CHAPTER V

MATERIALS AND METHODS
CHAPTER V

MATERIALS AND METHODS

5.1 Design:

Ideally, the risk of spontaneous abortion should be estimated following a cohort of women prospectively, from the time of conception until the end of pregnancy. In practice, however, estimation of the risk of spontaneous abortion faces many problems. Methodological problems involved in diagnosing an early pregnancy by means of measuring the concentration of HCG are substantial. A few problems are listed below:

- Some very early pregnancy losses are not identified by the woman or her doctor.

- Some conditions presenting as spontaneously aborted pregnancies are not true pregnancies. Some of the conditions may be due to menstrual irregularity, false positive pregnancy test, false clinical signs of pregnancy, or trophoblastic diseases.

- Not all women consult a doctor immediately after recognizing that they are pregnant, so some spontaneous abortions may occur even before the time of first prenatal visit and hence would never be registered.

- Some pregnancies are terminated deliberately, either because the pregnancy is not viable (viz. missed abortion, ectopic pregnancy or blighted ovum) or because of the woman's wish.
• Case-Control studies provide is an established and acceptable approach to study the risk factors. Although, susceptible to biases for its retrospective nature, it is quicker and economical approach as compared to Cohort Study.

Taking into consideration, the above methodological problems we performed this study in two phases:

Phase I: A hospital based unmatched case-control study for identification of the risk factors for spontaneous abortion; development and back validation of the devised risk scoring system.

Phase II. A cohort study for prospective validation of the scoring system

5.2 Setting:

The present study was carried out at Obstetrics and Gynaecology Department of Government Medical College Hospital, Nagpur a secondary and tertiary care referral center in Central India.

5.3 STUDY DESIGN -

5.3.1 Phase I- Case-Control design

An unmatched case-control study was conducted to identify the risk factors for spontaneous abortion, development of a risk scoring system and subsequent validation on the same data (back validation).
5.3.1a Cases:

The *cases* of spontaneous abortions were identified by trained female research associates (RAs) as the women admitted to the study hospital with bleeding during the pregnancy or the fetal loss up to 20 weeks of gestation.

The crucial methodological issue involved in this study was to ensure identification and recruitment of all, or as many cases of spontaneous abortion occurring in the study hospital as possible. Two trained Research Associates (RAs) used to visit the Family Planning Clinic, Abortion Ward and Operation Theatre of Obstetrics and Gynecology Department daily to track down the consecutive cases of spontaneous abortion. After the verification from the records and after confirming their abortion status from the Bed Head Tickets and the concerned Unit In-charge, the RAs then recruited the individual case in the study. The standard case definition was followed while recruiting the cases of spontaneous abortions.

**Definition of Spontaneous Abortion:** Spontaneous abortion is defined as a birth process terminating before the 20th week (or 139 days) of gestation. It implies the expulsion of all or any part of the placenta or membranes, with or without an identifiable fetus or a live born or still born infant weighing < 500 gm<sup>6</sup>.

5.3.1b Controls:

From the same ward to which the identified case belongs, one age and parity matched control per case was selected from the same hospital. The *controls* were normal delivered woman (with no history of bleeding during the pregnancy till 20th week of gestation) during the same period. Controls were selected by frequency matching with the cases on two important baseline characteristics, namely Age (± 2 years) and Parity (± 1).
5.3.1c Sample size for case control study

The sample size for the case-control study was calculated on the basis of anticipated moderate risk estimate (OR=2), and expected prevalence of exposure in the control population (p0=0.10). Thus, the minimum required number for this phase was 402 cases and 402 controls, for a specified level of significance (α=0.05) and power (0.90).

5.3.1d Back validation

Statistical methods like Multiple Logistic Regression technique was used to find the adjusted odds ratio for the statistically significant risk factors. On the basis of its regression coefficient value, each risk factor was assigned a statistical weight. A total score for each of the individual woman subject was obtained. The devised model was then validated on the same data set from which it was developed by calculating Sensitivity, Specificity, Positive predictive value, Negative predictive value, Agreement measures at various cut-offs of the total score. Then the best cut-off score was used to statistically predict high- and low- risk cases of spontaneous abortion among the clinically diagnosed cases and controls.

