Epidemiological Bulletin

Pan American Health Organization:
Celebrating 100 Years of Health

Vol. 24, No. 4 **December 2003**

On the Estimation of Mortality Rates For Countries of the Americas

Introduction

It is well known that statistics derived from registered mortality can be affected during any of the phases in their production: from collection of data and completion of forms, coding, data processing, to their subsequent enumeration. Indicators produced from this information (such as numbers of death and distribution of cases by cause) that have a role in the creation of rates can be altered in both the numerator and the denominator. Therefore, knowledge of the environment in which mortality statistics are produced and the problems that arise when producing them is indispensable for their correct interpretation and use. This knowledge allows for application of procedures to correct problems and improve the quality and credibility of the statistics.

Errors in collecting and processing databases can also give rise to problems that can be apparent only when data comparisons and their trends are studied. This implies a certain degree of knowledge in the field and a regular use of data. Estimation of rates requires a denominator that corresponds to the population by age groups on the one hand and to the registered live births, which are a part of maternal and child mortality rates, on the other hand. The population estimate for inter-census years is taken from projections, which could inadequately represent migration problems faced by some countries. Live births statistics also have some problems, the most important of which is extemporaneous registration of births. Consequently, observed maternal and child mortality rates will differ from actual rates if late registration of births and non-registration of births and deaths are not accounted for.

The quality of cause-specific mortality data is also affected by limitations in current medical knowledge, diagnostic errors, deficiencies of certification, and perhaps to a lesser extent, coding and other processing errors. The validity of the distribution by cause also is affected by under-registration of deaths. Cause of death certification, even when done by attending physicians, is often incomplete or of low quality for reasons such as lack of training on proper certification and insufficient understanding of the uses made of the information provided on the death certificate. Another problem frequently encountered is that physicians may prefer certain kinds of diagnoses, such as the ones in their specialty area; this bias may vary from country to country and over time. In many developing countries a sizable segment of the population lacks access to medical care. Consequently, non-attending physicians, who may have insufficient information for a diagnosis, may sign death certificates and non-medical witnesses may provide death reports. Both developing and developed countries face some of the same problems. For example, legal, societal, and other reasons may lead to the underreporting of causes of a sensitive nature, such as suicide or HIV/AIDS, on the death certificate. Moreover, physicians often do not understand how to adequately fill out the death certificate, especially in relation to the identification of direct, intervening, and underlying causes. Furthermore, the selection of a single underlying cause of death is often problematic in elderly decedents, who often suffer from several chronic diseases that concurrently lead to death.

Clearly, there is a real need to educate the public, physicians, and health sector decision-makers about both the im-

IN THIS ISSUE . . .

portance of accurate and complete reporting on the death certificate and the impact of erroneous reporting on aggregate mortality statistics. Practices differ from country to country as to whether deaths without medical certification are included or not on tabulations of deaths by cause. A World Health Organization (WHO) provision specifies that when deaths without medical certification constitute less than 2% of the total, they should be included in such tabulations under the category "ill-defined cause;" when they exceed this percentage, they should be tabulated separately. Countries sometimes apply different criteria, however. Deaths without medical certification are sometimes included in the national cause of death tabulations as follows: under codes 798.9 [International Classification of Diseases, Ninth Revision (ICD-9)][1] or R98 (ICD-10)[2], "unattended death," when the cause of death is not external but is unknown due to the lack of medical care at death or during the illness or condition leading to death; or under codes 799.9 (ICD-9) or R99 (ICD-10), "other unknown and unspecified cause of mortality". For medically certified cause of death data, the simplest indicator of quality is the proportion of deaths assigned to "symptoms, signs, and ill-defined conditions" (SSI), codes 780-799 (ICD-9) and R00-R99 (ICD-10). The "unknown" causes of death assigned to 798.9 and R98, or 799.9 and R99 account for a large proportion of deaths attributed to SSI, since most of these are without medical certification. Where registration coverage is incomplete, however, the proportion of deaths assigned to SSI will usually increase as coverage increases, without there having been a real drop in the quality of medical cause of death certification. In fact, under both ICD-9 and ICD-10, the magnitude of the proportion of deaths assigned to SSI is a lower bound estimate on the proportion of deaths from ill-defined causes, because a number of "defined" ICD-9 and ICD-10 categories, such as cardiac arrest and heart failure, lack diagnostic meaning. It should also be noted that deaths from "defined" causes are not necessarily "well" defined; they are subject to diagnostic, certification, and coding errors that cannot be detected after statistics are compiled. For most countries the proportion of deaths assigned to the category SSI, in combination with the proportion of deaths certified by attending and non-attending physicians, is useful for monitoring trends and differentials in access to medical care. Table 1 shows, by country, the total number of registered deaths and the percentage of deaths assigned to SSI around 2000 (or for the latest 3 data years available). In 21 countries of the Americas, less than 5.0 % of registered deaths were assigned to SSI around 2000.

Effect of the change of ICD revisions on mortality data

The introduction of the Tenth Revision of the ICD in the Americas, starting in 1996, marked the most sweeping changes in the Classification since the Sixth Revision was introduced in 1949 and reflects a conceptual shift in structure and content from previous revisions. Although each revision has produced some breaks in the comparability of cause of death statistics, the change from the Ninth Revision, in use since 1979, to the Tenth Revision, has had many consequences on the coding of mortality. The ICD-10 has considerably greater detail than ICD-9 (almost twice the number of codes); and includes shifts of inclusion terms and titles from one category, section, or chapter to another; new cause of death titles and corresponding cause of death codes and sections; regroupings of diseases; and changes in the coding rules to select the underlying cause of death. All of these result in a number of discontinuities in the comparability of cause of death statistics over time or in historical series. These discontinuities are best assessed at the national level from the analysis of the results of double-coding (or bridge-coding) studies on national data and observing comparability ratios.

Comparability ratios are derived from the dual classification of the underlying cause of death on mortality records for a single year, classified under the new revision and under the previous revision. They are calculated by dividing the number of deaths for a selected cause classified under the new revision by the number of deaths to the most comparable cause classified under the previous revision. A ratio of 1.0 indicates that the same number of deaths was classified to a particular cause or combination of causes regardless of the revision used; it does not necessarily mean that the cause was unaffected by changes in classification and coding procedures but that there was no net change. A ratio greater than 1.0 indicates that more deaths were assigned to a cause in ICD-10 than the comparable cause in ICD-9 and a ratio less than 1.0 indicates fewer deaths were assigned to a cause in ICD-10 than the comparable cause in ICD-9.

Completeness of Data

In many countries of the Americas, the coverage of the civil registration system is incomplete, and in some countries the population covered by available mortality data needs to be further clarified. Within countries, the completeness of registration is known to vary according to geographic area and age group. Registration of vital events is less complete in rural areas than in cities and, in general, is worse in areas with poor living conditions. Table 1 shows the estimated under registration of deaths in countries of the Americas

Table 1: Status of Death Registries in Countries of the Americas, around 2000 (last three years available)

Country	Last three years available	Cumulative registered deaths	Symptoms, signs and ill-defined causes around 2000 (%)	Crude death rate (per 1,000 pop.)		Estimated underregistration (%)
				registered	estimated	
Anguilla	1993-1995	169	30.2	7.2	7.2	-
Antigua	1993-1995	1,360	8.7	6.9	6.9	-
Argentina	1999-2001	852,632	6.6	7.7	8.0	3.9
Bahamas	1997,99,00	4,870	1.4	5.4	7.5	27.6
Barbados	1993-1995	7,327	3.0	9.3	9.1	-
Belize	1998-2000	4,073	3.8	6.1	6.1	-
Bermuda	1992-1994	1,468	0.7	8.3	...	...
Brazil	1998-2000	2,814,072	14.8	5.6	6.9	18.7
Canada	1998-2000	655,683	1.3	7.2	7.2	0.4
Chile	1997-1999	240,713	4.6	5.4	5.5	2.0
Colombia	1997-1999	529,448	3.0	4.3	5.8	24.6
Costa Rica	2000-2002	45,557	1.6	3.7	3.8	2.6
Cuba	1999-2001	235,357	0.7	7.0	7.2	2.1
Dominica	1992-1994	1,657	12.4	7.6	7.6	-
Ecuador	1998-2000	166,698	13.3	4.5	6.0	25.3
El Salvador	1997-1999	87,146	16.4	4.8	6.0	20.2
United States of America	1998-2000	7,132,006	1.2	8.5	8.4	-
Grenada	1994-1996	2,162	7.4	7.8	...	...
Guadeloupe	1997-1999	...	...	6.0	6.0	1.1
Guatemala	1997-1999	202,758	9.6	6.2	7.2	13.4
French Guiana	1997-1999	...	...	4.0	3.8	-
Guyana	1994-1996	14,293	2.3	6.4	8.2	21.8
Haiti	1997, 1999	13,250	44.7	0.8	10.6	92.1
Cayman Islands	1998-2000	382	1.8	3.4	...	...
Turks and Caicos Islands	1998-2000	156	6.5	3.1	...	...
Virgin Islands (US)	1998-2000	1,915	1.1	5.3	5.2	-
Virgin Islands (UK)	1996-1998	...	...	4.5	...	...
Jamaica	1989-1991	35,543	12.9	5.0	6.4	21.9
Martinique	1997-1999	...	...	6.5	6.5	-
Mexico	1999-2001	1,322,621	2.1	4.5	5.2	13.7
Montserrat	1992-1994	311	1.9	10.1	...	...
Nicaragua	1998-2000	42,127	3.7	2.8	5.7	49.9
Panama	1998-2000	35,701	9.3	4.2	5.1	16.9
Paraguay	1998-2000	54,202	19.4	3.4	5.4	37.0
Peru	1998-2000	262,401	15.8	3.5	6.4	46.2
Puerto Rico	1998-2000	87,193	0.7	7.5	7.9	5.1
Dominican Republic	1996-1998	76,230	10.6	3.2	5.0	36.3
Saint Kitts & Nevis	1994-1996	1,864	5.8	14.8	...	...
Saint Vincent	1997-1999	2,407	1.7	7.2	5.9	-
Saint Lucia	1993-1995	2,869	8.0	6.9	6.2	-
Suriname	1990-1992	6,171	14.1	5.1	6.2	17.8
Trinidad & Tabago	1994, 95, 98	27,942	2.1	7.4	5.9	-
Uruguay	1998-2000	94,803	7.5	9.5	9.5	-
Venezuela	1998-2000	311,536	1.4	4.4	4.4	-

