TREND ANALYSIS AND INTERPRETATION

KEY CONCEPTS AND METHODS FOR
MATERNAL AND CHILD HEALTH PROFESSIONALS

Division of Science, Education and Analysis
Maternal and Child Health Bureau

Maternal and Child Health Information Resource Center

Trend Analysis and Interpretation

Prepared by:

Deborah Rosenberg, Ph.D.
School of Public Health
University of Illinois at Chicago

December 1997

Contract No. 240-94-0047

Table of Contents

INTRODUCTION

Public health agencies have a long tradition of monitoring trends in rates of disease and death and trends in medical, social, and behavioral risk factors that may contribute to these adverse events. Trends in observed rates provide invaluable information for needs assessment, program planning, program evaluation, and policy development activities. Examining data over time also permits making predictions about future frequencies and rates of occurrence.

Typically in public health, trend data are presented for rates arising from large populations over relatively long periods of time (e.g. ten or more years). For example, the national vital records system is a source for trend analysis of infant mortality and other death rates. The national rates described in these analyses are very reliable and are available over many years insuring a precise characterization of changes over time. These rates are considered as the true underlying population parameters and therefore statistical assessment, which implies that the data are subject to sampling error, is rarely undertaken. If rates are assumed to be error-free, they can be presented "as is" in tables or graphs, and comparisons across populations or predictions of future occurrence can be made intuitively.

In contrast to descriptive trend analysis, research studies of changes over time have followed a somewhat different analytic course. Because research data are usually sample data, statistical procedures, including sophisticated approaches such as time series analysis and formal forecasting techniques, are commonly used. Time series analysis is also used when the analytic goal is not monitoring the trend in an outcome indicator per se, but rather describing the relationship between a risk factor and the outcome. For example, in a study of the association between air pollution levels and the hospitalization rates of children with asthma, time series analysis might be used to simultaneously *control for* the overall trends in both measures; here, the impact of the trends on the hypothesized association is the primary focus, rather than the trends themselves.

With the recognition that intervention strategies are more effective as they become more targeted and specific, and as public health decisionmaking shifts to the local level, the context for analyzing trends is changing. The public health community is increasingly interested in examining trends for smaller populations and in smaller geographic areas. In maternal and child health, for example, describing trends in perinatal outcomes in relation to trends in prenatal care utilization within and across local service delivery areas is critical for program planning. There is also interest in examining trends in indicators of emerging health problems which, by definition, are only available for short periods of time. Describing trends in substance abuse during pregnancy or trends in antepartum medical risk factors--indicators that in most states are only available since the revision of the birth certificate in 1989--is essential for monitoring health status.

Once the focus of trend analysis is on data from small areas, small populations, or for a narrow range of time, it is necessary to draw from both the classic descriptive methods and the statistical approaches used in research studies. As numbers get smaller, for example, confidence in their

accuracy is reduced and although no formal sampling has been carried out, it becomes more obvious that there is potential for error nonetheless. Reporting confidence intervals, or using other statistical methods to assess and compare trends becomes critical.

Consider the neonatal mortality rates of 3.9 and 3.0 per 1,000 live births to Native American women in 1990 in federal Regions V and VII respectively. These rates are based on 13 and 3 neonatal deaths and 3,364 and 993 live births. If, by chance, one less infant born to a Native American woman in Region V were to have died during the neonatal period, and one more infant born to a Native American woman in Region VII were to have died during the neonatal period, the neonatal mortality rates would have been 3.6 and 4.0, a reversal of the ordering seen in the observed data.

Region V

Actual Rate   **Hypothetical Rate**

$$\frac{13}{3,364}\times 1,000 = 3.9 \qquad\qquad \frac{12}{3,364}\times 1,000 = 3.6$$

Region VII

Actual Rate   **Hypothetical Rate**

$$\frac{3}{993}\times 1,000 = 3.0 \qquad\qquad \frac{4}{993}\times 1,000 = 4.0$$

The Region V and VII numbers and rates are drawn from data reported by the Midwest Maternal and Child Health Data Improvement Project (MMDIP) , University of Illinois at Chicago, School of Public Health.

This illustrates how erratic rates based on small numbers can be. If the number of neonatal deaths is considered an unbiased *sample estimate*, rather than a population parameter, common statistical procedures could be used to calculate confidence limits and to test for differences between Region V and Region VII. Actually, all population numbers and rates, regardless of population size, are subject to errors of this type. When numbers are large, however, the errors are negligible and ignoring them, as has been the usual practice when presenting public health trend data, has little impact on findings and interpretation. When numbers are small, on the other hand, the impact of errors can be important.

To better understand why statistical approaches may be appropriate for trend analysis, it is useful to think of population numbers and rates as samples in time and space. For example, ten years of infant mortality data for one geographic area can be viewed as a sample taken from a multi-year, multi-area population, and as such the rates are subject to sampling error. In addition, the number of infant deaths and/or the number of live births might be slightly different in a given time or place depending on indeterminate conditions. If a year could be "sampled" repeatedly,

an infant born at 11:59 on December 31st in one "sample" might be born at 12:00 on January 1st in another and thus be in the subsequent birth cohort, or an infant might survive in one "sample", but die in another since the probability of survival for an infant born with a given set of medical conditions is not fixed. These potential fluctuations may be termed *random error* as opposed to *sampling error*, but regardless, health events are seen as the result of dynamic, not fully predictable processes.

Error also plays a role when trend data are available for only a short period of time since the ability to accurately characterize the overall shape of the trend is compromised by the reduced number of data points. For example, the maternal mortality rate in the U.S. has declined from slightly over 20 per 100,000 in 1970 to 7.9 per 100,000 in 1990. Most of this decrease, however, occurred prior to 1980. Since then, the rate has barely declined at all. Both the long term and the short term information are important to understanding maternal mortality. If data were only available since 1980, our view of the pattern of decrease over time in this indicator would be limited. On the other hand, future maternal mortality rates may be predicted more accurately if only the data for recent years is used since it is a fair assumption that the rates in subsequent years will be more similar to those in close proximity than to those in the more distant past. Statistical of dynamic, not fully predictable processes.methods can be used to model trend data in various ways, incorporating appropriate assumptions about the nature of the trend in the past, present and future.
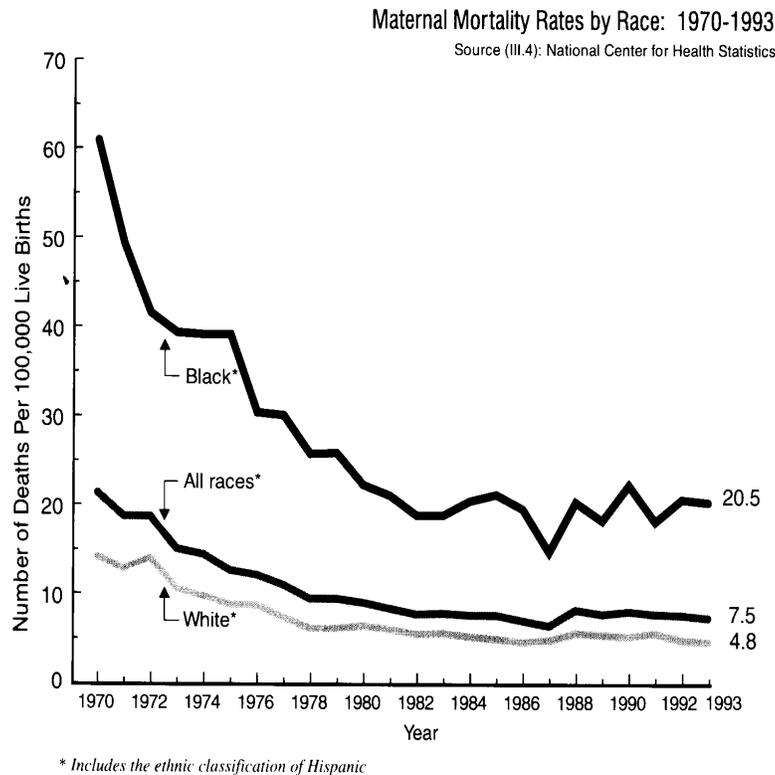
Maternal Mortality Rates by Race: 1970-1993
Source (III.4): National Center for Health Statistics

* Includes the ethnic classification of Hispanic

The changing terrain of public health calls for new strategies for examining trend data. Public health professionals need guidelines for presenting trend data based on small numbers or short time periods, and principles for interpreting these types of trend data. The number of observations that give rise to the rates of interest, the extent to which the number of observations varies across populations or geographic areas to be compared, and the number of time points available for study all influence how public health agencies will analyze and report trend data.

This discussion explores conceptual and methodological issues pertaining to analyzing trend data. Advantages and disadvantages of approaches for increasing the stability of rates such as averaging data over time and / or geography, and statistical methods such as regression analysis are described. Interpretation and presentation issues are addressed throughout. Sixteen years of infant mortality data reported by the Chicago Department of Public Health are used to illustrate the analytic approaches described in the text.

❑ Why Do Trend Analysis?

One of the hallmarks of epidemiologic analysis is the understanding that health outcomes in a population can only be fully understood if their frequency and distribution is examined in terms of person, place, and time. Trend analysis is one leg of this analytic triangle, and is used for public health surveillance and monitoring, for forecasting, for program evaluation, for policy analysis, and for etiologic analysis (investigation of potentially causal relationships between risk factors and outcomes). A study of time trends may focus, therefore, on one or more of the following:

The overall pattern of change in an indicator over time. The most general goal of trend analysis for public health surveillance is to discern whether the level of a health status, service, or systems indicator has increased or decreased over time, and if it has, how quickly or slowly the increase or decrease has occurred.

Comparing one time period to another time period.   This form of trend analysis is carried out in order to assess the level of an indicator before and after an event. Evaluating the impact of programs, policy shifts, or medical and other technical advances may call for what is sometimes called interrupted time series analysis.

Comparing one geographic area to another. When comparing the level of an indicator across geographic areas, only looking at one point in time can be misleading. For instance, one area may have a higher value on an indicator in one year, but a lower value in the next--analyzing the trend over several years can give a more precise comparison of the two areas.

Comparing one population to another. When comparing the level of an indicator across populations, both absolute and relative differences are important. For instance, one population may have consistently higher rates over time compared to another population and the rates in both populations may be decreasing over time, but the disparity between the rates in the two populations at each point in time may be increasing or decreasing. Analyzing the trend over time can provide information about the changing rates and the changing disparity in the rates.

Making future projections. Projecting rates into the future is a means of monitoring progress toward a national or local objective or simply providing an estimate of the rate of future occurrence. Projecting the potential number of future cases can aid in the planning of needed health and other related services and in defining corresponding resource requirements.

