

EPIDEMIOLOGIC IDENTIFICATION OF INFANTS WITH LOW BIRTH WEIGHT IN URBAN AREAS OF LATIN AMERICA: I. ORGANIZATION, POPULATION, AND METHODOLOGY OF THE GUATEMALAN PERINATAL STUDY¹

Edgar Kestler,² Joanne Dorgan,³
Ricardo Sibrián,³ Oscar Aquino,⁴ José Villar²

INTRODUCTION

A prospective epidemiologic study of pregnant women was conducted within a lower middle class population of Guatemala City residents. The study's aims were as follows:

- 1) to develop a statistically based risk score for identifying pregnant women at high risk of delivering low birth weight infants as early in gestation as possible;
- 2) to evaluate the sensitivity, specificity, and predictive value of this risk score, using a different sample of women from the same population; and
- 3) to develop a methodology for promoting use of the risk approach in prenatal care in other countries of the region.

This article provides detailed information about the study population,

organization of the study, methods and materials used, quality control of the data, and results of data reliability analyses.

THE STUDY POPULATION

Participants in the study were selected from women attending the prenatal care clinic at the Gynecology and Obstetrics Hospital of the Guatemalan Social Security Institute (IGSS). This is a 230-bed hospital where all deliveries by eligible women using the IGSS system take place. Women eligible to use the system include all women who are formally employed or whose husbands are formally employed.

¹ This article will also be published in Spanish in the *Boletín de la Oficina Sanitaria Panamericana*. Financial support for the work reported here was provided by the Board of Sciences and Technology for International Development (BOSTID), U.S. National Academy of Sciences (Grant No. REA-GT-4-84-2).

² Perinatal Nutrition Program, Division of Nutrition and Health, Institute of Nutrition of Central America and Panama (INCAP), Guatemala City, Guatemala; and Perinatal Research Program, Gynecology and Obstet-

rics Hospital, Guatemala Social Security Institute, Guatemala City.

³ Perinatal Nutrition Program, INCAP.

⁴ Perinatal Research Program, Guatemala Social Security Institute.

An average of 15,630 deliveries a year occurred at the IGSS hospital during the period 1979–1983. A central hospital clinic and peripheral clinics provide prenatal care to a large population that is referred to the hospital for delivery.

All pregnant women who made their first visit to the prenatal clinic at the IGSS hospital between 1 April 1984 and 10 January 1986, inclusive, were enrolled in the study. This group included 17,297 women. Women who delivered at the IGSS hospital but were not enrolled in the study included women who had received no prenatal care (about 9% of those delivering—1) and women who had received prenatal care outside the IGSS system or at the IGSS peripheral clinics (22% of those delivering). These women (some 31% of all those delivering at the IGSS hospital) were not enrolled because it was not feasible to implement standardized procedures and other data quality control measures in their cases.

A high percentage of the women eligible to receive IGSS services do deliver at the IGSS hospital. However, some (an estimated 3,000 per year—1) deliver at the two other general hospitals serving the public in Guatemala City; some others deliver at private hospitals; and some others deliver at home. Data are not available on the proportion of women eligible to receive IGSS services who deliver outside the IGSS hospital.

Table 1 shows demographic data for women delivering at the three public hospitals in Guatemala City, and also shows the percentages delivering infants with low birth weights. By and

TABLE 1. Demographic comparisons between women delivering babies at the three public hospitals in Guatemala City during 1983.

	% in indicated category or delivering a low birth weight infant at the:		
	IGSS hospital	University hospital	General hospital
% illiterate	18.8	18.8	17.0
% from homes lacking:			
electricity	8	20	17
piped water	23	34	34
sanitary facilities (plumbing)	24	36	38
Income (% <US\$100 per month)	11	26	27
% homemakers	70.8	86.2	92.6
% delivering infants with low birth weights (<2,500 g)	11.4	16.6	15.8

Source: Kestler, 1983 (1).

large (except for illiteracy) the data indicate that women attending the IGSS hospital tended to be in a better situation for delivering babies with adequate birth weights, and that the percentage of infants delivered with low birth weights was smaller at the IGSS facility than at either of the other two hospitals.

DATA COLLECTION

Patients were enrolled in the study when they registered at the IGSS perinatal clinic, following clinical or laboratory confirmation of pregnancy. Prenatal visits and interviews with social workers were scheduled to follow the hospital's routine procedures.