...: no data available

-: magnitude 0

around 2000. The estimates are based on a comparison of the crude death rates obtained using registered mortality, as reported to PAHO for the three-year period indicated, and the death rates estimated by using abridged life table central death rates (see section on estimation of death rates by cause, age and sex), where available, or from death rates estimated by the Population Division of the United Nations.[3]

Differences among countries in the time period used for calculation of registered death rates reflect differences in the availability of data from countries at the time the table was prepared. Country-wide registered mortality data are not available from Bolivia, Honduras, Netherlands Antilles and only for recent years and with limited coverage from Haiti. The estimates shown in Table 1 provide an indication of the magnitude of the existing under registration problem in the countries. The characteristics of, and underlying reasons for, under registration of deaths vary greatly among countries and also within each country. As can be seen in the table, there is little or no under registration in Anguilla, Antigua, Argentina, Barbados, Belize, Canada, Chile, Costa Rica, Cuba, Dominica, Guadeloupe, Martinique, Saint Lucia, Saint Vincent, and Trinidad and Tobago, the United States, Uruguay, Venezuela, and the Virgin Islands (USA). In these countries, the registered rate for the period shown is identical to, and sometimes greater than, the estimated rate for the quinquennium that contains the period. Under-registration is low in Puerto Rico (5.1%) and intermediate in Brazil, Guatemala, Mexico, Panama, and Suriname, which have estimated under-registration ranging between 13% and 19%, and appear to be on the way to achieving satisfactory levels of death registration. Another 11 countries continue to have serious under-registration problems, with estimates ranging between 20% and 92%. The level of under registration is unknown in 7 countries – Bermuda, Cayman Islands, Grenada, Montserrat, St. Kitts and Nevis, Turks and Caicos Islands and the Virgin Islands (UK). No data from civil registration sources are available for Bolivia, Honduras and Netherlands Antilles in recent years. Under registration is greater for infant deaths than for deaths occurring at older ages. Infants who live just a few hours or days may not be registered as either live births or infant deaths. At advanced ages there tends to be overstatement of age, which contributes to under estimation of mortality for some adult age groups and over estimation for older groups. Clustering of deaths in certain ages due to reporting preferences (such as ages ending in 0 or 5) is another well-known phenomenon that affects the age distribution of registered deaths.

Estimation of Death Rates by Cause, Age and Sex

In view of the above limitations in the coverage of civil registration systems and in the "quality" of mortality data as indicated by the proportion of deaths assigned to the category "signs, symptoms and ill-defined conditions," a general method to more accurately estimate mortality rates that addressed these limitations was required.

Estimation of mortality rates in PAHO is based on an estimation procedure first presented in the 1992 edition of *Health Statistics from the Americas.*[4] This procedure was updated to proportionately re-assign deaths not stated by age and sex and is described in the following paragraphs as well as in the 2003 edition of that publication, which is available on-line at www.paho.org.[5]

Assumptions and methodology

The procedure uses registered mortality data available in the PAHO regional mortality database. The data is tabulated for selected year(s), causes of death, age groups, and sex. The estimates of the central death rates ($_nM_x$) for the corresponding age groups and sex are obtained from life tables for 20 Latin American countries prepared and published by the Latin American and Caribbean Demographic Center (CELADE)[3] [For English speaking countries of the Caribbean, Canada, Puerto Rico, and United States, registered rates available from the PAHO database were used]; and corresponding annual population estimates by age groups and sex. The registered mortality data is first adjusted for deaths unknown by age and sex. The number of deaths unknown by age are redistributed into known age groups by multiplying the number of deaths for each sex and age group by an adjustment factor, $f_a = D/D_a$, where D is the total number of deaths and D_a is the number of deaths stated by age. A similar adjustment factor is used to redistribute the number of deaths in each age group not stated by sex.

The rate calculations make the following assumptions about the cause distribution of registered mortality data:

(a) All registered deaths coded to an external cause were in fact due to an external cause, and none of the registered deaths coded to other cause categories, including SSI, were really due to external causes. Consequently, all deaths assigned to SSI can be proportionately redistributed among other non-external cause categories, age groups, and sex, under the assumption that the SSI deaths follow the same distribution as that observed among registered deaths from non-external "defined" causes.

(b) An estimate of the total number of deaths that actually occurred in a given year or time period is obtained by applying the corresponding quinquennial central death rates for each age and sex group from the life table to the population estimates and totaling the number of deaths in each age group by sex. By subtracting the number of registered deaths, an estimate of the number of unregistered deaths is obtained. It is further assumed that the

distribution of unregistered deaths into cause categories, by age group and sex, is the same as that among registered deaths. Accordingly, unregistered deaths, including unregistered deaths due to external causes, are redistributed into corresponding cause categories by age and sex in the same proportions as the registered deaths.

Estimated age and sex specific rates are calculated by accumulating the estimated total deaths (registered and unregistered) in a given year or time period, by cause category and dividing by the sum of the corresponding estimated populations. The infant mortality rate is calculated using the estimated number of live births, if available. Otherwise, the estimated population under 1 year of age is used in the denominator.

The estimated number of deaths for a selected age-sex group, d'_i and the country's total estimated deaths, D' annually or for a given time period are defined in Box 1, as well as the estimated number of unregistered deaths, d'_{iU} in the ith age-sex group. The proportion of unregistered deaths due to external causes for the ith age-sex group d''_{iex} and the estimated total number of deaths due to external causes in the i^{th} age-sex group d'_{iex} are also shown.

The estimated total number of deaths, d'_{ic}, for a selected cause category, *c* and age-sex group *i,* can be calculated from the above. The second expression in the equation for d'_{ic} presented in box 1 reflects the proportionate redistribution of registered SSI deaths and unregistered deaths due to non-external causes in the i^{th} age-sex group that will be re-assigned to cause category *c*. By accumulating the estimated deaths in each age-sex cause grouping, the total estimated number of deaths can be determined.

Some limitations

In some instances, the number of registered deaths for a given year or time period was greater than the estimate obtained from the CELADE life tables. This indicates that the central death rate estimates of the life table for that country and time period do not adequately reflect the observed age patterns of mortality. In those instances and in countries where life table estimates are not available, the registered mortality data, adjusted for unknown age and sex, is used in estimating the rates. In effect, this assumes that there is no under registration present in that year or time period.

Since PAHO uses CELADE as its primary source for life tables, this information is not available for the English-speaking countries of the Caribbean, Canada, Puerto Rico and United States. Other sources of life table information could be consulted including the use of national life tables and model life tables and the feasibility of their use studied. The US Census Bureau's International database (www.census.gov/ipc/www/idbacc.html) also has this data for a few Caribbean countries (Guadeloupe, Martinique, St. Kitts and Nevis, Saint Lucia, and Trinidad and Tobago) but only for a year around 1980.

