Abstract
It is important to understand that any epidemiologic research study can only test some part of the observed association given that broader sets of factors (social, economic, environmental etc.) acting in the actual causal mechanism are largely unknown.[1] It is very important to understand concepts and principles that govern epidemiology to arrive at logical inferences.[1, 2] In developed countries, modern practice of Obstetrics and Gynecology (OBG) involves the integration of highly specialized disciplines such as reproductive epidemiology.[3] Although most of the developing countries are passing through a phase witnessing laudable achievements in upholding clinical skills in the field of OBG, conducting research studies, updating knowledge and implementing sound epidemiological skills still remained grey areas.[4, 5]

INTRODUCTION
"Not everything that can be counted counts, and not everything that counts can be counted." - Albert Einstein

It is important to understand that any epidemiologic research study can only test some part of the observed association given that broader sets of factors (social, economic, environmental etc.) acting in the actual causal mechanism are largely unknown.[1] It is very important to understand concepts and principles that govern epidemiology to arrive at logical inferences.[1, 2] In developed countries, modern practice of Obstetrics and Gynecology (OBG) involves the integration of highly specialized disciplines such as reproductive epidemiology.[3] Although most of the developing countries are passing through a phase witnessing laudable achievements in upholding clinical skills in the field of OBG, conducting research studies, updating knowledge and implementing sound epidemiological skills still remained grey areas.[4, 5]

A highly specialized discipline termed as reproductive epidemiology, distinct from other areas of epidemiology has emerged with highly critical outlook at range of issues addressing reproductive life. For detailed account of these epidemiological principles, readers are advised to refer to chapter on Reproductive Epidemiology by Jørn Olsen and Olga Basso.[3] The aim of this paper is to focus on introducing modern methods governing case control studies, intended to be part of series that covers range of topics for application of modern epidemiological principles in developing countries.

Regardless of the demanding need for fundamental data such as Low Birth Weight (LBW), it might be an impossible task in developing countries like India to estimate incidence of congenital malformations or low birth weight (LBW).[3] This is because we do not have the infrastructures that can generate such data validly. To estimate incidence of LBW, for example, we need to identify all the women who have conceived in a given area and follow them up till they deliver babies including measuring the early abortions which might otherwise be missed.[3] This might be too much to ask in a country like India wherein; there is difficulty for registering pregnancies for routine antenatal care. Hence conducting case control studies might only be the best option in developing countries. It is important that modern epidemiological methods are incorporated to enrich the body of evidence while doing any studies.[2]

RETROSPECTIVE OR PROSPECTIVE
Case control studies are routinely employed in the discipline of Obstetric research in India. However, case control studies need not be only retrospective as wrongly taught even currently. The terms retrospective and prospective are used only from investigator’s point of view with reference to direction of data collection. Readers would find several published studies with designs of prospective case control design (as done with density sampling) [6-9] as well as
INCIDENCE MEASURES OF ASSOCIATION FROM CASE CONTROL STUDIES

As an obstetrician, estimation of odds ratios may not be useful in communicating to patients and hence may prefer Risk Ratio (RR) and Incidence Rate Ratio (IRR). Case control studies are employed for very rare disease but, this by no way means that the odds ratio in case of rare diseases will automatically approximate to risk ratio much against conventional teaching in India. Odds ratios can approximate to risk or rate ratio only when some variants of case control studies are done,[14] which employ different sampling designs. Rate ratios are calculated when sampling of controls is done from “risk set”, wherein the subjects in the cohort who are at risk at the time of diagnosis of each case. Similarly, risk ratios are calculated as approximation of odds ratios when selection of controls is done such that exposure distribution among them will estimate without bias the exposure distribution in source population (Table-3-A).[14]

ANALOGY USED IN THIS PAPER

To understand better, it would be better to visualize a cohort of 1100 women who have conceived exactly on 1st January 2010 in a specific urban area named say ward.no.91, in some part of India. Let us also imagine that a hospital X caters to this specific area and provides preventive and curative services to this urban area and hence provides antenatal care to all these pregnant women and many pregnant women from other areas also attend this hospital. Out of 1100 conceptions, 50 conceptions were lost because of spontaneous abortion. Among other pregnant women (1050), who carried on with pregnancy, 25 women had abortion during first trimester and 25 women delivered dead born child at completion of full term. You as a researcher want to study the association of a drug called E (E for Exposure) and low Birth weight among live born children (D for Disease).

