2002. Appendicectomy, tonsillectomy and inflammatory bowel disease: a case-control record linkage study. *J. Epidemiol. Community Health* 56:551–54
50. Lewis S, Martens PJ, Barre L. 2009. Estimating the return on investment from health services research: a theoretical and empirical analysis. In *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*, ed. Panel on Return on Investment in Health Research, pp. A-21–40. Ottawa, ON: Can. Acad. Health Sci.
51. Li X, Sundquist J, Sundquist K. 2007. Specific familial risks of psychotic disorders and schizophrenia: a nation-wide epidemiological study from Sweden. *Schizophr. Res.* 97(1–3):43–50
52. Li X, Sundquist J, Sundquist K. 2010. Parental occupation and risk of small-for-gestational-age births: a nationwide epidemiological study in Sweden. *Hum. Reprod.* 25(4):1044–50
53. Lix LM, DeVerteuil G, Walker JR, Robinson JR, Hinds AM, Roos LL. 2007. Residential mobility of individuals with diagnosed schizophrenia: a comparison of single and multiple movers. *Soc. Psychiatry Psychiatr. Epidemiol.* 42:221–28
54. Lix LM, Yogendran MS, Leslie WD, Shaw SY, Baumgartner R, et al. 2008. Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. *J. Clin. Epidemiol.* 61:1250–60
55. Lix LM, Yogendran MS, Shaw SY, Burchill C, Metge C, Bond R. 2008. Population-based data sources for chronic disease surveillance. *Chronic Dis. Can.* 29(1):31–38
56. Lloyd JEV, Hertzman C. 2009. From kindergarten readiness to fourth-grade assessment: longitudinal analysis with linked population data. *Soc. Sci. Med.* 68(1):111–23

57. Ma X, Ma L, Bradley KD. 2008. Using multilevel modeling to investigate school effects. In *Multilevel Modeling of Educational Data*, ed. AA O'Connell, DB McCoach, pp. 59–110. Charlotte, NC: Inf. Age
58. Malacova E, Li J, Blair E, Leonard H, de Klerk N, Stanley F. 2008. Association of birth outcomes and maternal, school, and neighborhood characteristics with subsequent numeracy achievement. *Am. J. Epidemiol.* 168:21–29
59. Martens P, Frohlich N, Carriere K, Derksen S, Brownell M. 2002. Embedding child health within a framework of regional health: population health status and sociodemographic indicators. *Can. J. Public Health* 93(Suppl. 2):S15–20
60. Mechanic D, Tanner J. 2007. Vulnerable people, groups, and populations: societal view. *Health Aff. (Millwood)* 26(5):1220–30
61. Mekel M, Shortt SED. 2005. Coming of age and taking stock: the state of academic health policy research centres in Canada. *Healthc. Policy* 1(1):140–50
62. Miller E, Goldacre MJ, Pugh S, Colville A, Farrington P et al. 1993. Risk of aseptic meningitis after mumps vaccine in children. *Lancet* 341:979–82
63. Moffitt R. 2005. Remarks on the analysis of causal relationships in population research. *Demography* 1:91–108
64. Mosteller F, Tukey JW. 1977. *Data Analysis and Regression. A Second Course in Statistics*. Reading, MA: Addison-Wesley
65. Mustard CA, Derksen S, Berthelot JM, Wolfson M. 1999. Assessing ecologic proxies for household income: a comparison of household and neighbourhood level income measures in the study of population health status. *Health Place* 5:157–71
66. Newacheck P. 1994. Poverty and childhood chronic illness. *Arch. Pedatr. Adolesc. Med.* 48:1143–49
67. Oreopoulos P. 2003. The long-run consequences of living in a poor neighborhood. *Q. J. Econ.* 118(4):1533–75
68. Oreopoulos P, Stabile M, Walld R, Roos LL. 2008. Short, medium, and long term consequences of poor infant health: an analysis using siblings and twins. *J. Hum. Resour.* 43:88–138
69. Pamuk E, Makuc D, Heck K, Reuben C, Lochner K. 1998. *Socioeconomic Status and Health Chartbook, Health, United States, 1998*. Hyattsville, MD: Natl. Cent. Health Stat.
70. Pianta RC, Early D. 2001. Turnover in kindergarten classroom membership in a national sample. *Early Educ. Dev.* 12(2):239–52
71. Ray GT, Croen LA, Habel LA. 2009. Mothers of children diagnosed with attention-deficit/hyperactivity disorder: health conditions and medical care utilization in periods before and after birth of the child. *Med. Care.* 47:105–14
72. Roos LL, Brownell M, Lix L, Roos NP, Walld R, MacWilliam L. 2008. From health research to social research: privacy, methods, approaches. *Soc. Sci. Med.* 66:117–29
73. Roos LL, Menec V, Currie RJ. 2004. Policy analysis in an information-rich environment. *Soc. Sci. Med.* 58(11):2231–41
74. Roos LL, Walld R, Uhanova J, Bond R. 2005. Physician visits, hospitalizations, and socioeconomic status: ambulatory care sensitive conditions in a Canadian setting. *Health Serv. Res.* 40(4):1167–85
75. Roos NP, Black C, Frohlich N, De Coster C, Cohen MM, et al. 1996. Population health and health care use: an information system for policy makers. *Milbank Q.* 74(1):3–31
76. Roos NP, Brownell M, Guevremont A, Fransoo R, Levin B, et al. 2006. The complete story: a population-based perspective on school performance and educational testing. *Can. J. Educ.* 29(3):684–705
77. Roos NP, Roos LL, Brownell B, Fuller EL. 2010. Enhancing policymakers' understanding of disparities: relevant data from an information-rich environment. *Milbank Q.* 88(3):382–403
78. Rosella LC, Manuel DG, Burchill C, Stukel TA, PHIAT-DM team. 2010. A population-based risk algorithm for the development of diabetes: development and validation of the Diabetes Population Risk Tool (DPoRT). *J. Epidemiol. Community Health*. In press; doi: 10.1136/jech.2009.102244
79. Schroeder SA. 2007. Shattuck lecture. We can do better—improving the health of the American people. *N. Engl. J. Med.* 357(12):1221–28
80. Shu X, Ji J, Li X, Sundquist J, Sundquist K, Hemminki K. 2010. Cancer risk in patients hospitalised for Graves' disease: a population-based cohort study in Sweden. *Br. J. Cancer* 102(9):1397–99