Box 1: formulas for calculations

$d'_i = m_i * p_i$

m_i = Central death rate in the ith age group

p_i = corresponding population estimate

$D' = \sum d'_i$

$d'_{iU} = d'_i - d_{iR}$

d_{iR} = number of registered deaths in the ith age-sex group

$d''_{iex} = (d_{iex} / d_{iR}) * d'_{iU}$

d_{iex} = registered number of deaths due to external causes in the ith age-sex group

$d'_{iex} = d_{iex} + d''_{iex}$

$d'_{ic} = d_{ic} + [(d_{ic} / d_{iR}) - d_{issi} - d_{iex}] * [d_{issi} + (d'_{iU} - d''_{iex})]$

d_{ic} = registered number of deaths in the ith age-sex group due to cause c

d_{issi} = number of deaths in ith age-sex group assigned to «symptoms, signs and ill-defined conditions»

The estimation of rates utilizing this methodology is dependent on having suitable life tables that accurately account for a country's mortality patterns and can be used to assess the level of completeness of a country's vital registration system. It also is dependent on the accuracy in selecting and coding the underlying cause of death and on assumptions for the re-distribution of the cause category SSI and "unregistered" deaths to the cause of death structure for registered deaths. It is assumed that the registered deaths have negligible misclassification of the underlying cause of death.

References

(1) World Health Organization. *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death, Ninth Revision (1975),* Geneva, WHO, 1975. (Vol. 1 & 2).

(2) World Health Organization. *International Statistical Classification of Diseases and Related Health Problems.* Tenth Revision. Vols. 1-3. Geneva, WHO, 1992-1994.

(3) CELADE. Latin America: Life Tables 1950-2025. *Demographic Bulletin* (Santiago), 2001(Jan); 67.

(4) Pan American Health Organization. *Health Statistics from the Americas, 1992 edition*. Washington, D.C.:PAHO, 1992 (Scientific Publication 542).

(5) Pan American Health Organization. *Health Statistics from the Americas, 2003 Edition*. (Scientific Publication 591) [Web page]. Available at: http://www.paho.org/english/am/pub/SP_591.htm.

Source: Prepared by Mr. John Silvi from PAHO's Area of Health Analysis and Information Systems, and presented at the II Meeting of the Regional Advisory Committee on Health Statistics (CRAES) in September 2003.

Life Tables: A Technique to Summarize Mortality and Survival

Introduction

In a previous article of the *Epidemiological Bulletin* on years of potential life lost (YPLL)[1], emphasis was put on the importance of the age of death as a variable in mortality analysis. A concept closely linked to an individual's age at death is that of *survival*. While the YPLL consider the years of life lost as a result of *premature death*, another descriptive technique used in mortality analysis considers the *years lived* by individuals in a population before their death. This technique is that of mortality tables, more commonly known as *life tables*. It is used in public health to essentially measure mortality, but also in demographic, actuarial and other studies to examine longevity, fertility, migration, population growth and projections of population size and in studies of length of working life and length of disability-free life.[2]

In essence, life tables describe the process of extinction of a population experiencing the mortality observed at a given time, until the last of its components has died. A characteristic of life tables is that they end with the death of the last individual, and the fundamental difference between different life tables is the speed at which this end is reached.[3] Life tables can be calculated for a whole population or for a specific population subgroup (e.g., females, males, or Hispanics). In its simplest form, the entire table is generated from age specific mortality rates and the resulting values are used to measure mortality, survivorship and life expectancy, the most frequently used indicator provided by the life table. In other applications, the mortality rates are combined with demographic data to build a more complex model that measures the combined effect of mortality and changes in one or more socioeconomic characteristic.[2] One of the main advantages of life tables is that they do not reflect the effects of the age distribution of an actual population and do not require the use of a standard population for comparative analysis of levels of mortality in different populations.[2]

There are two classic forms of life tables: the *cohort* (or generation) and the *current* (or period) tables. Cohort life tables consist of monitoring a population *longitudinally* from a determining event (e.g., a birth cohort or a treatment cohort in a clinical trial) until all the individuals die or until the observation period is discontinued. Its use in the description of the survival of the whole population presents a series of practical difficulties, the most noteworthy being the large population needed to calculate a life table; the follow-up time required; and the losses due to migrations or other causes. The *cohort* table is usually used in survival analysis of clinical trials, which are carried out on smaller population samples and over a shorter period of time.

Current life tables provide a *transversal* view of mortality and survival experiences at all ages of a population during a short period of time, usually a year. They depend directly on the age-specific mortality rates for the year for which they are constructed. Thus, in a current life table the mortality experience of a population during a given year is applied to a hypothetical cohort of 10,000, 100,000 live births or in general 10^k individuals. Although the calculation is based on a "fictitious" population size, life tables reflect the "real" mortality experience of the population and are a very useful tool to compare mortality data at the international level and to assess mortality trends at the national level.[4,5]

The *complete* (or unabridged) life table is constructed using every single year of age from birth to the last applicable age. However, the *abbreviated* (abridged) life tables are more often used, in which each age is presented in groups, usually of children under 1 year, children 1 to 4 years, and 5-year age groups for the remainder of the ages until the final age interval, which remains open. The use of abbreviated tables expanded because mortality data are usually available and sufficiently accurate in the form of rates for 5-year age groups and not for each individual age. In all cases, it is assumed that deaths are distributed evenly throughout each age interval.

In addition to their general use, life tables can serve to study the impact of a cause or group of causes of death through the so-called *cause-elimination* or *multiple-decrement* life tables. They involve constructing a life table with all deaths and another one eliminating the cause or causes of interest. Upon comparing the two tables, the impact of the eliminated deaths can be observed in the different indicators of the life table.[4] The years of life expectancy lost (YLEL) are based on a similar concept and will be presented in a future issue of the *Epidemiological Bulletin*.

Limitations of life tables

Life table estimates have all the disadvantages of any statistical measure based on population censuses and vital records. Data on ages and mortality registries may be incomplete or biased. Infant mortality weighs heavily on life expectancy, which means that under-reporting of this indicator, a habitual fact in many countries, can have an important effect on the results of the tables. The same can be said about the procedure used in closing the final, open interval of the mortality table (e.g, 85 and more, 90 and more) and the information inaccuracies existing in these age intervals. Also, important differences in specific age/sex groups with high mortality may be overlooked, since this would have little effect on the overall life expectancy.

Constructing life tables for small populations, at the local or subregional level, is generally not recommended, since migratory movements affect the population structure more than

at the regional or national levels. In these cases, a very small number of deaths can be obtained, which may produce imprecise calculations of the table's columns.

Construction and interpretation of a life table

The construction of a life table is a simple process. It involves following a few routine steps that are repeated for each age group, which can be enormously facilitated by the use of a spreadsheet such as the one proposed by the United States Bureau of Census[6] or any other software offering this tool, such as Epidat 3.0.[7] The different components usually included in a life table are presented below, as well as their interpretation.[3,4] The formulas to calculate them are presented in box 1.

EXACT AGE (x). This column presents the lower limit of each age interval (usually 5-year periods), beginning with 0 and incrementing to 1, 5, 10, 15 and so on until the last, open interval is reached. As mentioned before, the first and second age groups are usually "under 1" and "1-4", therefore the values of the second and third rows of this column are 0 and 1. This reflects the importance and specific interest in mortality among children under 1, known as *infant mortality rate*[a]. Further, it is preferable to separate the calculation for age 0, and occasionally for age 1, from the age groups 1-4 or 2-4, due to the lack of homogeneity of mortality in this interval. Since the first stratum is a one-year age group, the following stratum from 1 to 4 is a 4-year age group. When adequate statistics are available, it is better to calculate directly the probabilities of death in the first and second years of life, using infant birth and death statistics.[3]

For a final, open interval, the most commonly used is 85 years and over, although it can vary depending on the life expectancy of the country.

WIDTH (IN YEARS) OF THE AGE INTERVAL (n). Usually, the first value is 1 (interval 0, 1), the second 4 (interval 1, 5) and the remaining values are 5 (5-year intervals), with the exception of the last value that normally is represented with the sign + indicating an open interval.

NUMBER OF DEATHS RECORDED IN THE INTERVAL (d_x). This column presents the number of subjects dying in that age group during the year corresponding to the life table.

NUMBER OF SUBJECTS IN THAT AGE GROUP (P_x). These numbers indicate the size of the corresponding age groups in the population under study, during the year considered.

AVERAGE NUMBER OF YEARS LIVED BY THOSE WHO DIE BETWEEN AGES X AND X+N, CALLED "SEPARATION FACTOR" ($_na_x$). Although it is necessary in its calculation, this factor is not typically presented as a column of the life table. Each person living in the interval (x, x+n) has lived x complete years plus some fraction of the interval (x, x+n). In a complete life table, a value of 0.5 (i.e. half of one year) is valid from the age of 5. For a simpler calculation, it is also assumed that those who die in the 5-year age intervals of an abridged life table live on average 2.5 years.[2] However, this is not necessarily the best value for the separation factor, because the value of this fraction depends on the mortality pattern over the entire interval and not the mortality rate for any single year. In addition, since a large proportion of infant deaths occur in the first weeks of life, this value is much smaller in the 0-1 age group and in the age group 1-4. Calculation of the separation factor is easy if the date of birth and date of death are available. When they are not, values from model life tables, such as those tabulated by Coale and Demeny, shown in Table 1, can be utilized for $_1a_0$ and $_4a_1$.