❑ Preparing to Analyze Trend Data

A series of conceptual issues must be addressed before analyzing and interpreting trend data regardless of the purpose of the analysis. These issues include:

- Sample size—the number of time periods being examined
- Presence of extreme observations or outliers
- Availability of numerator and denominator data
- Confounding--changes over time in factors related to the indicator of interest

Sample size. First, it is critical to understand the nature of the dataset being analyzed. In public health, trend analysis is typically carried out at the ecologic level. In other words, the observations, or units of analysis, are <u>time periods</u> (years, months, days) and not <u>individuals</u>. A simple dataset for use in a trend analysis might be organized in the following way:

	Year	Rate per 10,000	# of Health Events (numerator)	Population at Risk (Denominator)
1	1970	13.3	26	1,955
2	1971	12.8	24	1,877
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
25	1994	11.8	24	2,034
26	1995	12.0	26	2,164

In this hypothetical dataset, there are 26 observations, one for each year. In statistical terms, these 26 observations are a sample in time, and therefore 26 is the sample size for analysis regardless of the size of the population denominators. The fewer the number of time periods available, the smaller the sample size and as usual, the greater the potential for error. The longer the time period, the more information and therefore the more likely it is to precisely identify patterns of change. Referring again to the graph of maternal mortality shown earlier, we see that if only data from 1980 forward were examined, the lack of improvement during this time period could be misinterpreted as representing the longer term trend which, in contrast, shows considerable progress in preventing this adverse outcome.

Presence of extreme observations. Another consideration when analyzing and interpreting trends over time is whether there are extreme observations, or outliers, in the data. If there are, it is important to determine whether these are due to random variability or whether they reflect a real departure from the general trend. In the graph of the annual number of measles cases seen below, understanding the peak in 1989 and 1990 is as important as understanding the overall

long term trend. In this case, the maternal and child health community used this information to focus more intensively on immunization efforts.

Reported Number of Measles Cases: 1980-1992

Source (III.2): Centers for Disease Control and Prevention;
Interagency Committee to Improve Access to Immunization Services:
National Vaccine Advisory Committee
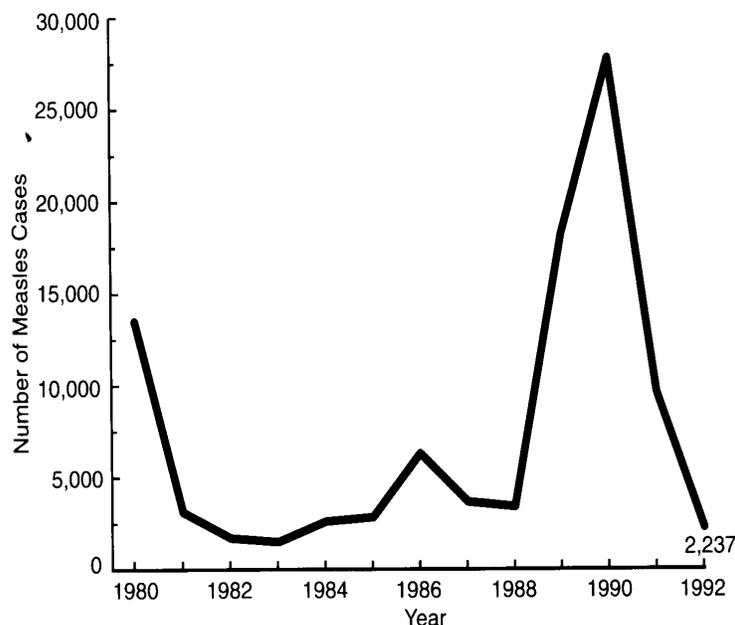
Reprinted from *Child Health USA '93*, U.S. Department of Health and Human Services, Public Health Service, Health Resources and Services Administration, Maternal and Child Health Bureau, DHHS Publication No. HRSA-MCH-94-1.

Availability of numerator and denominator data. The accuracy of numerator and denominator information over time is also very important in insuring meaningful interpretation of trend data. If both numerator and denominator data for an indicator are available for each time period being studied, then these can be easily analyzed. For instance, it is relatively straightforward to examine trends in indicators based on vital records data, such as low birthweight or infant mortality, because the numerator and denominator information is collected continuously. The number of low birthweight births or the number of infant deaths as well as the total number of live births can be accurately counted on an annual or even monthly basis.

In contrast, other indicators require population denominators which are not collected continuously. Census data are collected once every 10 years, with intercensal estimates being calculated routinely only for certain geographic areas and certain populations. If a trend analysis in a small area, or comparing different age groups or sociodemographic groups is planned, it is typically quite difficult to obtain estimates for the required denominators. For example, annual estimates of the population of children less than 18 years old in a small city are probably unavailable.

When intercensal estimates are available, they have typically been calculated using what is called the Component Method. This method uses data for births, deaths, and migration patterns (in and out) in a given population to yield revised population estimates. Data from a decennial census is used as baseline data (Year 1), and then the Component Method calculates estimates for subsequent years as follows:

$$\textbf{Estimate}_{\textbf{Year 2}} = \textbf{Population}_{\textbf{Year 1}} + \textbf{Births}_{\textbf{Year 2}} \quad \textbf{- Deaths}_{\textbf{Year 2}} \quad + \textbf{Net Migration}_{\textbf{Year 2}}$$

$$\textbf{Estimate}_{\textbf{Year 3}} = \textbf{Population}_{\textbf{Year 1}} + \textbf{Births}_{\textbf{Years 2 and 3}} \textbf{- Deaths}_{\textbf{Year 2 and 3}} + \textbf{Net Migration}_{\textbf{Year 2 and 3}}$$

and so on until the final year of interest is reached.

Birth and death information is readily available for all time periods, but migration patterns are quite difficult to estimate. When the Component Method is not feasible, other approaches turn to proxy indicators of population such as school enrollment, number of filed tax returns, or car registrations to derive population estimates.

Less precise, but easiest to carry out, are simple linear interpolation or extrapolation. Linear interpolation and extrapolation both involve the calculation of average annual % change for use in estimating population denominators.

<div align="center">Linear Interpolation</div>

When data are only available for the beginning and end of a time period of interest, the formula for calculating an average annual % change is as follows:

$$\textbf{Average Annual \% Change} = \left(\frac{\dfrac{\textbf{Final Value} - \textbf{First Value}}{\textbf{First Value}}}{\textbf{n} - \textbf{1}} \right) \times \textbf{100}$$

$$= \left(\frac{\dfrac{\textbf{Final Value}}{\textbf{First Value}} - \dfrac{\textbf{First Value}}{\textbf{First Value}}}{\textbf{n} - \textbf{1}} \right) \times \textbf{100}$$

$$= \left(\frac{\dfrac{\textbf{Final Value}}{\textbf{First Value}} - \textbf{1}}{\textbf{n} - \textbf{1}} \right) \times \textbf{100}$$

<div align="center">*where n = the number of years,*
inclusive of the first and final years</div>

Suppose available census estimates of the number of children under 18 years of age in a community are:

$$
\begin{array}{ll}
1990 & 3,224 \\
1995 & 3,943
\end{array}
$$

The average annual percent change can then be calculated as:

$$
\frac{\dfrac{3,943 - 3,224}{3,224}}{5} \times 100
$$

$$
= 4.46
$$

Once the average percent change is calculated, it can then be applied to the baseline value, yielding a constant for use in interpolating the population values for the intervening years. In other words, to get estimates for the 4 years between the known values for 1990 and 1995, the average annual percent change of 4.5 is applied to the number of children < 18 in 1990.

$$
\textbf{Constant} = \textbf{average annual \% change} \times \textbf{Baseline Value (1990)}
$$

$$
\textbf{144} = \textbf{0.0446} \times \textbf{3,224}
$$

This estimate of a yearly increase of 145 children < 18 can now be added year by year to obtain new annual estimates. In general this process occurs as follows:

$$
\textbf{Estimate}_{1991} = \textbf{Value}_{1990} + \textbf{Constant}
$$

And, then again:

$$
\textbf{Estimate}_{1992} = \textbf{Estimate}_{1991} + \textbf{Constant}
$$

This process continues until the known value for 1995 is reached. For estimating the population of children under 18, we obtain:

$$
\textbf{For 1991:} \quad \textbf{3,368} = \textbf{3,224} + \textbf{144}
$$

$$
\textbf{For 1992:} \quad \textbf{3,512} = \textbf{3,368} + \textbf{144}
$$

$$
\textbf{For 1993:} \quad \textbf{3,656} = \textbf{3,512} + \textbf{144}
$$

$$
\textbf{For 1994:} \quad \textbf{3,800} = \textbf{3,656} + \textbf{144}
$$

Linear interpolation assumes that the change over time is occurring uniformly--the change from the first to the second year is the same as the change from the second to the third year, and so on. Any year to year differences are unseen, but as long as these differences can be assumed to be small, this method will yield reasonable population estimates.

Even less precise than linear interpolation, sometimes decennial census figures are used as population denominators for a whole series of years. If the populations of interest are assumed to be fairly stable, using the decennial census figures as stand-ins may be reasonable, but if the population is changing rapidly, this will result in errors and misinterpretation of trends. If this approach is taken, it is incumbent upon the analyst to describe the potential bias in the results. For example, if a single decennial census figure is used as the population denominator for a five year period, but the population actually increased during that period, then any rates reported will be overestimates, and conversely, if the population actually decreased during that period, then any rates reported will be underestimates.

Confounding. Particularly when trend analysis is to be undertaken for a small area or small population, changes over time in other factors related to the indicator of interest must also be considered. For example, change in the sociodemographic characteristics of the population such as change in the age structure, the ethnic composition, or income level over time may be associated with the change over time in the indicator that is of primary interest. The question is whether comparing the health status of a community from 1970 to 1995, for example, is meaningful. Is it really the "same" community at the two endpoints of the trend analysis, or is any observed trend confounded by changes in factors other than the indicator being studied?

Changes in reporting definitions for an indicator, or reporting accuracy over time might also confound the trend information and lead to misinterpretation. For instance, varying criteria for reporting fetal deaths versus spontaneous abortions may produce spurious results when trend analysis for these indicators is carried out. Similarly, changes in medical technology which have an impact on the indicator, such as the impact on infant mortality of the widespread implementation of neonatal intensive care units in the late 1970's and early 1980's, must also be considered. Other changes over time that are of concern when conducting trend analysis include changes in laws or public policy such as a change in eligibility criteria for programs such as Medicaid, or rules concerning access to services for minors, as well as changes in cultural practices such as reduction (or increase) in substance abuse.

In general, accounting for change over time in other factors related to the indicator of interest increases in importance the longer the time period and/or the smaller the area or population size being studied. Demographic changes such as a shift in ethnic composition, for instance, may be more pronounced at the county or community level than in a state or the nation as a whole. On the other hand, a shift in cultural practices usually occurs gradually and will be identified only over a long period of time.