At the first prenatal visit, a social worker interviewed each woman to obtain sociodemographic data. These were recorded on a special social service

form developed by the hospital's social work service. Also, during each visit a prenatal form was completed by nurses or physicians providing patient care. Part of the hospital's perinatal record, this form is an expanded version of the perinatal record developed by the Latin American Perinatology Center (CLAP) (2) that has been adapted to local characteristics and the needs of the population served. These two forms, the social service form and the prenatal form, were the data sources used in this study.

After each prenatal visit, the patient's medical records were returned to the hospital's medical records department. At this time study staff members transferred information from these medical records onto precoded data collection forms especially developed for the study. (Copies of these forms are available upon request from the authors.) When missing data were detected, the record was flagged and attempts were made to obtain the missing information during the next prenatal visit. For variables that change over the course of pregnancy, such as blood pressure and uterine height, it was not possible to retrieve the missing data. Nevertheless, residents or obstetricians having incomplete prenatal records were easily identified and contacted by the project coordinator. This process not only minimized the number of missing values in the final data set, but also enabled in-service training to focus on standardization procedures and helped to identify physicians and nurses whose performance needed improvement.

At the time of delivery and before discharge the interviewers visited all mothers and collected any additional information required for the study. Then, before entry of the data into a computer, all the forms were visually reviewed by a special group of clerks for missing or invalid information. Any

problems detected were investigated at that time and corrected whenever possible, using the complete prenatal care record and direct contact with the patient.

DATA STANDARDIZATION AND CLEANING

To minimize extraneous variations, all measuring equipment (scales, tape measures, measuring boards, sphygmomanometers, etc.) was tested and calibrated before data collection started. Also, the results obtained by individual examiners were tested for comparability every two weeks.

The usual procedure for evaluating examiners involved two examiners (one of them the field director) who measured 10 pregnant women twice during the same prenatal clinic visit. Analysis of variance for repeated measurements (3) was performed to test for statistically significant variations and interactions. The standardization procedure was considered complete only when variations arising from all sources except the patient were found to lack statistical significance ($p > 0.05$).

In addition, measurement error was estimated by calculating the standard deviation of the error. Measurements were considered acceptable if the standard deviation of the error was less than or equal to values reported in the literature from similar studies (4).

Finally, using procedures suggested by Guzmán et al. (5), extensive statistical, logical, and computerized checks were performed to detect possible errors in the data gathered on each variable.

RELIABILITY TESTING

Standardized data collection procedures can be used to minimize the potential influence that variations in data collection methods employed by different personnel may have upon results. In the present instance, the data obtained by different "raters" were studied to determine how well these standardized procedures worked and to evaluate the reliability of the collected data.

The study of inter-rater agreement was divided into four sections defined by the types of variables investigated—these being sociodemographic, prenatal, labor and delivery, or laboratory variables. Information collected for the low birth weight study was used as one source of data for the inter-rater agreement study. Hence, these data were collected by hospital staff members applying usual study procedures. However, to test inter-rater agreement regarding prenatal, labor and delivery, and sociodemographic variables, the same data were also collected, respectively, by a specially trained physician, at a special prenatal care clinic, and by a social worker. Analyses of data on the last (laboratory) class of variables were performed at the INCAP laboratory and a private laboratory. All data collection procedures were performed separately on the same patient, and the results obtained by one rater were not available to another.

It was unclear whether the physician and social worker employed for the inter-rater agreement study gathered information more accurately than the regular study staff members. Similarly, it

was unclear whether the outside laboratories provided more accurate results than the IGSS laboratory. Therefore, the analyses that were performed considered all raters equally qualified.

The data were collected in two phases. The only difference between the two phases was the addition of a third laboratory in the second phase. Therefore, all the analyses were performed on a complete set of data.

Before starting these analyses, the data were checked for out of range values. When possible, invalid responses were checked and corrected; otherwise they were changed to "missing" values. Because the sample sizes were relatively small, the number of observations at different levels of categorical variables was often small. When reasonable, categories were collapsed to increase the size of these numbers and improve estimates.

The analysis of inter-rater agreement for categorical variables was performed using kappa statistics. Kappa is interpreted as a measure of the degree to which agreement appears above or below what would be expected by chance. In the case of two ratings per subject, the statistic varies from -1 to $+1$. Negative values of kappa indicate less than chance agreement, zero indicates chance agreement, and positive values indicate better than chance agreement. Perfect agreement is indicated by a kappa value equal to one.

