

Towards a Linkage of Census, Health, and Environmental Data

Sumi Mehta, M.P.H.

Environmental Health Sciences, School of Public Health
University of California at Berkeley

18th Population Census Conference
East-West Center
Honolulu, Hawaii
August 26-29, 1998

The United Nations Principles and Recommendations for Population and Housing Censuses (1) suggests that all countries collect information on a common set of characteristics, in addition to information which may be nation-specific. As a result, national censuses are increasingly becoming part of a global census. For example, all countries are encouraged to collect information on demographic and social characteristics such as sex, age, marital status, and citizenship. If relevant, countries may also collect information on suggested topics such as religion, language, or ethnicity.

Moreover, the increased availability of digitized census data means increased ability to convert data into information for research, planning, and policy. Therefore, making the most of the 2000 round census is not the sole responsibility of data creators. Users must also understand the benefits and limitations of using census data.

After providing a brief overview of the benefits and limitations of using census data, this paper will address some of the key methodological issues related to data linkage. This paper will then attempt to develop a working set of assumptions that are necessary for meaningful linking of census, environmental and health surveillance data.

Benefits of Using Census Data

- The census is the only source of data that attempts to obtain a complete count of the population.
- It provides detailed information on socio-economic factors (age, sex, educational attainment), as well as environmental factors (housing, sanitation, access to clean water).
- Information is provided for very small sub-populations, namely tracts, districts, or provinces. Spatial trends or differences can thus be easily detected.

- The census is a low cost source of information. Most researchers lack the funding to collect the wide variety of information included in the census. In addition, the independent collection of data already gathered by the census is not necessarily useful when research focuses on the population level.

Limitations of Census Data

- Only a limited set of information can be collected by the census.
- While multiple census results can be analyzed, the census only provides a cross-sectional view of the population.
- Confidentiality of data prohibits linking at the individual level.

Population-based data linkage in developing countries

In developing countries, where a great proportion of the burden of disease is due to diarrhea, malaria, pneumonia, and malnutrition (2), the ability to link environment and health data could have great consequences on health policy and resource allocation. This is especially true given that different populations have widely different social, environmental, and economic characteristics which result in different disease burdens.

One of the purposes of the Global Burden of Disease (GBD) (2) project is to estimate the fraction of mortality and disability attributable to major risk factors, including environmental risk factors, such as poor water supply, sanitation, and air pollution. GBD does not link environment, population, and health data; existing estimates of the relative risk associated with these various factors are used, in conjunction with information on the prevalence of exposure in the various regions, to get a sense of the attributable risk associated with the factors in each region.

However, rather than starting with assumptions about relative risk, data on health outcomes and relevant covariates from population-based surveillance systems can be used to gain information on the risks associated with environmental exposures. Population based methods of surveillance can provide information on the prevalence of multiple health outcomes. While these only offer a cross-sectional perspective, regular, periodic surveys can be a useful component of a data linkage project. If studies are designed so that information on environmental exposures and health outcomes can be linked to census data at high spatial resolution, the census can be an invaluable source of information on a wide range of covariates.

A working set of assumptions for environment and health data linkage

A working set of assumptions for environment and health data linkage is presented below. While this list is clearly not exhaustive, it can serve as the starting point for future endeavors in the linking of environment and health data.

1. Known or suspected / biologically plausible environmental pathway: Ecological bias could occur when group level associations do not reflect associations at the individual level. Since health survey data can only be used to make cross-sectional linkages of environment and health data, there is even greater potential for spurious, or inaccurate associations to occur in the face of differential confounding by region.
2. Short duration of time between exposure and outcome: Exposures which have a long duration before health outcomes can be detected cannot be meaningfully linked to health data collected during the same time frame.
3. Robust estimates of exposure levels: There is often great within person and within region variability in exposures. However, exposure levels must be robust enough so differences

between regions can be detected. Not all sources of surveillance data allow for time series analysis of changes in a single region with time. Estimates should therefore represent exposures which do not fluctuate much over time, or exposures where the short term variability does not impact health.

4. Exposure must provide a reasonable estimation of dose: Researchers have noted that water quality data is not directly associated with human health effects, as most water quality does not sample the water people actually consume (3). The right sort of environmental data must be linked to health data in order to develop accurate environmental health indicators.
5. Adequate treatment of confounders: While population based characteristics are sometimes more meaningful at the group level, bias can occur as a result of the grouping of individuals into regions. Bias can also occur due to risk factors which are distributed unevenly across groups. Confounding at the group level does not necessarily have the same structure as individual confounding.
6. Exposure should be uniformly distributed, or reflect the region as a whole (not just a simple average of different sub-regions): Estimates of exposure must be somehow representative of the population covered. This is especially important to keep in mind when dealing with large geographic regions.

Developing environmental health indicators

Environmental health indicators can be established once a linkage between environmental and health data has occurred. These indicators could form the basis of environmental health management and decision making (4). Developing reliable environmental health indicators is a complex matter. Linking of environment and health data results in the combined effects of

uncertainty in individual susceptibility (related to the health outcome data) and individual exposure or co-exposures (related to the environmental exposure data).

In addition, it is often difficult to link health outcomes with the time frame where environmental exposures actually played a role in the development of the health outcome. Researchers in the field of occupational health monitor sentinel health events, such as asbestosis and pesticide poisoning, that are likely due to acute environmental exposures in the work environment (4). However, environment-related diseases, such as cancer and respiratory disease, are often the result of chronic environmental exposures taking place over long periods of time.

The importance of integrated surveillance systems, especially in the arena of environmental health, cannot be overemphasized. Surveillance systems, by definition, should monitor for conditions which have public health interventions (5). Surveillance systems for health must continue to focus on outcomes. However, as long as a vast majority of diseases and illnesses are related to common environmental pathways, such as air or water, integrated systems of surveillance which include environmental exposures could potentially address the wide range of health outcomes which could be controlled by addressing a common environmental source.

The census can be an invaluable component of integrated environmental health surveillance systems. As mentioned above, studies must be carefully designed so that census data can be linked to other data sources at high spatial resolution. As access to census data continues to improve and expand on a global scale, health researchers and other census 'users' must be aware of the nature and form of data collected by the census. After all, improvements in the creation and dissemination of census data, which are the responsibility of the census 'creators', should result in the increased conversion of this data into useful information, which is the responsibility of the census 'users'.

References

1. United Nations, Department of Economic and Social Affairs, Statistics Division. Principles and Recommendations for Population and Housing Censuses. New York: United Nations, 1998.
2. Murray CJL, Lopez AD. The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: The Harvard School of Public Health, 1996
3. Kolsky PJ, Blumenthal UJ. Environmental health indicators and sanitation-related disease in developing countries: limitations to the use of routine data sources. *World Health Stat Q* 1995;48:132-9.
4. Kjellstrom T, Corvalan C. Framework for the development of environmental health indicators. *World Health Stat Q* 1995;48:144-154.
5. Teutsch SM, Churchill RE. Principles and Practice of Public Health Surveillance. New York: Oxford University Press, 1994.