Box 1: Formulas to calculate the life table*

$$_nM_x = d_x / P_x$$

$$_nq_x = [n * {}_nM_x] / [1 + (n - {}_na_x) * {}_nM_x]$$

$$_np_x = 1 - {}_nq_x$$

$$_nl_{x+n} = {}_nl_x * {}_np_x$$

The following formula can also be used: $_nl_{x+n} = {}_nl_x - {}_nd_x$

$$_nd_x = {}_nl_x * {}_nq_x$$

$$_nL_x = n * {}_nl_{x+n} + {}_na_x * {}_nd_x$$

($L_w = d_w / M_w$, w representing the most advanced age)

$$_nT_x = {}_nT_{x+n} + {}_nL_x$$

($T_w = L_w$, w representing the most advanced age)

$$_ne_x = {}_nT_x / {}_nl_x$$

* *Notation: the right subscript (x) refers to the initial point of the interval. The left subscript (n) refers to the interval width.*

Table 1: Separation factors for ages 0 and 1-4

	Zones	Separation factor for age 0			Separation factor for age 1-4		
		Men	Women	Both sexes	Men	Women	Both sexes
Infant mortality rate > 0,100	North[1]	0.33	0.35	0.3500	1.558	1.570	1.5700
	East[2]	0.29	0.31	0.3100	1.313	1.324	1.3240
	South[3]	0.33	0.35	0.3500	1.240	1.239	1.2390
	West[4]	0.33	0.35	0.3500	1.352	1.361	1.3610
Infant mortality rate < 0,100	North	0.0425	0.05	0.0500	1.859	1.733	1.7330
	East	0.0025	0.01	0.0100	1.614	1.487	1.4870
	South	0.0425	0.05	0.0500	1.541	1.402	1.4020
	West	0.0425	0.05	0.0500	1.653	1.524	1.5240

[1] Iceland, Norway and Switzerland; [2] Austria, Czechoslovakia, North-central Italy, Poland and Hungary; [3] South Italy, Portugal and Spain; [4] Rest of the World.

[a] Technical note: In a strict sense, the infant mortality rate is not equal to the under-one mortality rate, because they have different denominators. The first one is live births, and the second children under one year of age, which is more difficult to determine.

CENTRAL MORTALITY RATE (MORTALITY RATE) ($_nM_x$). This column results from dividing the deaths in the x, x+n interval (column d_x) by the number of people in this age group (column P_x).

PROBABILITY OF DYING BETWEEN THE AGES X AND X+N ($_nq_x$). The probabilities of dying are calculated based on the age-specific mortality rates for each age group. This column should be interpreted as the probability of dying between the two ages for the subject that has survived up to age x. For the last age group of the table, where death is unavoidable, the probability of dying is 1. For the other age groups, the calculation is more complicated (see Box 1).

PROBABILITY OF SURVIVAL BETWEEN THE AGES X AND X+N ($_np_x$). This column is the complement to 1 of $_nq_x$, and therefore it is sometimes not included in the life table. It should be interpreted as the probability of an individual who reaches age x to reach the exact age x+n alive.

SURVIVORS TO EXACT AGE X ($_nl_x$). l_0 is the initial number of newborns composing the generation, who are destined to die through

Box 2: Example of calculation of a life table: Brazil, 2000

Data from the death registry and population census:

Age group	Deaths[1]	Population[2]
Under 1	65,532	3,205,108*
1-4	11,271	13,084,650
5-9	5,366	16,533,114
10-14	6,294	17,406,984
15-19	19,255	17,847,032
20-24	26,620	16,500,057
25-29	25,404	14,534,868
30-34	28,162	13,533,472
35-39	33,578	12,953,294
40-44	39,855	10,942,252
45-49	45,880	9,106,099
50-54	52,276	7,139,958
55-59	58,078	5,425,966
60-64	72,044	4,553,017
65-69	81,641	3,365,780
70-74	93,339	2,588,020
75-79	90,927	1,602,984
80-84	80,847	857,170
85+	103,085	460,928

Questions related to the interpretation of the values in the life table:

1- What is the probability for an individual under 1 to die in Brazil in 2000?

The probability of dying between 0 and 1 is Brazil in 2000 ($_1q_0$) is 0.02006.

2- How many years can an individual born in 2000 in Brazil expect to live?

The number of years that a child born in 2000 may hope to live, i.e. the life expectancy at birth in Brazil (e_0) is 71.97 years.

3- What is the probability of dying of an individual between 5 and 10 years of age?

The probability that an individual die in 2000 in the 5-9 age group ($_5q_5$) is 0.00162.

4- What is the mortality rate between 5 and 10 years of age?

The central mortality rate in the 5-9 age group ($_5M_5$) is 0.00032.

5- What is the probability that an individual reaching 5 years of age reaches 10?

The probability that an individual in the 5-9 age group reaches the 10-14 age group ($_5p_5$) is 0.99838.

6- How many additional year is an individual between 5 and 10 years of age in 2000 in Brazil expected to live?

The life expectancy of the 5-9 age group is e_5 = 68.68.

NOTE: Because of differences in data sources or small variations in the methods used, the values obtained in this example may differ slightly from others published elsewhere. It should be noted in particular that the data presented here are not adjusted for deaths of unknown age, which represent 0.74% of all registered deaths. The values presented here were calculated using the formulas mentioned in this article in an Excel spreadsheet.

Life table:

x	n	d_x	P_x	$_na_x$**	$_nM_x$	$_nq_x$	$_np_x$	$_nl_x$	$_nd_x$	$_nL_x$	$_nT_x$	$_ne_x$
0	1	65,532	3,205,108*	0.05	0.02045	0.02006	0.97994	100,000	2,006	98,095	7,196,592	71.97
1	4	11,271	13,084,650	1.524	0.00086	0.00344	0.99656	97,994	337	391,143	7,098,498	72.44
5	5	5,366	16,533,114	2.5	0.00032	0.00162	0.99838	97,657	158	487,891	6,707,355	68.68
10	5	6,294	17,406,984	2.5	0.00036	0.00181	0.99819	97,499	176	487,055	6,219,463	63.79
15	5	19,255	17,847,032	2.5	0.00108	0.00538	0.99462	97,323	524	485,306	5,732,408	58.90
20	5	26,620	16,500,057	2.5	0.00161	0.00803	0.99197	96,799	778	482,053	5,247,103	54.21
25	5	25,404	14,534,868	2.5	0.00175	0.00870	0.99130	96,022	835	478,020	4,765,050	49.62
30	5	28,162	13,533,472	2.5	0.00208	0.01035	0.98965	95,186	985	473,468	4,287,030	45.04
35	5	33,578	12,953,294	2.5	0.00259	0.01288	0.98712	94,201	1,213	467,972	3,813,563	40.48
40	5	39,855	10,942,252	2.5	0.00364	0.01805	0.98195	92,988	1,678	460,744	3,345,591	35.98
45	5	45,880	9,106,099	2.5	0.00504	0.02488	0.97512	91,310	2,272	450,869	2,884,847	31.59
50	5	52,276	7,139,958	2.5	0.00732	0.03595	0.96405	89,038	3,201	437,188	2,433,978	27.34
55	5	58,078	5,425,966	2.5	0.01070	0.05212	0.94788	85,837	4,474	418,000	1,996,790	23.26
60	5	72,044	4,553,017	2.5	0.01582	0.07611	0.92389	81,363	6,192	391,334	1,578,790	19.40
65	5	81,641	3,365,780	2.5	0.02426	0.11435	0.88565	75,171	8,596	354,365	1,187,456	15.80
70	5	93,339	2,588,020	2.5	0.03607	0.16541	0.83459	66,575	11,012	305,345	833,091	12.51
75	5	90,927	1,602,984	2.5	0.05672	0.24839	0.75161	55,563	13,801	243,310	527,746	9.50
80	5	80,847	857,170	2.5	0.09432	0.38161	0.61839	41,761	15,937	168,965	284,436	6.81
85	+	103,085	460,928		0.22365	1.00000	0.00000	25,825	25,825	115,471	115,471	4.47

*Number of live births

** These values of the separation factor were selected because the infant mortality rate in Brazil is less than 0.1 (i.e. less than 100 deaths per 1,000 live births) and in the Coale y Demeny classification of countries, Brazil is part of the "West" group" (see table 1)

[1] PAHO. Technical Information System: Regional Mortality Database. AIS; Washington, D.C.; 2003.

[2] United Nations Population Division. World Population Prospects: The 2002 Revision. New York; 2003.

the process of mortality followed by the life table. It is called the radix of the table and has a value of 100,000 (or 10^k).

DEATHS BETWEEN THE EXACT AGES X AND X+N ($_nd_x$). In order to obtain $_nd_x$, l_x is multiplied by $_nq_x$.