❑ Analysis of Trend Data

The selection of a strategy for analyzing trend data will depend in part on the purpose of the analysis, and on careful consideration of all of the issues discussed above. Once there is a sound conceptual framework, tables, graphs and statistical analysis are tools for examining and analyzing trend data; graphs, in particular, are an effective tool for presenting the pattern of change over time.

Regardless of whether statistical techniques will be used for analyzing data over time, the most straightforward and intuitive first step in assessing a trend is to plot the actual observed numbers or rates of interest by year (or some other time period deemed appropriate). In addition, the numbers or rates should be examined in tabular form.

These initial steps are indispensable for understanding the general shape of the trend, for identifying any outliers in the data, and for becoming familiar with both the absolute and relative levels of the numbers and rates being studied. Inspection of the data provides the basis for making subsequent analysis choices and should never be bypassed.

Visual inspection of the data may indicate that use of statistical procedures is inappropriate. For example, an outlier at a particular point in time may be of intrinsic interest, representing circumstances in the community important to understand. Statistical procedures, such as regression analysis, would mute the effect of this outlier by assuming that its extreme value is attributable to statistical instability. For example, a sudden rise in the rate of injury deaths due to drowning may point to a problem at community swimming facilities and it would be unfortunate and inappropriate if a statistical examination of an overall trend were to obscure this information.

Visual inspection also permits a preliminary assessment of the overall direction and shape of the trend. Commonly used statistical approaches are designed for assessing *linear* trends; that is, a series of numbers or rates which change over time in a consistent, or uniform, fashion. If the trend appears to be different during distinct time periods, either in shape or direction, then an analytic method must be selected that will preserve and not obscure this important information.

In addition, when the number of health events is very small, individual case review may be called for rather than the application of statistical methods. Each event might appropriately be considered as a sentinel or unexpected occurrence, arising from different processes than when health events occur on a widespread basis.

In order to look more concretely at the process of analyzing trends, data reported by the Chicago Department of Public Health are used throughout the remainder of this discussion. The data presented here are for what is referred to as Chicago Community Area 33, representing a small area of the city. Chicago Community Area 33 is an aggregation of five census tracts. As recommended, we first look at the actual data, both in tabular and graphic form:

INFANT MORTALITY RATES
CHICAGO COMMUNITY AREA 33 1979-1994

	DEATHS	BIRTHS	RATE / 1,000	95% CI*
1979	8	213	37.6	11.6-63.6
1980	7	220	31.8	8.2-55.4
1981	6	179	33.5	6.7-60.3
1982	7	202	34.7	9.0-60.4
1983	5	172	29.1	3.6-54.6
1984	8	184	43.5	13.4-73.6
1985	4	183	21.9	0.5-43.3
1986	4	189	21.2	0.4-42.0
1987	7	194	36.1	9.4-62.8
1988	5	181	27.6	3.4-51.8
1989	9	227	39.6	13.7-65.5
1990	6	252	23.8	4.8-42.8
1991	4	247	16.2	0.3-32.1
1992	3	246	12.2	0.0-26.0
1993	10	244	41.0	15.6-66.4
1994	6	275	21.8	4.3-39.3

$$\textbf{* 95\% Confidence Interval (CI)} = \textbf{Rate} \pm \textbf{1.96} \sqrt{\frac{\textbf{rate}}{\textbf{pop}} \times \textbf{1,000}}$$

$$\textbf{For example, for 1994, the 95\% CI} = \textbf{21.8} \pm \textbf{1.96} \sqrt{\frac{\textbf{21.8}}{\textbf{275}} \times \textbf{1,000}}$$

$$= \textbf{21.8} \pm \textbf{1.96} \times \textbf{8.9}$$

$$= \textbf{21.8} \pm \textbf{17.5}$$

$$= \textbf{4.3} - \textbf{39.3}$$

Infant Mortality

Observed Rates by Year
Chicago Community Area 33

Infant Deaths per 1,000 Live Births
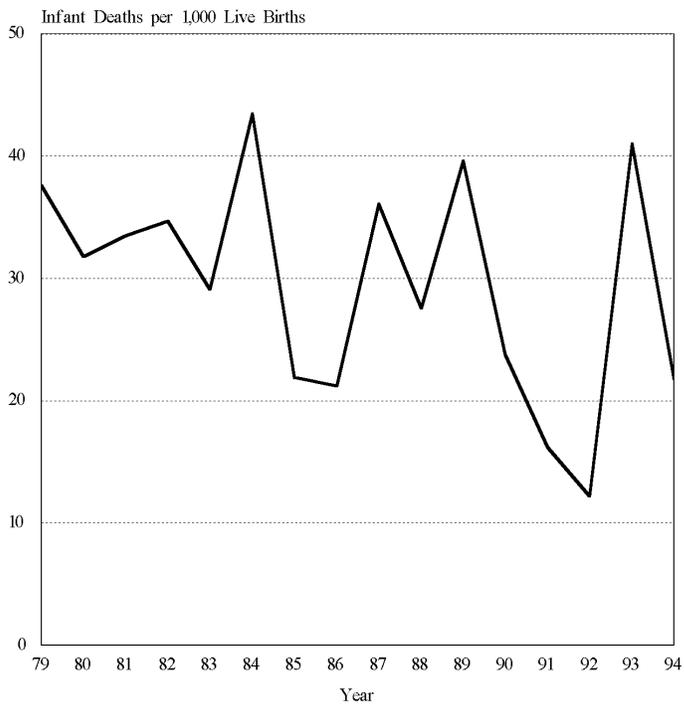
Infant Mortality

Observed Rates and Confidence Limits by Year
Chicago Community Area 33

Infant Deaths per 1,000 Live Births

13
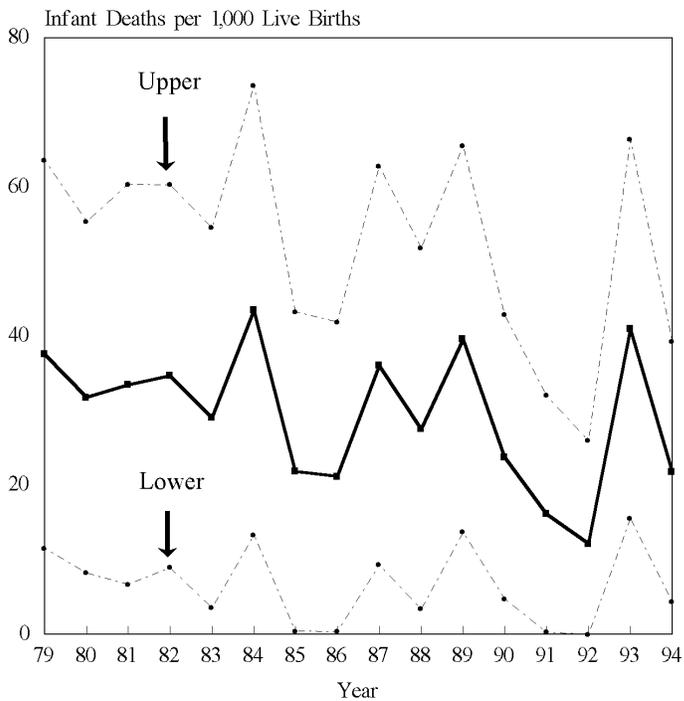
These preliminary views of the community area infant mortality data show a series of unstable rates—the confidence intervals around each rate are very wide and the pattern over time is quite jagged. Ten or fewer infant deaths and fewer than 300 live births occur each year in this area. Due to this instability, it is very difficult to meaningfully interpret the data and other techniques have to be applied in order to create a clearer picture of the pattern over time. (Note that the Y axis scaling on the two graphs is different; the scale on the second graph had to be expanded in order to accommodate the wide confidence limits.)

Data Transformation and Smoothing

One step toward improving the interpretability of the data is to put the rates on a logarithmic scale. A log transformation of the data provides more appropriate and realistic results because it "flattens" the series of rates. While the overall shape of the trend is unchanged, the rate of increase or decrease is somewhat altered. For example, if rates are decreasing over time and no transformation is carried out, future projections will eventually predict the occurrence of zero health events, but the log transformation will slow the approach to zero (and in fact never reach zero) making any projection of future rates more reasonable.

Infant Mortality
Observed Rates by Year on Linear and Log Scales
Chicago Community Area 33

14
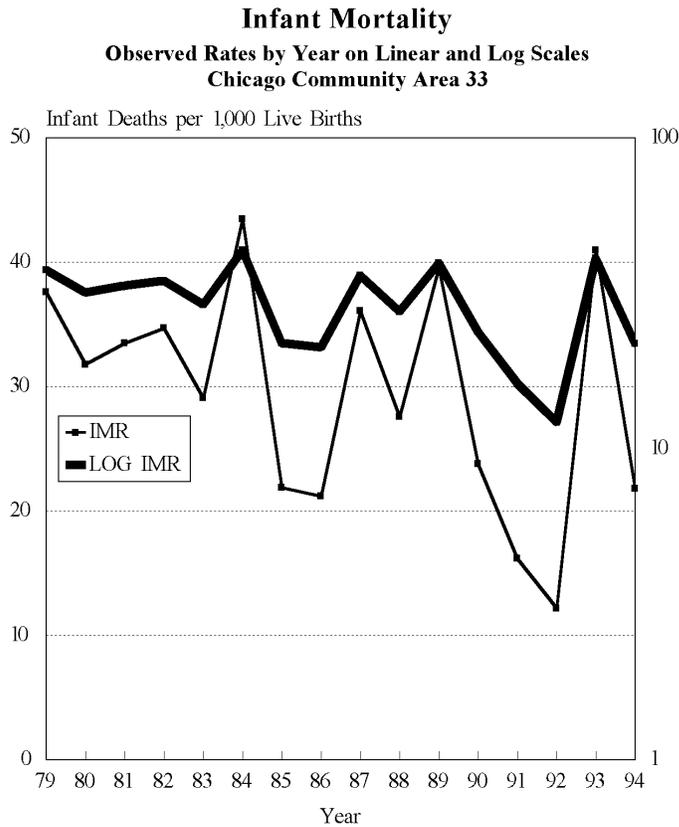
Next, if the initial assessment does not reveal any outliers in the data, more elaborate statistical methods beyond simple transformation of the rates may be worthwhile. In general, statistical approaches aim to "smooth" the data; that is, they aim to increase the stability of the rates and hence diminish their jagged appearance. Various forms of averaging, including use of multiple year rates, moving averages, and regression procedures can accomplish the desired smoothing.