Interpretation of the actual values of kappa is somewhat subjective. However, published guidelines suggest the following interpretation: Kappas greater than $+0.75$ represent excellent agreement beyond chance; kappas between $+0.40$ and $+0.75$ represent fair to good agreement beyond chance; and positive kappas below $+0.40$ represent poor agreement beyond chance (6). Based on these guidelines, an interpreta-

tion of the kappa values obtained for selected variables is shown in Table 2.

For continuous variables, the intraclass correlation coefficient was used as a measure of inter-rater agreement. This coefficient is the combined estimate of the correlation between different raters' measurements of the same subject. In the case of two raters, the intraclass correlation coefficient varies from -1 to +1. A value of zero indicates no agreement. A higher positive value signifies better agreement between the raters. In general, the intraclass correlation coefficient can be interpreted similarly to kappa, except for a factor involving $1/n$; so the intraclass correlation coefficient is equivalent to a weighted kappa (6).

Interpretations of intraclass correlation coefficients for selected variables, based on the same guidelines described previously for kappa values, are shown in Table 3.

This information (see Tables 2 and 3) can be used in deciding whether certain variables need further analysis,

and whether additional training of interviewers and health professionals or better standardization of their work is needed. On the basis of our experience, it is strongly recommended that periodic inter-rater agreement assessments be performed in the course of all epidemiologic studies.

DISCUSSION

We have described a very detailed data collection and quality control procedure. Given the size of the study population and the number of observers, we think that this procedure provided data of acceptable quality.

We have also described standardization procedures and reliability analyses including "inter-rater agreement" studies. Our work to date indicates that some or all of these two analytical processes, as described here and in the literature (3-7), can and should be implemented at all levels of prenatal care. Simplified computer programs for the most complex calculations, prepared for use by mid-level personnel, are available upon request. It is clear that all these steps must be monitored in any large-scale, multiple observer study.

TABLE 2. Kappa values for inter-rater agreement on certain dichotomized variables.

Variable ^a	Number of subjects	Kappa	Standard error	Rating
Marital status	140	0.94	.08	Excellent
Vitamins and minerals ^b	225	0.84	.07	Excellent
Analgesics	226	0.69	.07	Fair
Meconium during delivery	159	0.56	.08	Fair
Last child had low birth weight	143	0.35	.08	Poor
Edema during pregnancy	224	0.12	.07	Poor

^a Variables classified as yes or no except for marital status (married, unmarried partner, widowed, single).

^b Given in supplemental quantities prescribed in accord with hospital policy.

TABLE 3. Values of intraclass correlation coefficients for inter-rater agreement on certain continuous variables.

Variables	Number of subjects	Intraclass correlation coefficient	Confidence intervals	Rating
Gestational age at birth (weeks)	156	0.83	0.77, 0.87	Excellent
APGAR score at one minute	160	0.69	0.60, 0.76	Fair

However, it may not be necessary to implement longitudinal studies of risk factors in all health regions or countries. The biological association between risk factors and pregnancy outcome has been documented, and with some exceptions it appears to be relatively constant through populations. Thus, quantification of a given risk factor (determining its prevalence in a particular population) becomes the most important epidemiologic tool for planning and organizing perinatal health interventions. It should be noted that data standardization and quality control processes similar to those reported here also play a key role in these studies of risk factor prevalence.

Another question is whether all this methodology should be applied to routinely collected data. Large amounts of data are collected by perinatal services, most of which are only rarely used or considered when medical or public health decisions are taken (8). However, it seems clear that high data quality is needed in order for collected data to provide a sound basis for medical or public health decisions. And it seems equally clear that the quality control process recommended here is a nonroutine activity that could prove too expensive for many developing country institutions.

For all these reasons, we suggest that perinatal services should minimize the amount of information rou-

tinely collected and should apply the quality control and standardization process described here to all cases or a sample thereof. This will ensure that when collected data are used in the health decision-making process, those data will have an acceptable degree of reliability. A precoded data collection form can be used, and data can be entered into a computer at the local level, as recently suggested (9), permitting simple and rapid analysis.

This local level data analysis and interpretation should improve both data quality and data use, because it would permit local personnel to utilize the collected data directly and to verify the benefits derived from their collection efforts.

An alternative process, often used in the hospitals of developed countries, calls for extracting data from a complete medical record that deal only with those variables of interest and maintaining a quality control process only for these variables. We feel this alternative is not useful for large hospitals in Latin America. Implemented in our study, it required establishment of a new structure parallel to the hospital's bureaucracy, entailed excessive costs, and bur-

dened nurses and physicians without freeing them from responsibility for large-scale data collection.