NUMBER OF YEARS LIVED BY THE TOTAL OF THE COHORT OF 100,000 BIRTHS IN THE INTERVAL X, X+N ($_nL_x$). Each member of the cohort that survives the interval x, x+n contributes n years to L, while each member who dies in the interval x and x+n contributes the average number of years lived by those which die in this period, i.e. the separation factor of deaths mentioned previously. For the last, open group, L_w is used.

TOTAL YEARS LIVED AFTER EXACT AGE X (T_x). This number is essential for the calculation of life expectancy. It indicates the total number of years lived by the survivors l_x between the anniversary x and the extinction of the whole generation. The value T_0 is the total number of years lived by the cohort until the death of its last component.

LIFE EXPECTANCY AT AGE X ($_ne_x$). Among all the indicators provided by the life table, the most widely used is life expectancy (e_x), which represents the average number of years remaining to be lived by survivors to age x. As a result, life expectancy at birth (e_0) is the average number of years lived by a generation of newborns under given mortality conditions. This synthetic indicator is one of the most widely used to compare the general level of mortality between countries and over time.[2]

Life expectancy always decreases from the first row of the table to the last, with the exception of the second row (1-4), which can be greater than the first (0-1) in countries with very high infant mortality.[4] For a given population, life expectancy is greater in women that in men and the overall life expectancy should be approximately between the two. Exceptions to this rule could arise in countries with high fertility and high maternal mortality, or in populations in which, for cultural reasons, the nutritional and general living conditions of women are markedly worse than those of men.

Applications

The life table is a widely-used statistical table in demographic, social and health studies. The principal objective of a life table is to calculate life expectancy, at birth and at other ages. However life tables provide other interesting demographic data. Since the life table measures the probability of death (or some other end point) at each designated time interval, it thus provides the survival curve for a cohort of individuals. It is common to use the life table method to compare survival curves for two patient cohorts receiving different therapies in evaluating the differences or effectiveness of these therapies. It also allows calculating the *survival ratio*. This ratio, usually presented for a 5-year period ($_5P_x = {}_5L_{x+5} / {}_5L_x$), represents the survival between 2 age groups, i.e. the average chance that a person in an age group will survive 5 more years to the next age group. It is used in particular for making population projections.

Example

Box 2 presents data on deaths and population in Brazil in 2000. These data allow calculating the life table. The calculation starts with $_nM_x$.

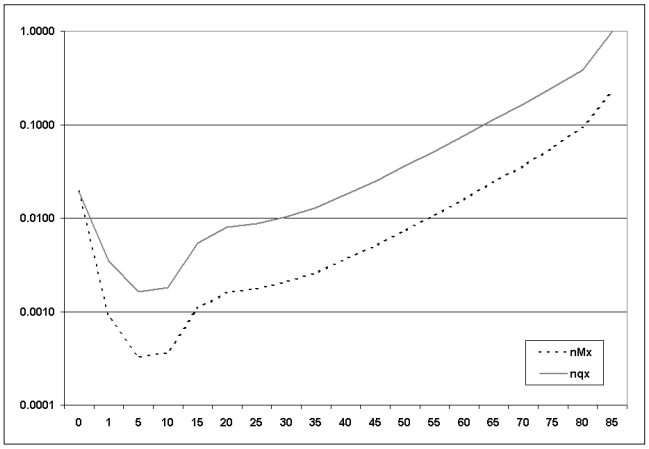
Figure 1 shows $_nq_x$ and $_nM_x$. They are presented on a logarithmic scale because the magnitude of the range of these two indicators is such that it cannot be visualized on a single graph with an arithmetic scale. The two curves are parallel, except in the extreme ages where they coincide or start to join. In effect, the probability of dying consistently overestimates the mortality rate, except in the group of children less than 1 year of age, where $_nM_x$ is above $_nq_x$. The two curves have the characteristic "J" shape, decreasing until the 5-9 interval, where they start to increase slightly until the 10-14 age group, then more rapidly until the 15 to 20 age group, and then regularly until they start joining at the 85-89 group.

Conclusion

Life tables present the mortality and survival experience of a whole population and permit evaluation of its effect on specific groups and over different periods. It is a simple instrument that is easily constructed with data collected routinely.

It is important to keep in mind that life tables are constructed based on population data from censuses and mortality registries, and therefore that the quality limitations of the latter will also affect , to different degrees, the validity of the estimations from the life table.

Figure 1: $_nM_x$ and $_nq_x$, Brazil, 2000 (logarithmic scale)



References:

(1) Pan American Health Organization. Area of Health Analysis and Information Systems. Techiques for the Measurement of the Impact of Mortality: Years of Potential Life Lost. *Epidemiological Bulletin*. 24(2):1-4; 2003

(2) United States Bureau of the Census. Shryock H, Siegel JS et al. *The Methods and materials of Demography*, Second Printing

(rev.). Washington, DC: United States Government Printing Office; 1973.

(3) Livi-Bacci M. *Introducción a la demografía*. Barcelona: Ed Ariel; 1993.

(4) Chang CJ. *Life Tables and Mortality Analysis*. Geneva: World Health Organization; 1980.

(5) Grundy EMD. Populations and population dynamics. In: Detels R, Holland WW, McEwen JMc and Omenn GS Eds. *Oxford textbook of Public Health, vol 1. The Scope of Public Health*. London: Oxford University Press; 1997.

(6) United States Census Bureau. Population Analysis Spreadsheets (PAS) [Internet page]. Available at: http://www.census.gov/ipc/www/pas.html. Accessed on 5 December 2003.

(7) Xunta de Galicia, Consellería de Sanidade e Servicios Sociais. Organización Panamericana de la Salud. Area de Análisis de Situación y Sistemas de Información. Análisis Epidemiológico de Datos Tabulados (Epidat), version 3.0 [Computer Program for Windows]; [To be published]

(8) Coale, Ansley J, Demeny P. *Regional Model Life Tables and Stable Populations*, Princeton University Press, 1966.

Source: Prepared by Dr. Enrique Vázquez from PAHO's Area of Health Situation Analysis and Information Systems (AIS) in the PAHO/WHO Argentina, Dr. Francisco Camaño (Universidad de Santiago de Compostela, Spain), Mr. John Silvi and Ms. Anne Roca (AIS - Washington, D. C.).

A Glossary for Multilevel Analysis

Ana V. Diez Roux
Divisions of Medicine and Epidemiology, Columbia University
New York, New York, United States

PART II

EMPIRICAL BAYES ESTIMATES

Estimates of parameters for a given group or higher level unit (for example, estimates of group specific intercepts or slopes, such as b_{0j} and b_{1j} in equation (1), under MULTILEVEL MODELS) obtained by combining information from the group itself with information from other similar groups investigated.[10, 19, 20] This is particularly useful when estimating parameters for a group with few within group observations. These estimates are "optimally" weighted averages that combine information derived from the group itself with the mean for all similar groups. The weighted average shifts the group specific estimate (derived using data only for that particular group) towards the mean for similar groups. The less precise the group specific estimate and the less the variability observed across groups, the greater the shift towards the overall group mean. Thus, the estimate for a given group is based not only on its own data but also takes into account estimates for other groups and the characteristics groups share.[20] Empirical Bayes estimates of parameters for a given group can be derived from multilevel models using estimates of the group level errors (for example, U_{0j} and U_{1j}, see MULTILEVEL MODELS) for that particular group. Empirical Bayes estimates are also sometimes referred to as "shrinkage estimates" because they "shrink" the group specific estimate towards the overall mean (although in fact when the overall mean is greater than the group specific estimate, the "shrunken" or empirical Bayes estimate may actually be greater than the group specific estimate). In public health, empirical Bayes estimation can be used, for example, to derive improved estimates of rates of death or diseases for small areas with few observations,[21] or to estimate rates of different health outcomes for individual providers (hospitals, physicians, etc.)[22] In other applications (which do not involve the structure of individuals within groups although they are analogous to it), empirical Bayes estimates of regression coefficients have been used to obtain improved estimates of associations in studies investigating the role of multiple exposures.[23]

ENVIRONMENTAL VARIABLES

In the context of ecological studies and multilevel analysis, the term "environmental variables" has sometimes been used to refer to group level measures of physical or chemical exposures. Environmental variables, so defined, have been proposed as a "type" of GROUP LEVEL VARIABLE, distinct from DERIVED VARIABLES and INTEGRAL VARIABLES.[11] These variables are not derived by aggregating the characteristics of individuals but they do have group level and individual level analogues (for example, days of sunlight in the community and individual level sunlight exposure information). In contrast with derived and integral variables, which may be used as indicators of group level constructs, group level environmental variables are used exclusively as proxies for individual level exposures (which may be more difficult to measure for logistic or methodological reasons), rather than as indicators of a group level property, which is conceptually different from the analogous measure at the individual level.