Perhaps the form of averaging most commonly carried out by MCH and other public health professionals is collapsing data across time periods--for instance, calculating rates by combining the numerators and denominators for two or three years of data rather than using the annual rates. This approach increases the stability of the resulting rates by increasing the sample size at each time point. The infant mortality rate for 1979-1980 in Chicago Community Area 33 is:

$$\frac{8+7}{213+220} \times 1{,}000 = \frac{15}{433} \times 1{,}000 = 34.6$$

Collapsing data in this fashion, however, though it increases stability, also means a loss of information. With 16 years of data, for example, using two year rates leaves only 8 data points to portray the pattern over time. If more years of data were combined, the loss of information would render it difficult to discern any pattern. Below is a plot of the infant mortality data for Chicago Community Area 33 using two year rates.

Infant Mortality

2 Year Observed Rates
Chicago Community Area 33

Infant Deaths per 1,000 Live Births

Year
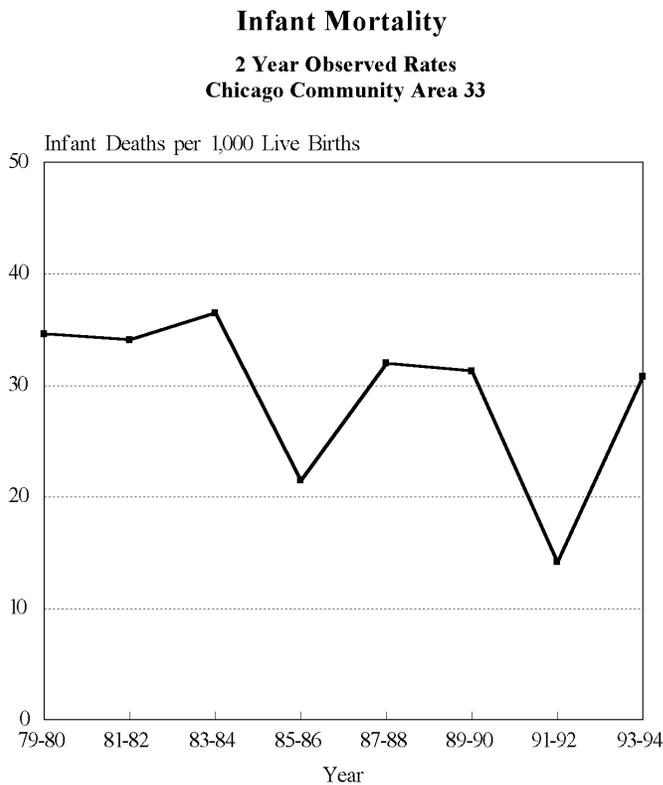
Using simple averages results in a "smoother" line
but also results in loss of information

15

Moving averages are also used to increase stability and "smooth" the data. Here, time periods are not combined in mutually exclusive groups as when simple multiple year rates are calculated, but rather in overlapping sequences:

> Years 1,2, and 3
> Years 2,3, and 4
> Years 3,4, and 5
> etc. until the last year is included

Moving averages have the advantage of increasing stability with a minimal loss of information. Compared with the 8 data points resulting from calculation of two year mutually exclusive rates as shown above, 15 data points result from calculation of rates based on two year moving averages. The maintenance of information with moving averages makes it possible to combine more years of data to maximize sample size at each point. For instance, calculation of three year moving averages with the 16 years of data results in only 2 data points being lost, leaving 14 to portray the pattern over time. Below is a plot of the infant mortality data for Chicago Community Area 33 using three year moving averages.
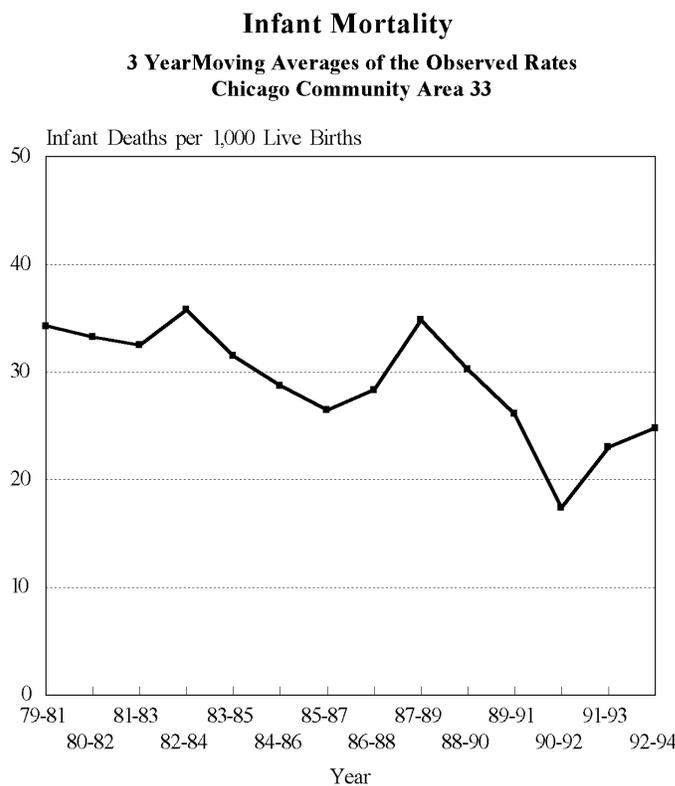
Infant Mortality
3 YearMoving Averages of the Observed Rates
Chicago Community Area 33

Using moving averages results in a "smoother" line
with minimal loss of information

16

Reporting Average Annual Percent Change in an Indicator

Earlier, average annual percent change was discussed in the context of linear interpolation when estimates of population denominators were required in order to assemble reasonable trend data. Using average annual percent change after the trend data has been assembled and organized is somewhat different. Here, there is information on the year to year changes rather than just the overall change from the beginning to the end of a time period and the goal is to estimate how fast or slow change is occurring. The formula for average annual % change in this context is:

$$\frac{\left(\sum_{i=2}^{n} \frac{Rate_{Year\ i}}{Rate_{Year\ i-1}} - 1\right) \times 100}{n-1}$$

where n is the total number of years
Year 1 is the first year
Year n is the final year

If numbers or rates are unstable and therefore the pattern very jagged, the average of the year to year changes if no preliminary smoothing is carried out can be misleading. For example, the average of the observed set of 15 annual % changes for the 16 infant mortality rates in Community Area 33 is calculated as:

$$\frac{\left[\left(\frac{31.8}{37.6}-1\right)+\left(\frac{33.5}{31.8}-1\right)+\ldots+\left(\frac{41.0}{12.1}-1\right)+\left(\frac{21.8}{41.0}-1\right)\right] \times 100}{15} = 10.5$$

The result of 10.5 % implies that the infant mortality rates are increasing over time, a result which contradicts the pattern of decreasing rates over time seen in the plots shown earlier.

Using the 13 three year moving averages to gain stability and smoothness, however, the average annual % change is:

$$\frac{\left[\left(\frac{33.3}{34.3}-1\right)+\left(\frac{32.5}{33.3}-1\right)+\ldots+\left(\frac{23.1}{17.4}-1\right)+\left(\frac{24.8}{23.1}-1\right)\right] \times 100}{13} = -1.1$$

This implies that the infant mortality rates are decreasing by approximately 1.1 %, a result much more consistent with what is seen in the plots of the data.

Average annual percent change can also be used to calculate a projected rate. This is accomplished through linear extrapolation, by iteratively increasing or decreasing the number or rate according to the average annual percent change. For the infant mortality rates based on 3 year moving averages in Community Area 33, the process would be as follows:

$$\textbf{Projected Rate}_{\textbf{93-94-95}} = \textbf{Observed Rate}_{\textbf{92-93-94}} + (\textbf{-0.011} \times \textbf{Observed Rate}_{\textbf{92-93-94}})$$

And, then again:

$$\textbf{Projected Rate}_{\textbf{94-95-96}} = \textbf{Projected Rate}_{\textbf{93-94-95}} + (\textbf{-0.011} \times \textbf{Projected Rate}_{\textbf{93-94-95}})$$

And this would continue for as many years as considered appropriate. For Community Area 33, the first two projected rates are:

$$\textbf{Projected Rate}_{\textbf{93-94-95}} : \textbf{24.5} = \textbf{24.8} + \left(-\textbf{0.011} \times \textbf{24.8} \right)$$

$$\textbf{Projected Rate}_{\textbf{94-95-96}} : \textbf{24.2} = \textbf{24.5} + \left(-\textbf{0.011} \times \textbf{24.5} \right)$$

And if this process of linear extrapolation were continued, the projected infant mortality rate for the three year period 1999-2000-2001 would be 22.7.

In order to calculate a confidence limit around a projected rate arrived at in this fashion, the average annual % change in the population denominator has to be computed and then linear extrapolation used to obtain an estimate for the projected year of interest. For the infant mortality data, for instance, the average annual percent change in total live births (or the average change in the 3 year moving total live births) would be calculated and then applied iteratively into the future. A 95 % confidence interval can then be generated with the usual formula:

$$\textbf{Projected Rate} \pm \textbf{1.96} \sqrt{\frac{\textbf{Projected Rate}}{\textbf{Projected Live Births}} \times \textbf{1000}}$$

18

Statistical Procedures

In order to test whether there is a statistically significant trend or whether two or more trends are statistically different, several approaches are available.

Chi-square test for linear trend. A chi-square test for trend can be obtained by organizing the observed data into a contingency table with one row for each time period and two columns: the first for the number of individuals who experienced the health outcome (the numerator of the rate); the second for the number of individuals who did not (the denominator minus the numerator). Following is modified output from Epi-Info showing the chi-square test for trend for the 16 annual infant mortality rates in Chicago Community Area 33:

Analysis For Linear Trend In Proportions

Exposure Score	Cases	Controls
1979.00	8.00	205.00
1980.00	7.00	213.00
1981.00	6.00	173.00
1982.00	7.00	195.00
1983.00	5.00	167.00
1984.00	8.00	176.00
1985.00	4.00	179.00
1986.00	4.00	185.00
1987.00	7.00	187.00
1988.00	8.00	176.00
1989.00	9.00	218.00
1990.00	6.00	247.00
1991.00	4.00	243.00
1992.00	3.00	243.00
1993.00	10.00	234.00
1994.00	6.00	269.00

Chi Square for linear trend : 1.583   p value : 0.20838

The linear trend in these data is not statistically significant according to this test since the p-value is 0.2, not less than the customary cutoff of 0.05.

Regression analysis. Regression analysis has several advantages over the other averaging or smoothing techniques discussed so far and over the chi-square test for trend. In general, regression modeling has the advantage of jointly considering the information contained in the series of counts or rates, rather than considering each time point separately. Analyzing the series of rates as a unit in effect imposes stability, and consequently, the confidence band around the set of predicted values from regression analysis will be narrower than the confidence limits

calculated around each count or rate separately; any statistical test based on regression results, therefore, will be more powerful.

In addition, regression procedures will generate estimates of future rates as well as estimates of average annual percent change, neither of which are automatically obtained when averaging without modeling is carried out or when a contingency table is organized to conduct a chi-square test for trend. Moreover, graphs of the predicted and projected values as well as the confidence band can be plotted.