5.3.2. Phase II- Cohort design

A separate cohort study was carried out for prospective validation of the developed risk scoring system in our setting.

5.3.2a Sample size for cohort study

The sample size for the second phase (cohort design) was calculated on the basis of an estimated predictive accuracy of the proposed risk scoring system in the first phase. A cohort of 440 women was considered sufficient to attain a 90% power at 5% level of significance if the tests were supposed to have predictive accuracy between 83% and 90%.
5.3.2b Cohort recruitment

Adjusting the sample size for an anticipated 15% non-response rate, a cumulative cohort of 510 pregnant women registered at Ante Natal Care (ANC) clinic of the study hospital was recruited, and prospectively followed over till the outcome of pregnancy i.e. delivery or abortion.

The same diagnostic criteria and procedures, used for assessment of the risk factors in the earlier case-control part of the study, were followed.

5.3.2c Prospective validation

At the recruitment stage of the study subjects in the cohort, a total score was calculated for an individual pregnant woman after the assessment of the risk factors under the devised risk scoring system. The outcome data was available after delivery or abortion.

The same analytical methods used in phase I for back-validation of the devised prediction model in phase I were also employed for the purpose of prospective validation of cohort data in the phase II. The validation was performed by estimating the sensitivity, specificity, predictive accuracy, likelihood ratios, and Cohen’s kappa, by using the proposed cut-point of the devised prediction model in the first phase.

5.4 Study Duration and Time line

The total study duration for both the phases was approx. 30 months. Data collection for the case-control phase of the study began in March 2001. The required sample size i.e. 402 cases and 402 controls was achieved by December 2001. Back validation was done in January 2002. Data Collection for the second phase began in January 2002 and continued till December 2002 wherein a cumulative cohort of a total 510 pregnant women was recruited and followed up from the date of registration till the outcome i.e. delivery or abortion. Prospective validation was done in January 2003. The next 9 months were devoted to final data editing, data analysis, report writing and preparation of the manuscript.
5.5 Data Collection:

The detailed information was collected in a pre-designed, pilot-tested, structured interview schedule by the trained interviewers (RAs) who were doctors and social scientists. To minimize the interviewer’s bias, the length of the interview for both cases and controls was recorded and compared. The study instrument was pilot-tested on subgroups of cases and controls for assessing the feasibility and validity of the questionnaire.

5.6 Study Instrument/Questionnaire:

The questionnaire was developed in relevance of study objectives and information to be collected on risk factors for spontaneous abortion. Questions assessing the risk factors for spontaneous abortion, in the context of study objectives were framed. The questionnaire mainly included socio-demographic characteristics and hypothesized risk factors.

Following factors are included in the study.

Socio-demographic Factors

- Age (Both husband and wife)
- Area of Residence
- Religion
- Socio-economic Status (SES)
- Family Size
- Education (Both husband and wife)
- Occupation (Both husband and wife)
- Per-capita Income

Maternal Factors

- Place of Delivery
- Age at Marriage
- Age at Menarche
- Obesity (Body Mass Index)
- Parity
- Gravidity
- Gestation in Weeks
Index Pregnancy
- Inter Pregnancy Interval
- Use of Contraceptives
- Pelvic Inflammatory Disease (PID)
- Menstrual History
- Previous History of Spontaneous or Induced Abortions
- Frequency of Previous Abortions
- Consanguineous Marriage
- Drug Use / Medication During Pregnancy

Life Style Factors
- Tobacco Use (Chewing / Smoking)
- Excessive Coffee / Tea Drinking
- Alcohol Consumption

Systemic Diseases
- Diabetes Mellitus
- Chronic Cerebro-vascular Disease
- Chronic Renal Disease
- Hypertension
- Tuberculosis

