United States
Environmental Protection
Agency

Health Effects Research
Laboratory
Research Triangle Park NC 27711

Research and Development

EPA-600/S1-84-017 Dec. 1984

EPA

Project Summary

Adjustment of Incidence Rates for Migration in Indirect Ecologic Studies

Chin Long Chiang and Paul M. Conforti

The overall objective of this research program was to develop a method for adjusting incidence rates for migration in studies relating environmental agents to diseases with long latent periods. Various methods of estimating migration and population change are considered.

An example of a situation requiring this adjustment is described. Cancer incidence rates were compared for census tracts with varying levels of asbestos in drinking water. Because cancer has a long latent period, recent in-migrants would not have been exposed for sufficient periods of time to be at risk for cancer. Unless the in-migrants were equally distributed across census tracts, an analysis of the relationship between asbestos and cancer based on incidence rates would be biased.

This report reviews a number of measures of migration and population change as well as stochastic models of migration and of population growth. The stochastic models of migration include models that are time-independent and time-dependent. They vary in complexity from a simple in-migrant model to one in which in-, out-, and within-migration are included. The stochastic models of population growth extend this work to include birth and death considerations.

Migration data available through the Census of Population and Housing of the Bureau of the Census are described. A method is developed that uses these data to estimate migration by census tract. This method is applied to data from a project supported by the U.S. Environmental Protection Agency (EPA) on the relationship between ingested abestos and cancer. A reanalysis of the data with the addition of migration information is compared to the original results.

This Project Summary was developed by EPA's Health Effects Research Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

The problem of migration arises in studying the relationship between a risk factor and disease of a long latent period by the indirect method. The indirect method uses groups of individuals as the observational units to compare risk factor presence or magnitude with disease outcome (morbidity or mortality). In studying a disease with a long latent period, unless groups are closed to migration, risk factor presence or risk factor exposure times of individuals within these groups will vary and effect the observed relationship.

An example in which the migration problem arises is an investigation supported by the U.S. Environmental Protection Agency (EPA) and carried out at the University of California, Berkeley, on the relationship between ingested asbestos and cancer. In this example, the study area was the five-county San Francisco/Oakland Standard Metropolitan Statistical Area (SMSA), which included 722 census tracts, of the 1970 Census Bureau Census tracts were compared for their cancer incidence rates over 3- and 6-year periods (1969-71 and 1969-74, respectively), and asbestos levels in drinking water supplies. Because cancer has a long latent period, it is important to know if the population of the study area had changed in the years preceding the collection of cancer incidence data. If there was an increase in the study area population such that census tracts grew in a uniform manner, then, although the relationship observed may be diluted, it would not be biased. Population of the San Francisco/Oakland SMSA by county and decade (for 1950, 1960, and 1970) is shown in Table 1.

Table 1. Population of San Francisco/Oakland SMSA by County and Decade for 1950, 1960, and 1970

County	1950	1960	1970
Alameda	740,315	908,209	1,073,184
Contra Costa	298,984	409,030	558,389
Marin	85,619	146,820	206,038
San Francisco	775,357	740,316	715,674
San Mateo	235,659	444,387	556,234
Total	2,135,934	2,648,762	3,109,519

During 1950-1970, the SMSA population grew from 2.1 million to 3.1 million. Furthermore, this growth is clearly not uniform across the counties and, therefore, not uniform across the census tracts of the study area.

Exposure time to asbestos levels in census tracts varies and may yield a biased view of the actual relationship between asbestos and cancer.

The present study reviews the methodology of measuring migration, presents stochastic modelling of population growth, and suggests a method for estimating census tract migration from available data.

The review of the asbestos-cancer study includes descriptions and sources of the variables analyzed. Asbestos levels in drinking water were determined from a water sampling plan throughout the SMSA. Samples were analyzed for asbestos by a well-developed method of electron microscopy. Cancer incidence rates were obtained by merging data on cancer incidence for the years 1969-1971 collected under the Third National Cancer Survey and data in the 1970 Census of Population. Race, sex, and site specific rates were age-adjusted by the indirect method using the entire SMSA population as the standard. Thirty-five cancer site and cancer site groupings were analyzed. Data on covariables such as socioeconomic status (median family income and median school years completed), marital status, and asbestos workers were also taken from the 1970 Census of Population.

A description of the observed cancer cases and study population shows the assumptions made about migration in the original asbestos-cancer study and the assumptions necessary to accurately portray the population at risk. The changes these assumptions make on the indirect age-adjusted rates are given.

A review is made of non-stochastic migration measures. The measures include direct methods, which involve data on mobility and prior residence from the census or surveys, and indirect methods, which require estimating net migration (the difference between in- and out-migrants) from population figures at two censuses or from natural increase (births minus deaths) or intercensal survival rates derived from life tables or comparison of age distributions of successive censuses. These indirect methods are called estimation by the "residual method." The difference between total change in population and change due to natural increase is imputed to net migration.