Another advantage of using regression methods for analyzing trends and making projections is that other variables can be included in a model. For instance, if annual infant mortality rates are to be modeled, annual data on rates of prenatal care utilization, or rates of substance abuse during pregnancy, or changes in demographic risk markers could be simultaneously examined. With trend data, regression procedures model the association between time and the outcome of interest; if other variables are confounders or effect modifiers of this association, the predicted rates, the projected rates, and the confidence bands around these will be adjusted appropriately. Without using some form of regression modeling, the impact of other variables cannot be accounted for in the results.

There are several regression approaches that can be employed to examine trend data. Following is a generic description of these approaches.

Ordinary Least Squares (OLS) Regression

Ordinary least squares (OLS) regression can be used to model the observed series of rates, but as stated above, it is often preferable to model the natural logarithm of the rates. The disparity between predicted rates based on linear and log-linear models increases as the time periods being forecasted move further from the time of observation. Multiple year rates or moving averages (or the log transformation of these) may also be modeled using the OLS method. A simple OLS model has the general form:

For modeling the actual rates

$$\textbf{rate}_\textbf{i} = \textbf{Intercept} + (\textbf{Slope} \times \textbf{Year}_\textbf{i})$$

For modeling the natural logarithm of the rates :

$$\textbf{ln}(\textbf{rate}_\textbf{i}) = \textbf{Intercept} + (\textbf{Slope} \times \textbf{Year}_\textbf{i})$$

where i = 1 to the number of years being analyzed

The figure below again shows the observed rates for Chicago Community Area 33 along with both the predicted values (regression line) obtained from modeling the actual rates and the predicted values (regression line) obtained from modeling the natural logarithm of the rates. You can see that the regression line from modeling the actual rates crosses the regression line

from modeling the transformed rates in approximately the year 2006. If the line for the actual rates is used, the infant mortality rate is predicted to be zero by the year 2023, whereas it is predicted to be approximately 8 per 1,000 live births in that same year if the regression line from the log transformed rates is used. In fact, using the log transformation, an infant mortality rate of zero will never be predicted, which unfortunately, is a more realistic result.

Infant Mortality
Observed Rates by Year
and Linear and Log Linear Regression Lines
Chicago Community Area 33

Infant Deaths per 1,000 Live Births

rate

ln(rate)

Year
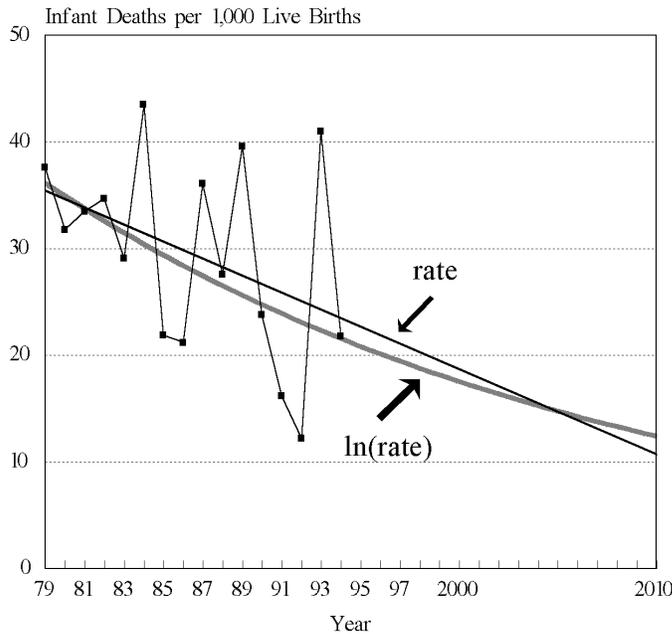
From the point of view of an OLS regression procedure, the sample size equals the number of time points being modeled. The procedure does not have access to information on the populations sizes that gave rise to the rates at each time point. In other words, OLS regression accounts for the variability across time periods, but cannot account for the variability or random error in each individual rate.

When modeling the 16 years of Chicago data with OLS, then, the sample size equals 16 regardless of whether the population (in this case, live births) in the denominator of each rate is in the 100s, 1,000s, or 100,000s. The confidence band around the OLS regression line will therefore be identical for sets of the same rates even if these arise from populations of different size. For example, if another community area in Chicago had exactly the same set of rates as those for Community Area 33, but these were based on many more live births each year, OLS regression would not distinguish between the two areas.

21

To illustrate, suppose that the 1994 rate of 21.8 in another community area was based on 27,500 rather than on 275 live births. The OLS procedure would not recognize this difference; the sample size remains 16, and the regression lines, the confidence bands, and other statistics such as average annual percent change would be identical. This is despite the fact that there is much less potential for error in the set of more stable rates compared with the set of unstable ones.

Further, if two year rates were modeled in order to gain stability, the sample size would be reduced from 16 to only 8 data points. While the confidence band around the line would be narrower as a result of the smoothing, there would be loss of statistical power due to the smaller sample size. Again, modeling moving averages has the advantage of both increasing stability and maintaining sample size. Below is the plot of the ordinary least square regression results of modeling the natural logarithm of the 3 year moving averages for Chicago Community Area 33.

Infant Mortality
3 Year Moving Averages, Regression Line and Confidence Band
Modeling the Natural Log Transformation
Chicago Community Area 33
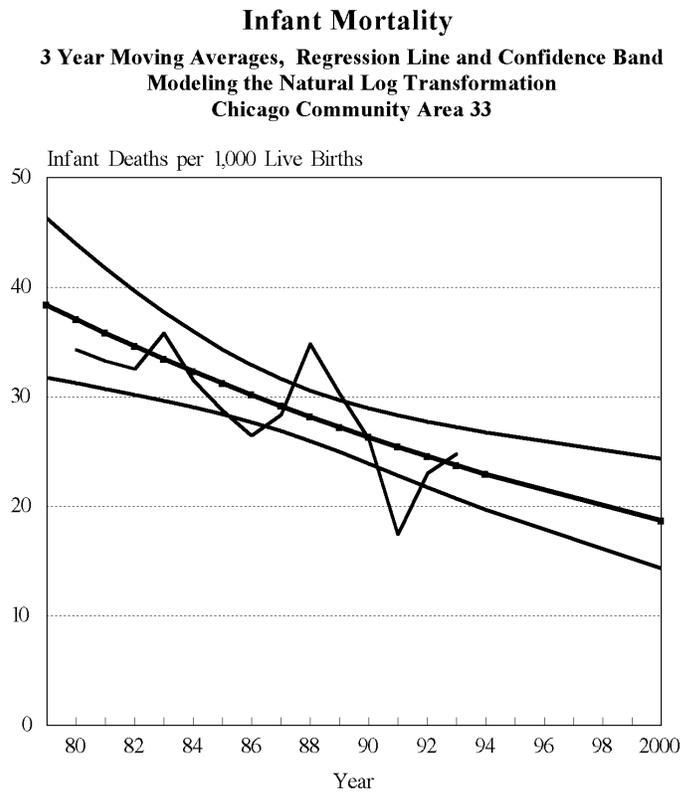
Poisson Regression

Poisson regression can also be used to model the observed rates, collapsed data, or moving averages. Because of the nature of the Poisson distribution, by definition a log transformation of the data is employed.

In contrast to ordinary least squares regression, Poisson regression has the advantage of accounting both for the fluctuation across time and the variability at each time point. The procedure models the counts in the numerator and denominator for each time period rather than the pre-calculated rates. In essence, this approach is equivalent to having data at the individual level, with sample size equaling the number of *individuals* in the denominators of the rates instead of the number of *time periods* for which rates are available.

For example, a Poisson regression procedure will model the 6 infant deaths and 275 births that occurred in Chicago Community Area 33 in 1994 or the 600 infant deaths and 27,500 live births that occurred in some hypothetical community area, rather than modeling the rate representation of 21.8 for both areas as is done in OLS. The confidence band around a Poisson regression line, therefore, will be different for sets of the same rates when they arise from populations of different size. A simple Poisson model has the general form:

$$\ln\left(\frac{\#\text{ of health events}_i}{\#\text{ in population at risk}_i}\right) = \textbf{Intercept} + \left(\textbf{Slope} \times \textbf{Year}_i\right)$$

$$\textbf{where i} = 1 \textbf{ to the number of years being analyzed}$$

Despite the differences between the OLS and Poisson procedures, in most circumstances the results from using OLS to model the natural logarithm of rates or Poisson regression will be quite similar. Understanding when and how these two regression procedures might differ, however, is important for choosing an appropriate method to use.

When a set of rates is based on small numbers, the variability of each rate is large, and the variability across time will probably also be large (a jagged pattern). In this scenario, the confidence band from either Poisson or OLS regression will be wide since both the small numbers and the jagged pattern over time imply instability. The confidence band from Poisson regression will be somewhat wider because it incorporates information about both kinds of variability, while OLS only incorporates information about the variability across time, but the difference will not be great.

On the other hand, in the improbable scenario that a set of rates based on small numbers shows little variability across time (a smooth pattern), the confidence band from Poisson regression will be quite a bit wider than that from OLS. Here, the small numbers and the smooth pattern carry contradictory information about stability. Poisson regression calculates a confidence band based on both pieces of information, but places heavier weight on the population sizes with, in this case, their implication of instability, while OLS regression calculates a narrower confidence band based solely on the smooth pattern over time with its implication of stability.

When a set of rates is based on large numbers, the variability of each rate is small, and the variability across time will probably also be small (smooth pattern). In this scenario, the confidence band from either Poisson or OLS regression will be narrow since both the large

numbers and the smooth pattern over time imply stability. In fact, the confidence bands will be almost identical because the variability of each rate is very close to zero.

On the other hand, in the improbable scenario that a set of rates based on large numbers shows a lot of variability across time (jagged pattern), the confidence band from Poisson regression will be quite a bit narrower than that from OLS. Now the large numbers and the jagged pattern carry contradictory information about stability. Poisson regression calculates a confidence band based on both pieces of information, but again places heavier weight on the population sizes with, in this case, their implication of stability, while OLS regression calculates a wider confidence band based solely on the jagged pattern over time with its implication of instability.

Because it is much more likely that the variability across time and the variability in the rates are consistent with each other, rather than contradictory, the results from Poisson and OLS regression will usually yield similar results. Often, in public health agencies, OLS regression is used because this procedure is available in many different software packages, whereas Poisson regression is available only in more specialized software.

Time Series Analysis

Time Series Analysis refers to a collection of specialized regression methods that use integrated moving averages and other smoothing techniques and have different assumptions about the error structure of the data. The term "moving average" as used in time series analysis should not be confused with the calculation of moving averages as has been demonstrated above. Here, the term refers to a complex process that incorporates information from past observations and past errors in those observations into the estimation of predicted values.