Overall, the experience gained from our large perinatal project in Guatemala suggests a need for continuous data quality control, standardization of obstetric and neonatal procedures, limitation of the data collected to routinely gathered information about a group of important variables, and organization of local primary data analysis systems. Implementing all this will be difficult; but we believe that such action provides the only means of ensuring that the information collected will provide a sound basis for improving maternal and infant health care systems.

SUMMARY

A prospective epidemiologic survey of pregnant women performed in Guatemala City in 1984–1986 sought to develop a "risk score" for identifying mothers at high risk of delivering low birth weight infants as early in gestation as possible. This article provides detailed information about the study population and steps taken to ensure that the data gathered were precise and reliable.

All pregnant women making their first visit to the prenatal clinic at the Gynecology and Obstetrics Hospital of the Guatemalan Social Security Institute (IGSS) between 1 April 1984 and 10 January 1986 were included in the study. During this visit a social worker interviewed the patient to obtain sociodemographic information, which was recorded on a special social service form. In addition, during this and all subsequent visits a prenatal form was completed by nurses or physicians attending the patient. These two forms, supplemented by interviews at the time of delivery and be-

fore discharge, were the main data sources employed in the survey.

These forms were reviewed for completeness soon after each visit, and efforts were made to obtain any missing information. Also, to minimize extraneous variations, all measuring equipment was tested and calibrated before the survey started, and the results obtained by individual examiners were tested for comparability every two weeks. In addition, extensive checks were performed to detect possible logical errors or inconsistencies in the collected data.

Finally, to test the reliability of these data, limited numbers of patient examinations and laboratory tests were done outside the IGSS hospital. The resulting data were then tested statistically for agreement with data collected routinely from the same patient by hospital staff members.

In general, the authors feel that the data quality control processes described should be performed at all levels of prenatal care, for it seems clear that high-quality data are needed to provide a sound basis for medical and public health decisions. On the other hand, these quality control processes are not now routinely performed and could prove too expensive for many developing country institutions. Therefore, the authors suggest that perinatal services should minimize the number of variables about which data are collected routinely and should apply the recommended procedures to this smaller body of data.

REFERENCES

- 1 Kestler, E. Cuán distinta es la población que asiste a las principales maternidades de la ciudad capital de Guatemala? Influencia sobre bajo peso al nacer? *Bol Med Instituto Guatemalteco de Seguridad Social* 1(1):20-28, 1983.
- 2 Belizán, J. M., A. G. Díaz, H. Gioconmini, H. Horcher, M. Martell, P. Quaranta, and R. Schwarcz. *Historia clínica perinatal: Propuesta de un modelo*. Ministerio de Bienestar Social, Secretaría de Estado de Salud Pública, Dirección Nacional de Maternidad e Infancia, Centro Latinoamericano de Administración Médica, Buenos Aires, 1976.
- 3 Kirk, R. E. *Experimental Design: Procedures for the Behavioral Sciences*. Brook/Cole; Belmont, California, 1968.
- 4 Martorell, R., J.-P. Habicht, C. Yarbrough, G. Guzmán, and R. E. Klein. The identification and evaluation of measurement variability in the anthropometry of preschool children. *Am J Phys Anthropol* 43:347-352, 1975.
- 5 Guzmán, M. A., R. Sibrián, and R. Flores. Procedimientos básicos en el registro de proceso de datos. *Arch Latinoam Nutr* 33:257-267, 1983.
- 6 Fleiss, J. L. *Statistical Methods for Rates and Proportions*. John Wiley and Sons, New York, 1981, 223 p.
- 7 Snedecor, G. W., and W. G. Cochran. *Statistical Methods (sixth edition)*. Iowa State University Press; Ames, Iowa, 1967.
- 8 Hail, M., and P. K. Chng. Chapter 3: Antenatal Care in Practice. In: E. Murray and I. Chalmers. *Effectiveness and Satisfaction in Antenatal Care (first edition)*. William Heinemann Medical Books, London, 1982.
- 9 Schwarcz, R., A. G. Díaz, R. H. Fescina, J. L. Díaz Rossello, M. Martell, and S. M. Tenzer. Historia clínica perinatal simplificada. *Bol Of Sanit Panam* 95:163-172, 1983.