FIXED EFFECTS/FIXED COEFFICIENTS

Regression coefficients (intercepts or covariate effects) that are not allowed to vary randomly across higher level units (see MULTILEVEL MODELS). For example, in the case of persons nested within neighborhoods, two options are available for modelling the effects of neighborhood. One option is to include a dummy variable for each neighborhood. In this case the neighborhood coefficients are modelled as fixed (sometimes called "fixed effects"). Another option is to assume that the neighborhoods in the sample are a random sample of a larger population of neighborhoods and that the coefficients for the "neighborhood effect" vary randomly

around an overall mean (for example, as reflected by U_{0j} in equation 2 under the entry for MULTILEVEL MODELS). In this case, the neighborhood effects are modelled as random (sometimes called "random effects", see RANDOM EFFECTS MODELS). In the same example, the coefficients for individual level covariates can also be modelled as fixed or random. For example, if the relation between individual level income and blood pressure is not allowed to vary randomly across neighborhoods, the coefficient for individual level income is fixed ("fixed coefficient"). On the other hand, if the coefficient for individual level income is allowed to vary randomly across neighborhoods around an overall mean effect (as reflected by U_{1j} in equation 3 under the entry for MULTILEVEL MODELS), the coefficient for income is modelled as random (sometimes called a "random coefficient", see RANDOM COEFFICIENT MODELS). Although the terms "fixed effects" and "fixed coefficients " are sometimes distinguished as noted above, they are often used interchangeably. Fixed effects models or fixed coefficient models are models in which all effects or coefficients are fixed. See also RANDOM EFFECTS/RANDOM COEFFICIENTS.

GROUP LEVEL VARIABLES

Term used to refer to variables that characterize groups. The terms group level variables, macro variables and ecological variables are often used interchangeably.[2, 6, 11, 14, 24] Group level variables may be used as proxies for unavailable or unreliable individual level data (for example, when neighborhood mean income is used as a proxy for the individual level income of individuals living in the neighborhood) or as indicators of group level constructs (for example, when mean neighborhood income is used as an indicator of neighborhood characteristics that may be related to individual level outcomes independently of individual level income). It is the second usage (as indicators of group level constructs) that is of particular interest in multilevel analysis. Group level variables have been classified into two basic types,[11, 13, 24] DERIVED VARIABLES and INTEGRAL VARIABLES. Two additional types of group level variables, STRUCTURAL VARIABLES [13] and ENVIRONMENTAL VARIABLES [11] are sometimes distinguished. The term contextual variables has been used as a synonym for group level variables generally [6, 13] although it is sometimes reserved for derived group level variables.[11, 14]

HIERARCHICAL (LINEAR) MODELS

See MULTILEVEL MODELS

INDIVIDUAL LEVEL VARIABLES

Term used to refer to variables that characterize individuals and refer to individual level constructs (for example, age or personal income).

INDIVIDUALISTIC FALLACY

Term used as a synonym for the ATOMISTIC FALLACY. May sometimes also be used as a synonym for the PSYCHOLOGISTIC FALLACY.

INTEGRAL VARIABLES

A type of GROUP LEVEL VARIABLE. Integral variables differ from DERIVED VARIABLES (another type of group level variable) in that they are not summaries of the characteristics of individuals in the group. Integral variables have no individual level analogues and necessarily refer to group level constructs. Examples of integral variables include the existence of certain types of laws, political or economic system, social disorganization, or population density.[11, 13] Integral variables have also been referred to as primary or global variables.

INTRACLASS CORRELATION

A measure of the degree of resemblance between lower level units belonging to the same higher level unit or cluster.[25] In the case of individuals nested within groups (for example, neighborhoods), the intraclass correlation measures the extent to which values of the dependent variable are similar for individuals belonging to the same group. It can be thought of as the average correlation between values of two randomly drawn lower level units (for example, individuals) in the same, randomly drawn higher level unit (for example, neighborhood). It can also be defined as the proportion of the variance in the outcome that is between the groups or higher level units. In the case of a simple random intercept model, the intraclass correlation coefficient is estimated by the ratio of population variance between groups (τ_{00}) to the total variance ($\tau_{00} + \sigma^2$).[25] (see MULTILEVEL MODELS) The estimation of the intraclass correlation coefficient in models including random covariate effects, or in the case of non-normally distributed dependent variables, is more complex and not always straightforward .

MARGINAL MODELS

See POPULATION-AVERAGE MODELS.

MIXED MODELS

Term used to refer to models that contain a mixture of FIXED EFFECTS (or fixed coefficients) and RANDOM EFFECTS (or random coefficients). In mixed models some of the regression coefficients (intercepts or covariate effects) are allowed to vary randomly across higher level units but others are not (see MULTILEVEL MODELS). Thus mixed models can be thought of as a particular case of the more general multilevel models (although the term is also occasionally used as a synonym of multilevel models generally). Sometimes the term mixed models is also used to encompass models that account for correlation between lower level units (for example, individuals) within higher level units (for example, neighborhoods) in

other ways—that is, by modelling the correlations or covariances themselves rather than by allowing for random effects or random coefficients.[26] These models (which are not multilevel models) have also been called covariance pattern models,[26] marginal models, or POPULATION AVERAGE MODELS.

MULTILEVEL ANALYSIS

An analytical approach that is appropriate for data with nested sources of variability—that is, involving units at a lower level or micro units (for example, individuals) nested within units at a higher level or macro units (for example, groups such as schools or neighborhoods).[5, 10, 19, 24, 25, 27–30] Multilevel analysis allows the simultaneous examination of the effects of group level and individual level variables on individual level outcomes while accounting for the non-independence of observations within groups. Multilevel analysis also allows the examination of both between group and within group variability as well as how group level and individual level variables are related to variability at both levels. Thus, multilevel models can be used to draw inferences regarding the causes of inter-individual variation (or the relation of group and individual level variables to individual level outcomes) but inferences can also be made regarding inter-group variation, whether it exists in the data, and to what extent it is accounted for by group and individual level characteristics. In multilevel analysis, groups or contexts are not treated as unrelated but are conceived as coming from a larger population of groups about which inferences want to be made. Multilevel analysis thus allows researchers to deal with the micro-level of individuals and the macro-level of groups or contexts simultaneously.[5]

Multilevel analysis has a broad range of applications in many situations involving nested sources of random variability such as persons nested within neighborhoods,[5, 30] patients nested within providers,[31] meta analysis (observations nested within sites),[19, 32] longitudinal data analysis (repeat measurements over time nested within persons),[28, 33, 34] multivariate responses (multiple outcomes nested within individuals),[5] the analysis of repeat cross sectional surveys (multiple observations nested within time periods),[35] the examination of geographical variations in rates (rates for smaller areas nested within regions or larger areas)[36] and the examination of interviewer effects (respondents nested within interviewers).[37] Multilevel analysis can also be used in situations involving multiple nested contexts[19, 28] (for example, multiple measures over time on individuals nested within neighborhoods) as well as overlapping or cross classified contexts (for example, children nested within neighborhoods and schools).[38] The statistical models used in multilevel analysis are referred to as MULTILEVEL MODELS [25, 28, 29] or hierarchical linear models.[19, 39]

MULTILEVEL MODELS

The statistical models used in MULTILEVEL ANALYSIS.[19, 25, 28, 29] The terms "hierarchical models" and "multilevel models" are often used synonymously. These models (or variants of them) have previously appeared in different literatures under a variety of names including RANDOM EFFECTS MODELS or RANDOM COEFFICIENT MODELS [40–42] "covariance components models" or "variance components models",[43, 44] and MIXED MODELS.[26] A simplified example for the case of a normally distributed dependent variable, a single individual level (lower level unit) predictor and a single group level (higher level unit) predictor is provided below. Analogous models can be formulated for non-normally distributed dependent variables.[10, 28, 39, 45]

In the case of multilevel analysis involving two levels (for example, individuals nested within groups), the multilevel model can be conceptualized as a two stage system of equations.

In the first stage (level 1), a separate individual level regression is defined for each group or higher level unit.

(1) $Y_{ij} = b_{0j} + b_{1j}I_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$

Y_{ij} = outcome variable for i^{th} individual in jth group

I_{ij} = individual level variable for ith individual in jth group

b_{0j} is the group specific intercept

b_{1j} is the group specific effect of the individual level variable

Individual level errors (e_{ij}) are assumed to be independent and identically distributed with a mean of 0 and a variance of σ^2. The same regressors are generally used in all groups, but regression coefficients (b_{0j} and b_{1j}) allowed to vary from one group to another.

In a second stage (level 2), each of the group or context specific regression coefficients defined in equation (1) (b_{0j} and b_{1j} in this example) are modelled as a function of group level (or higher level) variables.