Unlike ordinary least squares, logistic or Poisson regression which assume that errors in the modeled observations are independent (uncorrelated), time series methods assume that the errors are correlated. Time series methods can diagnose the precise nature of the correlation and adjust for it. In general, correlation in the errors increases the variance of the regression estimates, and therefore using these methods will yield wider confidence bands than the other approaches already discussed. The regression estimates themselves will also be somewhat different, and despite the wider confidence band, the efficiency of time series methods may result in a steeper slope than that generated by OLS or Poisson regression.

In many public health datasets, the extent of correlation in the errors is not always high. If possible, statistical software such as SAS should be used to diagnose whether correlated errors are a significant problem. If they are not, then OLS or Poisson regression can be used without seriously violating the assumption of independence inherent in these approaches.

There are many other features of time series methods that aid in precisely characterizing a trend, including identification and adjustment for non-linear components of the trend and differential

weighting of observations for forecasting. While time series methods yield more precise results, their use also requires specialized software and more advanced technical expertise, and therefore, building them into routine surveillance and monitoring activities is rarely feasible. If these methods are to be used, the assistance of a statistician should be sought.

The following table provides a comparison of the results obtained from using various regression approaches to model the sixteen years of infant mortality data for Chicago Community Area 33. For simplicity, only the results for 1994 are shown. The OLS and Poisson results shown are quite similar. The time series approach yields a similar predicted value, but a wider confidence band as expected.   Despite the differences, these approaches will probably not lead to different public health conclusions.

Estimated Rates and 95% Confidence Bands
Various Regression Approaches
Chicago Community Area 33

Method	1994
OLS Regression, Modeling the Natural Logarithm of the Rates	21.5 15.4-30.2
Poisson Regression	23.7 16.4-34.2
Time Series Analysis--the AR(1) Model Using the Natural Logarithm of the Rates	19.6 8.9-43.2
OLS Regression, Modeling the Natural Logarithm of 3 Year Moving Averages.	22.9 (19.7-26.7)

An estimate of average annual % change and projected rates can be obtained directly from the output of a regression procedure rather than having to conduct the iterative calculations demonstrated earlier. The following are regression results generated by SAS for modeling the natural logarithm of the 16 infant mortality rates for Chicago Community Area 33:

ANNUAL IM RATES FOR CHICAGO COMMUNITY AREAS

AREA=33

Model: MODEL1
Dependent Variable: LNIMRATE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.40427	0.40427	3.722	0.0742
Error	14	1.52046	0.10860		
C Total	15	1.92472			

Root MSE	0.32955	R-square	0.2100
Dep Mean	3.32866	Adj R-sq	0.1536
C.V.	9.90043		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >]T]
INTERCEP	1	6.311364	1.54815934	4.077	0.0011
YEAR	1	-0.034482	0.01787243	-1.929	0.0742

As was seen in the table comparing different regression approaches, the predicted value for 1994 using OLS regression to model the natural logarithm of the infant mortality rates is:

$$e^{a+b \times YEAR}$$

$$e^{6.311+(-0.0345 \times 94)} = 21.5$$

Remember that because the natural logarithm of the infant mortality rates were modeled, the regression estimates need to be exponentiated in order to report the results in the usual units.

The average annual percent change based on the predicted values from the log linear regression analysis is:

$$(e^{-0.0345} - 1) \times 100 = -3.4$$

This formula is exactly the same as the one used before when the calculation for average annual percent change was carried out iteratively with the observed data--here the beta coefficient (slope or parameter estimate) from the regression results is the ratio of the rate in one year to the rate in the previous year (the relative risk). This is seen more clearly if the formula is written as follows:

$$\left(\frac{e^{6.311+(-0.0345\times94)}}{e^{6.311+(-0.0345\times93)}} - 1 \right) \times 100$$

$$= \left(e^{-0.0345\times(94-93)} - 1 \right) \times 100$$

$$= \left(e^{-0.0345} - 1 \right) \times 100$$

$$= -3.4$$

Notice that modeling the 16 annual log transformed rates, the average annual % change is negative, indicating a decreasing infant mortality rate. This is in contrast to the result obtained earlier when the average annual percent change was computed without any modeling or smoothing. Without modeling, moving averages had to be used in order to generate a reasonable result, but here the regression procedure itself accomplished the necessary smoothing. Of note is that the regression estimate of a 3.4% annual decrease in infant mortality is quite a bit higher than the estimate of a 1.1 % annual decrease obtained earlier when using the 3 year moving averages without any modeling. This result reflects the assumption of linear change imposed by the regression procedure.

Using the results of regression analysis, projected rates can also be calculated directly from the regression equation. The calculation is the same as for any predicted value, except that the time period of interest is outside the range of the observed data. Following are 2 projected rates using the log linear regression results:

The projected rate for 1995 : $e^{\,6.311+(-0.0345\times95)} = 20.8$

The projected rate for 2000 : $e^{\,6.311+(-0.0345\times100)} = 17.5$

Consistent with its higher estimate of average annual percent change, the regression projected rate of 17.5 for the year 2000 is lower than the projected rate of 22.7 calculated earlier for the year 2000 when using the 3 year moving averages without any modeling.

Other Considerations When Selecting Graphing or Statistical Approaches

Whether only one series of rates or multiple series of rates are being analyzed. For example, the choice of how many years, if any, to combine when calculating moving averages may change. It may be determined that 3 year moving averages are appropriate for analyzing data from a particular geographic area. Another area, however, may require the use of 4 year moving averages in order to achieve adequate smoothing. If the two areas are to be compared, the same amount of averaging should be applied to both.

If Chicago data were to be analyzed for the city as a whole, no averaging at all would be necessary since the stability of the city rates yields a smooth pattern. If, however, a comparison were to be made between the city and one or more community areas with their jagged patterns, then averaging would have to be applied to all of the data, and applied in a comparable fashion. Below are three plots of the infant mortality rates in Chicago Community Area 33, Chicago Community Area 35, and the city as a whole. The first examines the three sets of annual observed rates:

Infant Mortality

Observed Rates by Year
Chicago Community Areas 33, 35, and All of Chicago
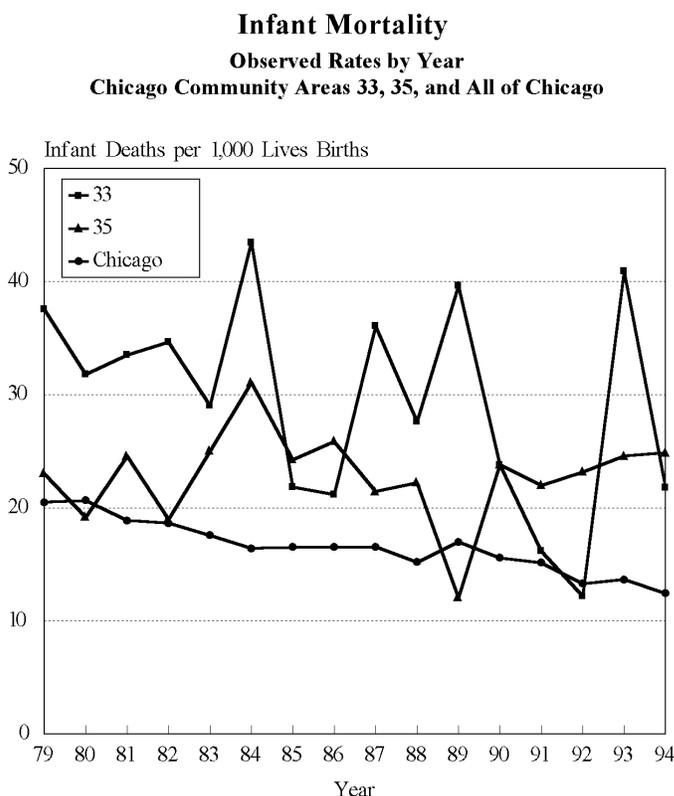
The second plot examines the three sets of 3 year moving averages, and the third the predicted values from modeling the natural logarithm of these using ordinary least squares regression:

Infant Mortality

3 Year Moving Averages of the Observed Rates by Year
Chicago Community Areas 33, 35, and All of Chicago

Infant Deaths per 1,000 Live Births
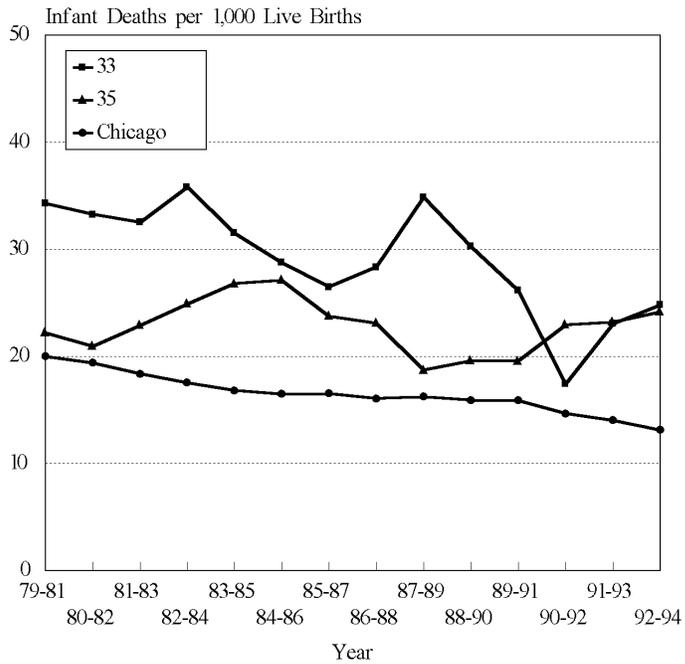
Using moving averages results in a "smoother" line
with minimal loss of information

Infant Mortality

Regession Lines
Modeling the Natural Log of the 3 Year Moving Averages
Chicago Community Areas 33, 35, and All of Chicago

Infant Deaths per 1,000 Live Births
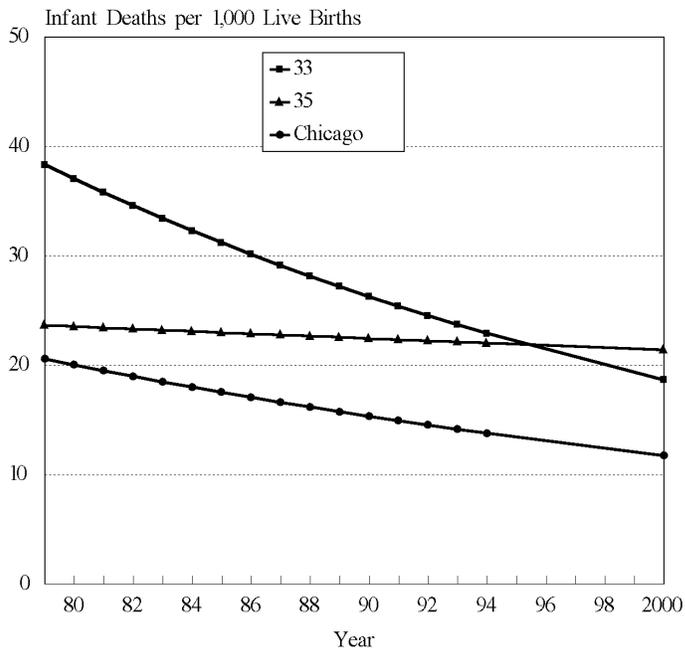
Note the projection for the Year 2000

29

The first plot is difficult to interpret since there is so much fluctuation in the rates in the two community areas. The comparison between the city and the two community areas is easier to see in the second and third plots. It appears that Community Area 33 has generally had a higher infant mortality rate than Community Area 35, although, particularly in the plot of the regression analysis, it appears that the rate is decreasing faster in Community Area 33.