Environmental Factors
- Radiation exposures
- Exposure to Toxic Substances

Work-related Factors (Both for employed women and housewives)
- Current Employment
- Standing for more than two hours at work
- Walking for more than two hours at work
- Sitting for more than two hours at work
- Vibrations at work
- Commute to work
- Reaching over the shoulders at the job
- Carrying loads over 9 kg on the job
- Night shifts at work
- Care of young children (< 5 years old)
- Full-time household work (as a housewife)
- Inadequate Rest

Other Factors
- Physical Trauma
- Other Gynecological Problems
- Abuse During Pregnancy
5.7 Statistical Analysis:

Subsequent to the data collection, code-book was devised for the purpose of data entry. Each question in the study instrument was analysed for possible options (based on responses) and based on this codes were developed. Data were then entered in Microsoft Excel sheet, and analysis was carried out in STATA (7.0, 2001) statistical software. In the first step, descriptive analysis was carried out. Subsequently, in bivariate analysis, Pearson’s chi-square, crude Odds Ratios and 95% Confidence Intervals, and if necessary Fisher’s Exact Test and Exact Confidence Intervals were calculated for small frequency.

As the outcome of interest (Spontaneous abortion) was a dichotomous variable, and we wanted to study its relationship with an exposure variable and a set of covariates — some of which were measured on different measurement scales — Logistic Regression approach was used for model building purpose. The logistic regression model is a remarkably flexible model. Another advantage in using logistic regression approach was the well known fact that analysis of data from case-control studies via logistic regression may proceed in the same way and using same computer programs as cohort studies. (as per Hosmer and Lameshow on “Sampling models” in their text book on “Applied Logistic Regression”, 3rd edition, 1989).

Unconditional Multiple Logistic Regression Analysis was carried out to estimate adjusted Odds Ratios. Full Model of logistic regression included risk factors, which were significant on bivariate analyses at 0.2 level of significance ($\alpha$). Subsequently, the factors which were significant at $\alpha = 0.2$ in the full model were included in final model I. Final model II or the reduced model included risk factors significant in the final model I at 0.05 level of significance. Attributable Risk Proportion (ARP) and Population Attributable Risk Proportion (PARP) and their 95% CI were calculated on the basis of the adjusted odds ratio estimates obtained from the final (reduced) model for significant risk factors.

A prediction model was developed to identify the women who were at risk of spontaneous abortion. Each of the risk factors retaining significance in the final model
of the logistic regression analysis was assigned an appropriate statistical weight. The weight for a particular factor was calculated using a suitable linear transform on the respective regression coefficient ($\beta$) of that variable in the final model. The statistical weight was calculated by rounding the linear transform to the nearest integer. Such a transform was necessary to make the risk scoring system easy to use in actual clinical practice. Thus, the total score for an individual subject was obtained by adding the corresponding statistical weights for the factors present in the individual woman subject. The devised prediction model was then back-validated on the same data set from which the system was developed. All the study subjects were scored individually using the devised risk scoring system. The sensitivity, specificity, positive and negative predictive values, likelihood ratio, percent agreement, and Cohen’s kappa statistics were calculated at various cut-offs of the total risk score. The overall predictive accuracy of the prediction model was calculated as an equivalent of the area under Receiver Operating Characteristic (ROC) curve. The best cut off for the total score was obtained graphically by plotting the ROC curve. Predicted probability for an individual woman subject was obtained by applying the prediction rule for presence of the identified risk factors. The capability of the model to correctly classify the identified women at risk of abortion was also ascertained. The utility of the prediction model for various sub-groups (e.g. age) was also tested.

The same analytical methods used in phase I for back-validation of the devised prediction model in phase I were also employed for the purpose of prospective validation of cohort data in the phase II. The prospective validation was performed by estimating the sensitivity, specificity, predictive accuracy, likelihood ratios, and Cohen’s kappa, by using the proposed cut-point of the devised prediction model in the first phase. Area under ROC Curve and predicted probabilities were also estimated, and classification capability of the model was assessed. Crude Analysis of cohort study was also done by calculating estimates of relative risks (RR) and 95% Confidence Intervals for important modifiable risk factors.