(2) $b_{0j} = \gamma_{00} + \gamma_{01}G_j + U_{0j} \qquad U_{0j} \sim N(0, \tau_{00})$

(3) $b_{1j} = \gamma_{10} + \gamma_{11}G_j + U_{1j} \qquad U_{1j} \sim N(0, \tau_{11})$

$\text{cov}(U_{0j}, U_{1j}) = \tau_{10}$

G_j group level variable

γ_{00} is the common intercept across groups

γ_{01} is the effect of the group level predictor on the group specific intercepts

γ_{10} is the common slope associated with the individual level variable across groups

γ_{11} is the effect of the group level predictor on the group specific slopes

The errors in the level 2 equations (U_{0j} and U_{1j}), sometimes called "macro errors", are assumed to be normally dis-

tributed with mean 0 and variances τ_{00} and τ_{11} respectively. τ_{01} represents the covariance between intercepts and slopes. Thus, multilevel analysis summarizes the distribution of the group specific coefficients in terms of two parts: a "fixed" part that is common across groups (γ_{00} and γ_{01} for the intercept, and γ_{10} and γ_{11} for the slope) and a "random" part (U_{0j} for the intercept and U_{1j} for the slope) that is allowed to vary from group to group (see also FIXED COEFFICIENTS and RANDOM COEFFICIENTS).

By including an error term in the group level equations (equations (2) and (3)), these models allow for sampling variability in the group specific coefficients (b_{0j} and b_{1j}) and also for the fact that the group level equations are not deterministic (that is, the possibility that not all relevant macro-level variables have been included in the model). The underlying assumption is that group specific intercepts and slopes are random samples from a normally distributed population of group specific intercepts and slopes, or alternatively, that the macro errors are exchangeable—that is, that the residual variation in group specific coefficients across groups is unsystematic.[10]

An alternative way to present the model fitted in multilevel analysis is to substitute equations (2) and (3) in (1) to obtain:

$$Y_{ij} = \gamma_{00} + \gamma_{01}C_j + \gamma_{10}I_{ij} + \gamma_{11}C_jI_{ij} + U_{0j} + U_{1j}I_{ij} + \varepsilon_{ij}$$

The model includes the effects of group level variables (γ_{01}), individual level variables (γ_{10}) and their interaction (γ_{11}) on the individual level outcome Y_{ij}. These coefficients (γ_{01}, γ_{10} and γ_{11}), which are common to all individuals regardless of the group to which they belong, are often called the FIXED COEFFICIENTS (or fixed effects). The model also includes a random intercept component (U_{0j}), and a random slope component (U_{1j}). The values of these components vary randomly across groups, and hence U_{0j} and U_{1j} referred to as the RANDOM COEFFICIENTS (or random effects). The parameters of the above equations (fixed effects, random effects, variances of the random effects, and residual variance) are simultaneously estimated using iterative methods. The level 1 and level 2 variances (σ^2, τ_{00}, τ_{11}, τ_{10}) are called the (CO)VARIANCE COMPONENTS.

Many variants of the more general model illustrated above are possible. For example, only group specific intercepts (b_{0j}) may be modelled as random (these models have also been called RANDOM EFFECTS MODELS). When covariate effects (b_{1j} in the example above) are modelled as random these models have also been called RANDOM COEFFICIENT MODELS. When some of the coefficients are fixed and others are random, these models have also been called "mixed effects models" or simply MIXED MODELS. When all coefficients are modelled as fixed (no random errors are included in level 2 equations), these models are reduced to traditional CONTEXTUAL EFFECTS MODELS. Multilevel models can also account for multiple nested contexts (or levels)[19, 28] allowing fixed and random coefficients to be associated with variables measured at different levels of the data hierarchy being analyzed. Multilevel models can also be modified to allow for non-hierarchical, overlapping or cross classified contexts (for example, children simultaneously nested within neighborhoods and schools).[38]

References:

NOTE: References 1-18 were included in Part I of the Glossary, in Vol. 24, No. 3 (2003) of the *Epidemiological Bulletin.*

(19) Bryk AS, Raudenbush SW. *Hierarchichal linear models: applications and data analysis methods*.Newbury Park: Sage, 1992.

(20) Rice N, Jones A. *Multilevel models and health economics*. Health Econ 1997;6:561–75.

(21) Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987;43:671–81.

(22) Thomas N, Lonford N, Rolph J. Empirical Bayes methods for estimating hosptial-specific morality rates. *Stat Med*1994;13:889–903.

(23) Witte JS, Greenland S, Haile RW, et al. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology* 1994;5:612–21.

(24) Von Korff M, Koepsell T, Curry S, et al. Multi-level research in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol* 1992;135:1077–82.

(25) Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London: Sage, 1999.

(26) Brown H, Prescott R . *Applied mixed models in medicine*. New York: Wiley, 2000.

(27) Mason W, Wong G, Entwisle B. Contextual analysis through the multilevel linear model. In: Leinhardt S, ed. *Sociological methodology*. San Francisco: Josey Bass, 1983–1984: 72–103.

(28) Goldstein H. *Multilevel statistical models*. New York: Halsted Press, 1995.

(29) Kreft I, deLeeuw J. *Introducing multilevel modeling*. London: Sage, 1998.

(30) Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;21:171–92.

(31) Sixma HJ, Spreeuwenberg PM, Pasch MAvd. Patient satisfaction with the general practitioner: a two-level analysis. *Med Care* 1998;36:212–29.

(32) Hedeker D, Gibbons R, Davis J. Random regression models for multicenter clinical trials data. *Psychopharmacol Bull*1991;27:73–7.

(33) Rutter C, Elashoff R. Analysis of longitudinal data: random coefficient regression modelling. *Stat Med*1994;13:1211–31.

(34) Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med* 1997;16:2349–80.

(35) DiPrete T, Grusky D. The multi-level analysis of trends with repeated cross-sectional data. *Sociol Methodol* 1990;20:337–68.

(36) Langford I, Bentham G, McDonald A. Multi-level modelling of geographically aggregated health data: a case study of malignant melanoma mortality and uv exposure in the European community. *Stat Med*1998;17:41–57.

(37) Hox JP, de Leeuw ED, Kreft IGG. The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. In: Biemer PP, Lyberg LE, Mathiowetz NA, et al, eds. *Measurement errors in surveys*. New York: Wiley, 1991.

(38) Goldstein H. Multilevel cross-classified models. *Sociol Methods Res* 1994;22:364–75.

Source: Published initially as "A glossary for multilevel analysis" in the *Journal of Epidemiology and Community Health*, 56:588-594, 2002.

Revision of the International Health Regulations

Background

The International Health Regulations (IHR) represents the earliest multilateral initiative by countries to develop an effective framework to prevent cross-border transmission of diseases. The IHR strives to harmonize public health, trade, and traffic, and today remains the only binding set of regulations on global surveillance for infectious diseases by the World Health Organization's (WHO) Member States.

The current IHR was adopted in 1969, amended in 1973 with additional provisions for cholera, and subsequently revised in 1981 to exclude smallpox. Today, only cholera, plague and yellow fever are notifiable diseases under the IHR. Its fundamental purpose is to ensure *maximum security against the international spread of diseases with a minimum interference with world traffic.*

Because of extensive globalization in travel and trade, diseases from even remote parts of the world could spread to other areas. Potentially damaging traffic and trade embargoes may be imposed, sometimes based only on the perception of risk for disease importation, and potentially reach global proportions as happened during the cholera epidemic in the Americas in the early 1990's.

To address the threat posed by substantial increases in international travel and the potential for the rapid spread of infectious diseases, especially by air travel, the World Health Assembly (WHA) requested the revision of the International Health Regulations (IHR) in the 1995 resolution WHA 48.7.

Progress

The current revision is a collaborative process that was initiated in 1995. Its essence is to review the gaps in the present IHR and transform it into an effective regulatory tool for WHO Member States to strengthen global disease surveillance and to be proactive in dealing with international outbreaks. Proposed changes are being developed and fine-tuned to adapt to contemporary global surveillance demands and control of international outbreaks. All of the items introduced are proposals, and as such require extensive consultation before presentation to the WHA and ultimate acceptance by Member States.

The revision approach is based on three specific principles[1]:

– Ensuring that all public health risks (mainly of infectious origin) that are of urgent international importance are reported under the Regulations
– Avoiding stigmatization and unnecessary negative impact on international travel and trade and invalid reporting from sources other than Member States, which can have serious economic consequences for countries
– Ensuring that the system is sensitive enough to detect new or re-emerging public health events.

To this end, three key changes are being proposed. First, the scope of reported events will be expanded to include all *public health emergencies of international concern.* There will be a clear link between reporting and established mechanisms for action**.**

To define an event as a public health emergency of international concern a set of specific criteria is being proposed:

(1) *Severity*: The health event produces an abnormal increase of case fatality and/or incidence rates
(2) *Unusual or unexpected*: An emerging health event or a known health event showing an abnormal behavior
(3) Risk of *international propagation*
(4) The event will lead, eventually, to *international restrictions of travel and trade*

Second, a *National Focal Point* will be designated to facilitate the greater flow of information between the WHO and the different national levels in both directions. Specifically, this focal point should be able to: manage international surveillance and response requirements; advise senior health officials regarding notification to the WHO, and implementation of WHO recommended measures, distribution of information, and coordination of input from several key national areas, such as disease surveillance, ports, airports, and ground crossings' public health services, as well as other government departments, such as agriculture and customs; and finally, act as the technical resource coordinating body during the revision and implementation processes.