Characteristics of other independent variables to be included in a regression model. It might be desirable to build a more complicated regression model that would include several independent variables in addition to the time variable. Inclusion of other variables can yield a more precise picture of a trend by accounting for changes over time in other factors. A model of this type might look like:

$$\ln(\text{rate}_i) = \text{intercept} + (\text{slope}_1 \times \text{Year}_i) + (\text{slope}_2 \times \text{Var1}_i) + (\text{slope}_3 \times \text{Var2}_i)$$

Remember that trend analysis is generally carried out using aggregate data. Independent variables in the model, therefore, are typically proportions and not values for individuals. For example, a model might include the annual percent of poverty in a community or the percent of women who did not receive prenatal care in a community.

Sometimes there are as many data points for these other variables as there are for the rates of the outcome being modeled; sometimes, however, there are fewer data points. For example, with 16 years of data, there may be only 2 census values available for another variable of interest. In addition, an independent variable may also need to be transformed, for instance, to the log scale prior to analysis. Moreover, in epidemiologic terms, time is the "exposure" variable in trend analysis, and therefore another variable must be related to both time and the outcome being studied in order to be a confounder of the association between these two (the trend). To be an effect modifier, the rate of change in the outcome over time (the slope of the trend) must be different at each level of another variable.

Achieving stability by combining geographic areas. In addition to or instead of smoothing trend data by using multiple year rates or moving averages, geographic areas can be combined to gain stability. As with any averaging technique, there is loss of information when areas are combined, but if the areas are similar with respect to the characteristics of their residents and / or the characteristics of their health service delivery systems, this is a viable approach.

Combining geographic areas can reduce the data burden when many comparisons are of interest, and can provide a more concise picture of the trend in a health indicator. Further, areas may be combined in a way that is most informative for health assessment, planning, and evaluation. Below are two plots of the infant mortality rates in Chicago Community Areas 33 and 35 combined, and the city as a whole. Although these two areas showed somewhat different trends in infant mortality as seen in the previous graphs, they are very similar demographically, and combining them may make more sense from the standpoint of intervention strategies. The first plot examines the 3 year moving averages of the observed rates, and the second examines the regression lines and confidence bands from modeling the natural logarithm of these.

Infant Mortality

3 Year Moving Averages of the Observed Rates by Year
Chicago Community Areas 33 & 35 combined, and Chicago

Infant Deaths per 1,000 Live Births
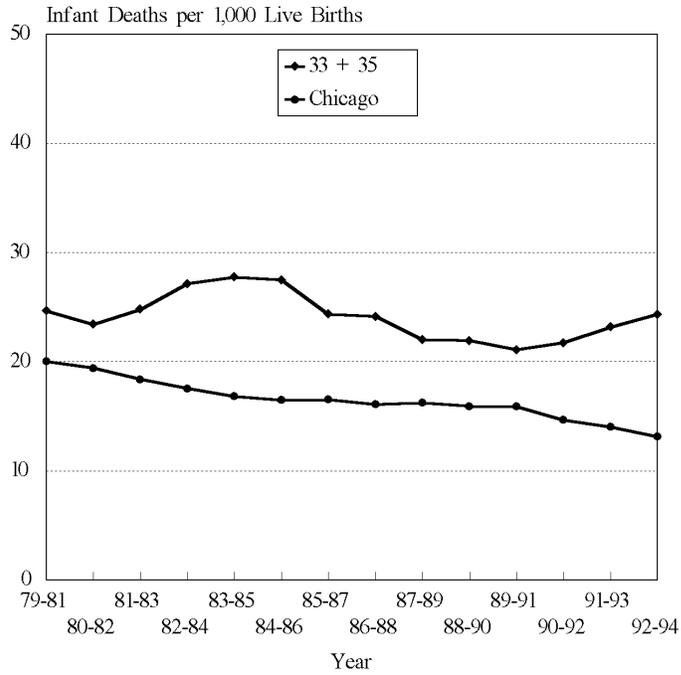
Year

Infant Mortality

Regression Lines and Confidence Bands
Modeling the Natural Logarithm of the 3 YearMoving Averages
Chicago Community Areas 33 & 35 combined, and Chicago

Infant Deaths per 1,000 Live Births

Year

31
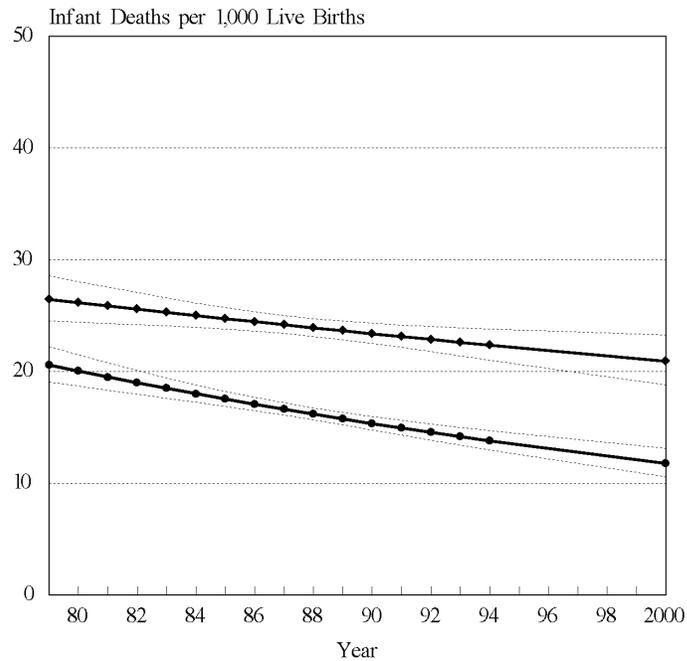
Trend Analysis and Use of Local, State, or National Objectives

Using trend data, several indicators can be projected out to the point at which each meets or surpasses the appropriate Year 2000 Objective (or any other objective), and then a comparison across these indicators can be made according to the number of years behind or ahead of the objective. For example, suppose trend analysis yielded the following projections for three indicators:

$$\text{Indicator \#1 -- 2004}$$
$$\text{Indicator \#2 -- 1999}$$
$$\text{Indicator \#3 -- 2003}$$

In this scenario, priority might be given to improving indicator #1 (assuming other factors have been taken into account) since it appears to be farther from its Year 2000 Objective than the other two indicators.

In addition, statistical testing can be carried out to assess the current level of indicators with respect to the appropriate Year 2000 (or other) Objective. The resulting z-tests can then be compared across indicators. The formula for a z-test, using an objective as a standard is as follows:

$$z_1 = \frac{\textbf{Indicator}_{1\,\textbf{current}} - \textbf{Objective}_1}{\sqrt{\dfrac{\textbf{Objective}_1}{\textbf{n}_{1\,\textbf{current}}} \times \textbf{multiplier}}}$$

$$z_2 = \frac{\textbf{Indicator}_{2\,\textbf{current}} - \textbf{Objective}_2}{\sqrt{\dfrac{\textbf{Objective}_2}{\textbf{n}_{2\,\textbf{current}}} \times \textbf{multiplier}}}$$

where n_1 and n_2 are the population denominators
giving rise to Indicator$_1$ and Indicator$_2$ respectively

Further, z-tests can be generated using projected levels of the indicators instead of the current levels. Comparisons can then be made across indicators according to these scores:

$$z_1 = \frac{\text{Indicator}_{1\,\text{projected}} - \text{Objective}_1}{\sqrt{\dfrac{\text{Objective}_1}{n_{1\,\text{projected}}} \times \text{multiplier}}}$$

$$z_2 = \frac{\text{Indicator}_{2\,\text{projected}} - \text{Objective}_2}{\sqrt{\dfrac{\text{Objective}_2}{n_{2\,\text{projected}}} \times \text{multiplier}}}$$

**where n_1 and n_2 are the population denominators
giving rise to Indicator_1 and Indicator_2 respectively**

The results of using the current level of the indicator and the projected level of the indicator may yield different, but equally important information. One indicator may currently be farther from an objective than another indicator, but it may also be exhibiting a faster rate of decline over time and therefore be projected to be closer to its objective than the other indicator at some point in the future. These types of analyses can complement the information obtained from examining plots of trend data, comparing average annual percent changes or other statistical results.

❑ **Presentation of Trend Data**

The discussion thus far has described a range of analytic methods useful for understanding and interpreting trend data. Deciding which results to present and the form in which to present them is an important aspect of turning the data into information. As a general rule, it is important that public health analysts within a health agency carry out as refined and as detailed an analysis as is feasible, even though the final reports or forums for presenting the information may incorporate only portions, sometimes only small portions, of the work. Examining the data in multiple and varied forms allows for exploration of different presentation strategies, and is necessary for selecting an effective and useful approach. If analytic shortcuts are taken, this process cannot occur and the final product may suffer.

