Che-Fu Lee, The Catholic University of America

All of the five contributed papers just presented deal more directly with issues of assessing differences and changes in "vital rates" of some kind and less directly with estimates of such demographic rates. No one addresses problems of paucity of basic demographic data, which have in the past been the preoccupation of demographer-statisticians studying the demography of developing countries. If the subjects chosen by these contributors serve as an index, it perhaps signifies the recent improvement of the data situation in most of today's developing countries. Moreover, the subjects presented in these papers are of general methodological interest in demographic analysis aside from their ramification for studying the "vital rates in developing countries", as the title of this session seems to delimit. I shall now comment on each of these fine studies in their turn.

1. The title of Udry et.al.'s paper, "Random Variation in Rates Based on Total Enumeration of Events", suggests in effect three aspects of the problem in constructing a rate: a) ascertaining the random process underlying the occurrence of an event, b) identifying the appropriate population "at risk" of the event occurrence for the denominator, and c) total enumeration of the numerator, observed occurrences. Ideally, the latter two problems should have been resolved when a model random (or stochastic) process is conceptualized, and data are collected accordingly. To estimate a rate in practice, however, one often has to rely on the available records. Socially significant events like births, marriages, and deaths or accidents tend to be registered as they occur and thus are the result of a complete count. The denominator population "at risk" on the other hand may be obtained from a different source, which is not infrequently based on a sample estimate. An even-handed treatment of sampling fluctuation in the denominator and errors involved in enumeration of the numerator, which is supposedly subject to no sampling error, is itself a difficult task. Udry et.al. did not focus their study on this issue and tacitly assumed no sampling error for a computed rate when the numerator represents a total count.

The paper moved directly into a demonstration that observed variance of crude birth rates exceeds the variance that can be expected from a binomial model, even after the variations across comparative units and over time periods have been taken into account. The authors went on to suggest other sources of non-sampling variations: correlations between random error of birth rates and time; and unequal risks of birth among individuals, i.e., heterogeneity in the probability process or compound probability distributions. These, in other words, are equivalent to saying that a simple binomial model is inadequate for depicting the "true" process of birth. This is hardly a surprise to those researchers who are inclined to model building. However, the well organized exposition in this article serves its purpose in calling the practitioners' attention to an unconservative inference on difference or change in birth rates at their face values.

One may well ask a logical question: so what should be done then? The authors seemed to suggest two ways. One is to set a minimally required sample size ensuring stability of rate estimates; the other, as the authors put it, "predicted variances made from detailed data on the actual population being studied". More concrete suggestions than these open-ended ones may require more work than the scope of the paper intended by the authors. It will suffice to point out that these suggested directions for tackling the remaining problem may be more complex than they appear at first blush.

One of such difficulties was touched by the authors in their statement, "assuming that we are usually dealing with populations in which p (the rate) has some unknown distribution, our predicted variances based on the simple binomial model seem doomed to be over-estimates" (pp. 15-16). The implications of this were not pursued. Let me extend it as a query. As we conceive of the random process in terms of a more realistic and usually more complex model, the larger will become the predicted variance due to random variation. While comparing to a simple binomial assumption we tend to draw non-conservative assertions of true differences in rate. Wouldn't any refined conceptual model quickly "step up" the predicted variance and render us a "too conservative" inference, as observed variations in rates hardly ever exceed the predicted variations based on a complex model? This strikes me as a major caveat in most of the model-building exercises, e.g., birth-interval models considering the elements of fecundability, postpartum lapsed period and various outcomes of a pregnancy, etc. The most common fate of an elegantly constructed model is being shelved and never becoming useful in data confrontation, especially for detecting differences or changes in demographic rates.

The alternative to a preconceived model depicting the random process lies in data exploratory-confirmatory approach (Tukey 1970). The authors' call for studying the detailed data on the actual population may be interpreted as suggesting this line of approach, but I am not sure from reading their paper. Exploration of data distributions and boundaries of homogeneity and heterogeneity requires detailed information on differentials in rates. A pragmatic approach without having to specify the underlying random (probability) process in the first place is suggested by Allen and Avery; this leads our discussion to the next paper.

2. If a sample is drawn from the population at risk of birth, the binomial distribution of births and no births or a multinomial distribution by number of births during a period of observation can be handled as a discretely measured dependent variable, without being converted into rate measurement; and such a frequency distribution can be cross-tabulated with other categorical factors in a multiple-way contingency table and analyzed by a log-linear analysis for significant factors or interactions among them in differentiating the distributions of the observed frequency of event occurrence. This is exactly what Allen and Avery proposed in their paper.

It seems to me a promising new way of analyzing differentials in demographic measures by discretizing the occurrence of events. Allen and Avery treated the period fertility in terms of frequencies of mothers falling into a dichotomy of having had no birth and one or more births during the past five years before the Costa Rican census. This dependent variable was alternatively measured in terms of a polytomy of 0 to 5+ births. The odds of falling into one category vs. the other(s) is actually the criterion quantity to be analyzed. Since all the predicting factors selected (rural-urban residence, marital status, labor force status, and education), and the control variables (previous parity and age) are all represented as categorical measures, the log-linear model for discrete multivariate analysis seems to be suitable. The analytical results were then presented in terms of variations in odds of having any births vs. no births (or having one particular number of births vs. all others in a polytomous measure of the dependent variable), which are attributable to differences in the selected predicting factors and their interaction effects, which have been identified as statistically significant.