81. Slaughter PM, Collins PK, Weisbaum KM, Laupacis A, Roos NP, Williams JI. 2004. *Harmonizing research and privacy: standards for a collaborative future. Final workshop summary, May 2004*. **http://www.cihr-irsc.gc.ca/e/24065.html**
82. Souhami R. 2006. Governance of research that uses identifiable personal data will improve if the public and researchers collaborate to raise standards. *BMJ* 333:315–16
83. Strohschein L, Roos NP, Brownell M. 2009. Family structure histories and high school completion: evidence from a population-based registry. *Can. J. Sociol.* 34(1):83–103
84. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. 2006. Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: a multilevel analysis of Massachusetts births, 1989–1991. *Am. J. Epidemiol.* 164(9):823–34
85. Sundquist K, Li X. 2006. Differences in maternal and paternal transmission of coronary heart disease. *Am. J. Prev. Med.* 30(6):480–86
86. Sundquist J, Johansson S, Yang M, Sundquist K. 2006. Low linking social capital as a predictor of coronary heart disease in Sweden: a cohort study of 2.8 million people. *Soc. Sci. Med.* 62(4):954–63
87. Sundquist K, Malmström M, Johansson SE. 2004. Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden. *J. Epidemiol. Community Health* 58:71–77
88. Syme SL. 2008. Reducing racial and social-class inequalities in health: the need for a new approach. *Health Aff.* (*Millwood*) 27(2):456–59
89. Trewin D. 2004. Technical paper: Census of Population and Housing: socio-economic indexes for areas (SEIFA), Australia 2001. *ABS No. 2039.0.55.001*, Aust. Bur. Stat., Canberra
90. Trutwein B, Holman CDJ, Rosman DL. 2006. Health data linkage conserves privacy in a research-rich environment. *Ann. Epidemiol.* 16(4):279–80
91. Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, et al. 2004. Impracticability of informed consent in the registry of the Canadian Stroke Network. *N. Engl. J. Med.* 350(14):1414–21
92. Upshur RE, Morin B, Goel V. 2001. The privacy paradox: laying Orwell's ghost to rest. *CMAJ* 165:307–9. Erratum 165:888
93. Weisbaum KM, Slaughter PM, Collins PK. 2005. A voluntary privacy standard for health services and policy research: legal, ethical and social policy issues in the Canadian context. *Alta. Health Law Rev.* 14(1):42–46
94. Winkleby M, Sundquist K, Cubbin C. 2007. Inequities in CHD incidence and case fatality by neighborhood deprivation. *Am. J. Prev. Med.* 32(2):97–106

Annual Review of Public Health

Volume 32, 2011

Contents

Symposium: Determinants of Changes in Cardiovascular Disease

Epidemiology and Biostatistics

Cardiovascular Disease: Rise, Fall, and Future Prospects
Russell V. Luepker 1

Proportion of the Decline in Cardiovascular Mortality Disease due to Prevention Versus Treatment: Public Health Versus Clinical Care
Earl S. Ford and Simon Capewell 5

Social Determinants and the Decline of Cardiovascular Diseases: Understanding the Links
Sam Harper, John Lynch, and George Davey Smith 39

Sodium Intake and Cardiovascular Disease
Alanna C. Morrison and Roberta B. Ness 71

Environmental and Occupational Health

Public Health Practice

Social Environment and Behavior

Health Services

Indexes

Errata

An online log of corrections to *Annual Review of Public Health* articles may be found at http://publhealth.annualreviews.org/