Third, *core country capacities required in surveillance and response, including at points of entry* will be defined and included in the IHR. In order for urgent national events to be picked up early, each country will require a surveillance system informing on unusual and unexpected events from the periphery into the center in a very short time, including the capacity to analyze rapidly such data. In many countries, this surveillance/analysis capacity may already be in place. Others may need a grace period to fulfill this future IHR requirement, and external assistance and funding may become necessary.

The 43rd Meeting of the Pan American Health Organization (PAHO) Directing Council adopted Resolution CD43.R13 in support of the revision of the International Health Regulations (IHR), urging Member States to participate actively in the review process both nationally and through the regional integration systems.

In the face of the risk posed by the emergence and re-emergence of infectious diseases, PAHO has focused its tech-

nical cooperation efforts on building a national and subregional capacity to detect, investigate, and control events related to epidemic-prone diseases through emerging disease surveillance networks.

PAHO has also been working with Member States to obtain their comments on the proposed revisions and to keep them informed on the progress made. Moreover, PAHO has taken the opportunity to discuss the IHR revision in working groups on health that were created within the subregional integration systems.

One of these groups has been the *Mercado Común del Sur* (MERCOSUR), which includes the Southern Cone countries (Argentina, Brazil, Paraguay, and Uruguay, with Bolivia and Chile as observers). This group has provided insight into the proposed changes and has taken concrete steps regarding the IHR, such as: including the Regulations as a priority topic of its Surveillance Working Group; pledging unanimous support to the revision process, especially as it refers to border health and its trade components; conducting four workshops resulting in resolutions and agreements signed by the Ministries of Health; carrying out country activities including the revision of national norms for port-of-entry sanitation and travelers' health certificates; testing syndromic surveillance at the national level; and testing the algorithm for reporting events of international public health concern.

Another is the Andean subregion comprising Bolivia, Chile, Colombia, Ecuador, Peru and Venezuela; its *Organismo Andino de Salud*, has included the IHR revision on its health agenda. Through a cooperative agreement with PAHO, it has organized two workshops on the subject to inform the countries of the revision of the IHR, to initiate a national process to bring together interested parties, and to obtain national views regarding the proposed changes. Two ministerial resolutions emerged from this initiative. The first one established national technical task forces and the second urged countries to review and strengthen epidemiological surveillance, especially in border areas.

Next steps

According to the present schedule, the next major milestones in the revision process include:

- Distributing the first draft of the reviewed IHR in all official WHO languages by December 2003
- Convening regional and subregional consultation meetings regarding the proposed changes by June 2004
- Delivering the final draft of the IHR to every country by November 2004
- Discussing the Project Proposal of the new IHR at the World Health Assembly in May 2005

Reference:
(1) World Health Organization. Resolution of the Executive Board of the WHO. Revision of the International Health Regulations. Geneva: WHO; January 2003. (EB111.R13)

Source: Prepared by PAHO's Area of Disease Prevention and Control, Communicable Diseases Unit (DPC/CD).

Summer Courses in Epidemiology, 2004

McGill University, Department of Epidemiology, Biostatistics and Occupational Health - XIX Annual Summer Program in Epidemiology and Biostatistics

Place: Montreal, Quebec, Canada

Dates: 31 May to 20 June 2004.

The information on courses offered will be available soon at: http://www.mcgill.ca/epi-biostat/summer/

For more information, please contact: Summer Programme Coordinator, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Purvis Hall, 1020 Pine Avenue West, Room 38-B, Montreal, Québec, Canada, H3A 1A2, Tel: (514) 398-3973, Fax: (514) 398-4503
Email: summer.epid@mcgill.ca,
Website: http://www.mcgill.ca/epi-biostat/summer/

Harvard University School of Public Health - XI Annual Summer Session for Public Health Studies

Place: Boston, Massachussets, USA
Dates: 1 July to 13 August 2004.

Courses in biostatistics, environmental health, epidemiology, health policy and management and social behavior.

For more information, please contact: Hildi Keary, Administrative Assistant for Summer Programs, Summer Session for Public Health Studies, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115-6023, Tel: (617) 432-1052, Fax: (617) 432-2009
Email: hkeary@hsph.harvard.edu,
Website: http://www.hsph.harvard.edu/summer/brochure/

Johns Hopkins University Bloomberg School of Public Health - XXII Graduate Summer Institute in Epidemiology and Biostatistics

Place: Baltimore, MD, USA

Dates: 21 June to 9 July 2004.

Courses offered:

Three-week courses
Principles of Epidemiology

Intermediate Epidemiology
Infectious Disease Epidemiology
Epidemiologic Basis for Tuberculosis Control
Statistical Reasoning in Public Health I
Statistical Reasoning in Public Health II

Two-week courses
Regression Analysis in Public Health Research
Regression Analysis Laboratory

One-week courses
Applications of the Case-Control Method
Methods and Applications of Cohort Studies
Conducting Epidemiological Research
Clinical Trials: Issues and Controversies
New Perspectives on Management of Epidemiologic Studies
Data Analysis Workshop I and II
Survival Analysis
Multilevel Models
Introduction to the SAS Statistical Package
Epidemiologic Applications of GIS
Family Based Genetic Epidemiology
Molecular Biology for Genetic Epidemiology
Genetic Epidemiology in Populations
Gene Expression Data Analysis
Epidemiology in Evidence-Based Decisions
Outcomes and Effectiveness Research
Pharmacoepidemiology
Epidemiologic Methods for Planning and Evaluating Health Services
Epidemiology of HIV/AIDS
Nutritional Epidemiology
Tobacco Control: National and International Approaches
Ethics Issues of Human Subjects Research in Developing Countries

Weekend course
Clinical Research: An Introduction to Design and Analysis

For more information, please contact: Ayesha Khan, Program Coordinator, Graduate Summer Institute of Epidemiology and Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD, 21205, USA. Tel: (410) 955-7158; fax: (410) 955-0898.
Email: akhan@jhsph.edu
Website: www.jhsph.edu/summerepi

UNIVERSITY OF MICHIGAN SCHOOL OF PUBLIC HEALTH - XXXIX GRADUATE SUMMER SESSION IN EPIDEMIOLOGY

Place: Ann Arbor, Michigan, USA

Dates: 11 to 30 July 2004

Topics offered:

One and three-week courses: Fundamentals of Biostatistics and Epidemiology, Infectious Diseases, Epidemiology in Public Health Practice, Cancer, Injuries, Clinical Trials, Computer Applications, Epidemiologic Measures, Logistic Model, Environmental and Occupational Epidemiology, Behavioral Change, Violence, Health Economics, Social Epidemiology, Longitudinal Studies, PC-SUDAAN, Global Health, Genetics, Epidemiology of Bioterrorism, Pharmacoepidemiology, and Geographic Information Systems.

Special Weekend Courses: Introduction to Genetics in Epidemiology, Occupational Risks for Health Workers, Advances in the Epidemiology of Diabetes, and Integration of Biomarkers in Epidemiologic Services.

For more information, please contact: Jody Gray, Graduate Summer Session in Epidemiology, The University of Michigan School of Public Health, 109 Observatory St., Ann Arbor, MI 48109-2029, Tel: (734) 764-5454, Fax: (734) 764-3192
Email: umichgss@umich.edu

The *Epidemiological Bulletin* is grateful to the following individuals for their reviews and comments on articles published in 2003: Jaume Canela (PAHO), Vance Dietz (PAHO), Ana Diez-Roux (Columbia University, New York, USA), John Ehrenberg (PAHO), Charles Eisner (PAHO), José Antonio Escamilla (PAHO/Brazil), Gabriela Fernández (PAHO), Margaret Hazlewood (PAHO), Patricia Ruiz (PAHO), Roberto Salvatella (PAHO/Panaftosa), Gabriel Schmunis (PAHO), Juan Sentís (Universidad de Barcelona, Spain), Clovis H. Tigre (PAHO), Enrique Vázquez (PAHO/Argentina), Helen Walters (Johns Hopkins University, Baltimore, USA), Thomas Yerg (PAHO).

PAHO's Epidemiological Bulletin is published quarterly in English and Spanish. Catalogued and indexed by the United States National Library of Medicine. Printed on acid-free paper.

Internet: http://www.paho.org/english/DD/AIS/beindexe.htm

PAN AMERICAN HEALTH ORGANIZATION
Pan American Sanitary Bureau, Regional Office of the
WORLD HEALTH ORGANIZATION
525 Twenty-Third Street, N.W.
Washington, DC 20037