Judicious presentation of the statistics resulting from the log-linear analysis is essential in making important findings recognizable, as Davis (1975) once complained that such an analysis generated "too many results". Allen and Avery presented at length the variations in odds by various-ordered effects. Such odds figures filled almost six full pages of table presentation, and their graphs attempted for facilitating a visual summary of the results did not seem to alleviate much of the reader's burden in putting these tremendous numbers of odds figures into perspective. Showing the possibility of constructing the probability of birth (convertible into the familiar birth rate) from the odds figures in the appendix, the authors unfortunately failed to see the importance of presenting the "smoothed" fertility rates as differentiable by the tested factors. I am inclined to think of discrete multivariate statistics as the means and the vital rates arrivable through statistical testing and smoothing being the end. I am sure that Allen and Avery can easily produce the familiar differential birth rates following their log-linear analysis if they elected to do so. It would involve use of the model predicted frequencies, rather than the observed frequencies, and computing the rate thereof.

Just a point of information: the odds measure in the case of a dichotomous variable like mortality - death or no death, is easily interpretable. The odds measure in the case of polytomy is limited in the log-linear analysis to the odds of a chosen category to all others. This sometimes may not be easy to interpret. There is at least one alternative method for polytomous dependent variables like fertility observed over a longer period. The weighted least square approach (Grizzle et.al. 1969) to multivariate analysis of categorical data allows for the flexibility of converting a polytomous dependent variable into its expectation, i.e., the birth rate in this case. Moreover, the predicting factors are not limited to categorical measures, and the hypotheses need not be hierarchical in the weighted least square method.

3. Rashid and McElroy's comparison of the labor force separation rate for Saudi Arabia obtained through a longitudinal study and that obtained through standard working life table, seemed to have raised more questions than it answered. My first expectation from the title of this paper was in seeing a discrepancy which often results in comparisons between period and cohort rates: one being cross-sectional rates for different age groups synthesized, and the other tracing the flow through ages of an actual cohort. However, the so-called "longitudinal study" in this paper refers to a two time observation apart only by a period of 9 months. Without a detailed explanation of the computation procedures in the paper, I am at a loss in finding justification for calling such a short period data "longitudinal". The period withdrawl rates for an occupation-age category group were not clearly explained either. I was puzzled as to whether or

not the "longitudinal" meant a prospective measure comparing the job status at the end of 9 months subsequent to the beginning of the survey; and whether the rate used for working life table analysis was a retrospective job status last year compared with that at the beginning of the survey. The results of 182% difference in withdrawl rates obtained from the life table and the longitudinal data in professional, technical, and managerial category, and 60% in production workers, operatives and laborers, were indeed alarming as expressed by the authors, but no adequate explanation for these discrepancies were given. Could they be due to the current age structure of the occupational make-up: modern sector occupations are filled by younger males (e.g., the professional), so that the relatively small proportions of higher aged males overrepresent the withdrawl rate from one age level to another, in a cross-sectional comparison? Questions like these must be answered by the authors in a fuller presentation of their computational details.

4. The two papers by Kripalani and his associates on the model of population growth may be discussed together. Their simulation analysis reminded me of Frejka's (1973) work entitled "The Future of Population Growth: Alternative Paths to Equilibrium". Frejka used vital statistics available around 1965-1970 and projected the population growth to the year 2150 following alternative assumptions of reaching an equilibrium (just replacement rate, NRR = 1) immediately, in 10, 20 years or a longer period. The major innovation here is to take the initial rate of growth, instead of NRR=1, as a point of reference. It also quantifies the "lagged" response of fertility decline to the initial "disturbance" of reduced mortality by a parameter between 0 and 1; the immediate fertility reduction to offset the effect of declined mortality at one end, and no fertility response hence allowing for the full effect of the initial mortality change on the growth of population, at the other. Of course, there are finer manipulations of the input variables in the present simulation analysis than the abstract linear adjustment of fertility and mortality schedules in the population projection as conducted by Frejka. Like other well conceived projection exercises, Kripalani and his associates have added to the material that is useful for population education needed by decision makers and development planners, who are concerned with the dynamics of population growth and want to be told about differences in terms of quantitative magnitude.

It may be interesting to note that Kripalani and Smith's projections of the Indian population based on alternative assumptions of lagged fertility response ("L" ranging from 0.6 to 1.0) fall into a rather small range of variation, compared to the alternative projections carried out by Frejka, who assumed various lengths for the lapsed period before NRR=1 is reached. In terms of the projected total population in the year 2000, the four projections of Kripalani and Smith come very close to Frejka's project no. 2, which assumes that a just replacement fertility rate is attained in the years 1980-1985. I admit that there are technical problems involved in such a comparison across projections done by different demographers who all have their respective justifications in generating the projected figures of their own. What I fear is: can we expect non-demographers to understand our projection exercises, or simply tell them to make their own choice according to their own taste.

REFERENCES

Davis, James A. 1975. "Key Concepts in 'The Goodman System' for Analysing Contingency Tables: An Outline" read at Meetings of the American Sociological Association, San Francisco, Calif. August 1975

Frejka, Tomas. 1973. The Future of Population Growth: Alternative Paths to Equilibrium. New York: John Wiley.

Grizzle, J. E., C. F. Stramer and G. C. Koch. 1969. "Analysis of Categorical Data by Linear Models" Biometrics 25(Sept.):489-504.

Tukey, J. W. 1970. Exploratory Data Analysis. Reading, Mass.: Addison-Wesley.