Because direct methods are fairly straightforward when the proper questions are asked in the census or surveys, the concentration here is on the indirect methods. The indirect methods of measuring migration presented are (1) the national growth rate method, which uses data from two censuses, and (2) the residual method, comprising (a) the vital statistics method, which requires complete registration of births and deaths in intercensal periods, and (b) the survival rate method, which requires census data with survival rates obtained from either life tables or censuses.

Procedures

The stochastic models of migration that are presented begin with a simple time independent process for the probability of observing k in-migrants in an area during a time interval (O,t). This results in a Poisson process. This process is extended to a system of Poisson processes that might represent the probabilities of observing varying numbers of in-migrants in the 722 census tracts of the San Francisco/Oakland SMSA. Since it is more realistic to assume that the probability of migrating depends on time, the system of Poisson processes is modified to include this assumption.

A model which yields the probability of observing k individuals in an area while allowing for both in- and out-migration is developed. It is assumed that the probabilities of in- and out-migration are dependent on time and the probability of out-migration is also dependent on the number of persons in an area at a particular time. The result is a process in which the population in the area of interest at time t is the sum of two random variables. One of these random variables is binomial and represents the number of survivors of an initial number of people from time O and the other random variable is Poisson and represents the total number of surviving immigrants in (O,t).

A process is presented that includes parameters for births, deaths, and migration. These parameters are time independent. The probability of having k individuals in the area of interest at time t is the sum of two random variables, one of which is negative binomial and one of which is unnamed.

The most realistic stochastic model is one in which birth, death, and migration are considered in a linear growth, time-dependent process. In this model the 722 census tracts and the area outside of the San Francisco/Oakland SMSA are the areas of interest. The parameters include possibilities for increases and decreases within the areas of interest and allows for individuals to move from one area to another. Each time-dependent parameter is multiplied by the current population of respective areas to yield a linear growth model. The differential equation for the probability generating function of the number of people in an area cannot be solved explicitly.

When the aforementioned system of Poisson processes is modified to allow parameters to depend on the number of individuals in an area such that growth is considered linear, the resulting process is the time-dependent Yule process. An unnamed process results when the probability of in-migration is dependent on time and dependent on the number of individuals in an area in a non-linear way. Models are also presented which include parameters for in- and out-migration and parameters for in-, out-, and within-migration. A model is discussed that extends the in-, out-, and within-migration process to include births and deaths.

Conclusions

Unless special surveys are made to obtain information about migration pertaining to census tracts, data from the Census of Population and Housing from the Bureau of the Census of the U.S. Department of Commerce must be used to estimate migration at the census tract level.

Under the reports of the Bureau of the Census there are two variables that relate to the subject of migration or population mobility. Under the Population portion of the 1970 census report is an item entitled "Residence in 1965." Residence on April 1, 1965, is the usual place of residence five years before enumeration. The category "same house" includes all persons five-years-old and over who did not move during the five years as well as those who had moved, but by 1970 had returned to their 1965 residence. The category "different house" includes persons who, on April 1, 1965, lived in the United States in a different house from the one they occupied on April 1, 1970, and for whom sufficient information concerning the 1965 residence was collected. These persons were subdivided into three groups according to their 1965 residence in or outside a standard metropolitan statistical area: "in central city of this SMSA," "in other parts of this SMSA," and "outside this SMSA." The category "abroad" includes those with residence in a foreign country or outlying area of the United States in 1965.

The Housing Characteristics portion of the 1970 Bureau of the Census reports lists a variable entitled "year moved into unit." Data on year moved into unit are based on the information reported for the head of the household. The question refers to the year of the latest move. Thus, if the head of the household moved back into a unit he had previously occupied or if he moved from one apartment to another in the same building, the year he moved into his present unit was to be reported. "Year moved into unit" was reported in five categories: 1968 to March 1970, 1965 to 1967, 1960 to 1964, 1950 to 1959, and 1949 or earlier.

A procedure is presented for estimating migrants by census tract using the 1970 census item "year moved into unit." The assumptions necessary for making this estimation are discussed.

Chin Long Chiang and Paul M. Conforti are with the University of California, Berkeley, CA 94720.
***Judy A. Stober** is the EPA Project Officer (see below).*
The complete report, entitled "Adjustment of Incidence Rates for Migration in Indirect Ecologic Studies," (Order No. PB 85-124 139; Cost: $11.50, subject to change) will be available only from:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officer can be contacted at:
Health Effects Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

☆ U S GOVERNMENT PRINTING OFFICE, 559-016/7862

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

BULK RATE
POSTAGE & FEES
EPA
PERMIT No. G-3

Official Business
Penalty for Private Use $300