For instance, initial analysis may explore approaches to combining years or geographic areas. One averaging method may obscure an important trend, another may highlight it; unless the preliminary work is carried out, inappropriate choices for presentation may be made. In addition, having multiple analyses provides a pool from which to draw depending on the audience and circumstances of the presentation. If time and resources permit, then, the following analyses should be conducted for internal use:

1. Produce plots of the observed data along with confidence limits
2. Produce tables of the observed data along with confidence limits
3. Assess whether there are extreme observations, or outliers, in the data
4. Consider transforming the data to a logarithmic scale
5. Assess whether the trend is linear or whether its shape or direction changes over time
6. Consider examining trends separately for subsets of the population
7. If data are based on small numbers, explore different smoothing strategies:
   a. Combine multiple years of data into mutually exclusive groups
   b. Combine multiple years of data into moving averages
   c. Combine geographic areas
8. Calculate average percent change, either annual or some other time increment
9. Calculate projected rates, particularly with respect to relevant health objectives
10. Use statistical testing, such as the chi-square test for trend or regression analysis
11. Explore different formulations for regression modeling
    a. Produce plots of the predicted and projected values, and confidence bands
    b. Explore the impact on the trend of including other variables in a model
    c. Calculate average percent change, either annual or some other time increment

Once the above analyses are generated, the following are general guidelines for presentation to a wider audience:

At Minimum	Optional
Display plots of the observed data over time	Display a table of the rates along with confidence limits
Comment in narrative form on the stability of the rates and approaches used to increase it	Display plots of predicted and projected values from regression analysis
Report average percent change	Overlay plots for different populations or geographic areas
Interpret in narrative form the relationship of the trend to reaching health status objectives, to health services utilization, and to systems functioning	Report the results of statistical testing, either for one trend or for comparisons across populations or geographic areas

Although the observed data, if based on small numbers, may be very hard to interpret due to a jagged pattern over time, most audiences need to see the original "real" numbers or rates before they can understand transformed data or regression results. For example, a community group may wish to see data for its area alone, even though a clearer picture might be obtained if data for several communities were combined. Often, it is decided not to show plots of regression results, but to simply report some of the information that a regression analysis yields such as an average percent change or a prediction of a future rate.

SUMMARY

Trend data provide a dynamic rather than a fixed view of the health status of the MCH population and of the services and systems that can have an impact on that health status. For trend data to be most useful, it is critical that an analysis be conceptually tied to program and policy issues. The job of the analyst, therefore, is to present graphs, tables, statistical results, and narrative that make these connections. In particular, the ability to appropriately analyze and interpret trends for small geographic areas or for small populations is essential if program intervention strategies are to be more targeted and thus more effective.

The scientific literature contains many examples of trend analysis, both theoretical and applied. Attached is a list of articles selected for their relevance to public health practice and in particular for their relevance to maternal and child health. The list is neither comprehensive nor exhaustive, but includes articles that are illustrative of the range of analytic approaches used when examining trend data.

SELECTED BIBLIOGRAPHY

1. Alexander, J.A., M.T. Halperin, and S.Y. Lee. 1996. The short-term effects of merger on hospital operations. Health Services Research. 30: 827-847.

2. Bailey, N.T.J. 1994. Estimating HIV incidence and AIDS Projections: prediction and validation in the public health modeling of HIV/AIDS. Statistics in Medicine. 13: 1933-1943.

3. Becker, S.R., F. Diop, and J.N. Thornton. 1993. Infant and child mortality in two counties of Liberia: Results of a survey in 1988 and trends since 1984. Int. J. Epidemiol. 22: S56-S63.

4. Bhat, P.N., K. Navaneetham, and S.I. Rajan. 1995. Maternal mortality in India: estimates from a regression model. Stud. Fam. Plann. 26: 217-232.

5. Boyle, P. and C. Robertson. 1987. Statistical modeling of lung cancer and laryngeal cancer incidence in Scotland, 1960-1979. Amer. J. Epidemiol. 125: 731-744.

6. Breslow, N.E. 1984. Extra-Poisson variation in log-linear models. Appl. Statist. 33:38-44.

7. Brillinger, D.R. 1986. A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. Biometrics 42: 693-734.

8. Brookmeyer, R. and M.H. Gail. 1988. A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. J. Amer. Statist. Assoc. 83: 301-308.

9. Buescher, P.A. 1996. Smoking in pregnancy in North Carolina: maternal characteristics and trends 1988-1994. State Center for Health Statistics No. 101.

10. Catalano, R. and S. Serxner. 1987. Time series designs of potential interest to epidemiologists. Amer. J. Epidemiol. 126: 724-731.

11. Chaulk, C.P., K. Moore-Rice, R. Rizzo, and R.R. Chaisson. 1995. Eleven years of community-based directly observed therapy for tuberculosis. JAMA 274: 945-951.

12. Chin, J. and S.K. Lwanga. 1991. Estimation and projection of adult AIDS cases: a simple epidemiological model. WHO Bulletin OMS 69: 399-406.

13. Davis, H., K.C. Schoendorf, P.J. Gergen and R.M. Moore. 1997. National trends in the mortality of children with Sickle Cell Disease, 1968 through 1992. Amer. J. Public Health 87:1317-1322.

14. Devesa, S.S., J Donaldson, and T. Fears. 1995. Graphical presentation of trends in rates. Amer. J. Epidemiol. 141: 300-304.

15. Downs, A.M., R.A. Ancelle-Park, and J.B. Brunet. 1990. Surveillance of AIDS in the European community: recent trends and predictions to 1991. AIDS 4: 1117-1124.

16. Geddes, J.R. and E. Juszczak. 1995. Period trends in rate of suicide in first 28 days after discharge from psychiatric hospital in Scotland, 1968-1992. BMJ 311: 357-360.

17. Glynn, R.J., T.A. Stukel, S.M. Sharp, T.A. Bubolz, J.L. Freeman, and E.S. Fisher. 1993. Estimating the variance of standardized rates of recurrent events, with application to hospitalizations among the elderly in New England. Amer. J. Epidemiol. 137: 776-786.

18. Greenland, S. 1995. Dose-response and trend analysis in Epidemiology: alternatives to categorical analysis. Epidemiol 6: 356-365.

19. Hakulinen, T. and T. Dyba. 1994. Precision of incidence predictions based on Poisson distributed observations. Statist. In Med. 13: 1513-1523.

20. Hamers, F.F., T.A. Peterman, A.A. Zaida, R.L. Ranson, J.E. Wrotten, and J.J. Witte. 1995. Syphilis and gonorrhea in Miami: similar clustering, different trends. Amer. J. Public Health 85: 1104-1108.

21. Helfenstein, U. 1991. The use of transfer function models, intervention analysis and related time series methods in epidemiology. Intern. J. Epidemiol. 20: 808-815.

22. Hodge, M.J., G.E. Dougherty, I.B. Pless. 1995. Pediatric mortality and hospital use in Canada and the United States, 1971 through 1987. Amer. J. Public Health 85: 1276-1279.

23. Janerich, D.T. 1984. Forecasting cancer trends to optimize control strategies. JNCI 72: 1317-1321.

24. Katzenellenbogen, J., D. Yach, and R.E. Dorrington. 1993. Mortality in a rural South African mission, 1837-1909: an historical cohort study using church records. Intern, J. Epidemiol. 22: 965-973.

25. Keiding, N., P.K. Anderson, and K. Frederiksen. 1990. Modeling excess mortality of the unemployed: choice of scale and extra-Poisson variability. Appl. Stats. 39: 63-74.

26. Kleinman, J. 1991. Methodological issues in the analysis of vital statistics. In *Reproductive and Perinatal Epidemiology*, Kiely, M. ed., Boston: CRC press.

27. Kuhn, L., L.L. Davidson, and M.S. Durkin. 1994. Use of Poisson regression and time series analysis for detecting changes over time in rates of child injury following a prevention program. Amer. J. Epidemiol. 140: 943 – 955.

28. Land, G.H. and J.W. Stockbauer. 1993. Smoking and pregnancy outcome: trends among black teenage mothers in Missouri. Amer. J. Public Health. 83: 1121-1124.

29. MacNeill, I.B., J.M. Elwood, D. Miller, and Y. Mao. 1995. Trends in mortality from melanoma in Canada and prediction of future rates. Stat. Med. 14: 821-839.

30. Malcolm, M.S. and C.E. Salmond. 1993. Trends in amenable mortality in New Zealand 1968-1987. Intern. J. Epidemiol. 22: 468-474.

31. Morbidity and Mortality Weekly Report. 1994. Surveillance for selected tobacco-use behaviors – United States, 1900-1994. CDC Surveillance Summaries Vol. 43, No. SS-3.

32. Morbidity and Mortality Weekly Report. 1995. Trends in smoking initiation among adolescents and young adults-–United Stated, 1980-1989. 44: 521-525.

33. Morbidity and Mortality Weekly Report. 1995. Pertussis--United States, January, 1992-June, 1995. 44: 525-529.

34. Morbidity and Mortality Weekly Report. 1995. Pneumonia and influenza death rates–United States, 1979-1994. 44: 535-537.

35. Morbidity and Mortality Weekly Report. 1995. State-specific changes in physical activity among persons aged 65 or older--United States, 1987-1992. 44: 669-673.

36. Murray, D.M. 1995. Design and analysis of community trials: lessons from the Minnesota heart health program. Amer. J. Epidemiol. 142: 569-575.

37. Nam. J.M. 1995. Interval estimation and significance testing for cyclic trends in seasonality studies. Biometrics 51: 1411-1417.

38. Nelson, D.E., G.A. Giovino, D.R. Shopland, P.D. Mowery, S.L. Mills, and M.P. Eriksen. 1995. Trends in cigarette smoking among US adolescents, 1974 through 1991. Amer. J. Public Health. 85:34-40.

39. Opitz, W. and H. Nelson. 1996. Short-term, population-based forecasting in the public sector. Popul. Res. Policy Rev. 15:549-563.

40. Parker, R.A. 1989. Analysis of surveillance data with Poisson regression: a case study. Statist. In Med. 8: 285-294.

41. Polednak, A.P. 1994. Projected numbers of cancers diagnosed in the US elderly population, 1990-2030. Amer. J. Public Health 84: 1313-1316.

42. Rockett, I.R.H. and J.H. Pollard. 1995. Life table analysis of the United States' year 2000 mortality objectives. Intern. J. Epidemiol. 24: 547-551.

43. Ruwaar, D., R.T. Hoogenveen, H. Verkleif, D. Kromhout, A.F. Casparie, and E. A. van der Veen. 1993. Forecasting the number of diabetic patients in the Netherlands in 2005. Amer. J. Public Health 83: 989-995.

44. Sankrithi, U, I. Emanuel, and G. van Belle. 1991. Comparison of linear and exponential multivariate models for explaining national infant and child mortality. Intern. J. Epidemiol 20: 565-570.

45. Singh, G.K., and S.M. Yu. 1995. Infant mortality in the United States: trends, differentials, and projections, 1950-2010. Amer. J. Public Health 85:957-964.

46. Siskind, V., C. Del Mar, and F. Schofield. 1993. Infant feeding in Queensland, Australia: long-term trends. Amer. J. Public Health. 83: 103-106.

47. Thomas, J.C., A.L. Kulik and V.J. Schoenbach. 1995. Syphilis in the south: rural rates surpass urban rates in North Carolina. Amer. J. Public Health 85:1119-1122.

48. Tuomilehto, J., E. Virtala, M. Karvonen, R. Lounamaa, J. Pitkamiemi, R. Reunanen, E. Tuomilehto-Wolf, L. Toivanen, and the Dime Study Group. 1995. Increase in incidence of insulin-dependent diabetes mellitus among children in Finland. Intern. J. Public Health 24: 984-992.