At the completion of full term, 900 babies born have normal weight and 100 of them are low birth weight babies. (Figure.1) (Table.1) We also assume that among the 100 women who gave birth to babies with low birth weight, 90 were exposed to drug E for a certain amount of time around 1st January 2010 and other 10 women were not.

<table>
<thead>
<tr>
<th>Exposure</th>
<th>No. of cases</th>
<th>Non Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>90</td>
<td>500</td>
<td>590</td>
</tr>
<tr>
<td>Absent</td>
<td>10</td>
<td>500</td>
<td>510</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>1000</td>
<td>1100</td>
</tr>
</tbody>
</table>

APPROACH TO DESIGN CASE CONTROL STUDY

If a cohort study were done from baseline (1st January 2010), we would then get the estimates as depicted in Table.2. It can be noted that risk ratio and odds ratio are same when we do a cohort study including exposure experiences of all the people in defined source population. Ideally, to be valid, any Case control study done to study the association between exposure (E) and low birth weight (D) should produce odds ratio similar to table no.2. However, this rarely happens due to incorrect source population, incorrect choice of controls, incorrect methods, and incorrect
statistical procedures and lack of enabling systems for research. We deal with these problems in subsequent steps.

**Table 2**

Measures of association of source population

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Numbers in formula</th>
<th>Value of Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk among exposed</td>
<td>Disease among exposed</td>
<td>90/150</td>
<td>0.66</td>
</tr>
<tr>
<td>Risk among unexposed</td>
<td>Disease among unexposed</td>
<td>10/150</td>
<td>0.07</td>
</tr>
<tr>
<td>Odds of (Disease) among exposed</td>
<td>Disease positive-exposed</td>
<td>90/150</td>
<td>0.58</td>
</tr>
<tr>
<td>Odds of (Disease) among unexposed</td>
<td>Disease negative-unexposed</td>
<td>10/150</td>
<td>0.07</td>
</tr>
<tr>
<td>Risk Ratio</td>
<td>Risk among exposed/Risk among unexposed</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>Odds among exposed/Odds among unexposed</td>
<td>5.8</td>
<td></td>
</tr>
</tbody>
</table>

**IDENTIFYING THE SOURCE POPULATION**

Source Population is the population that gives rise to cases, if they develop disease (D) within that population (ward no. 91 in our example) would end up in the hospital X. One has to note that the source population for an outcome of pregnancy such as low birth weight is the cohort of all conceptions in a given area (1100 in our example) and not the number of pregnant women who carried their live fetuses until full term.

In practice, hospitals in India do not cater to a defined area from where the cases can potentially come from and the catchment area spreads across different districts, states and even countries. Identification of source population is important first step in case control studies because it is this population, which will provide ideal controls from, and also to which the inferences from study will apply to. In developing countries such as India, we can only obtain data on some pregnant women (those who attend antenatal care as prescribed), on some deliveries (that are carried out in the same hospital) and on few abortions (and usually only on abortions of clinically recognized pregnancies, thus missing the early ones). Hence most of the processes leading to these events are hidden or altogether absent from the data we possess while doing case control studies. 3

**IDENTIFICATION OF VALID CONTROLS**

Controls should be selected from the same source population that gives rise to the cases. Further, controls should be selected independently of their exposure status; that is, drug E in our example should neither cause nor prevent the disease suffered by controls.

**Table 3a**

Selection of controls: Case-non Case method from source population

<table>
<thead>
<tr>
<th>Exposure</th>
<th>No. of cases</th>
<th>Non cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>90</td>
<td>51</td>
<td>141</td>
</tr>
<tr>
<td>Absent</td>
<td>10</td>
<td>49</td>
<td>59</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>(90 X 49) / (10 X 51)= 8.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3b**

Selection of controls: Case-non Case method from hospital control

<table>
<thead>
<tr>
<th>Exposure</th>
<th>No. of cases</th>
<th>Non Cases-1</th>
<th>Non-Cases-2</th>
<th>Total-1</th>
<th>Total-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>90</td>
<td>75</td>
<td>2</td>
<td>165</td>
<td>92</td>
</tr>
<tr>
<td>Absent</td>
<td>10</td>
<td>25</td>
<td>98</td>
<td>35</td>
<td>108</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Odds Ratio1</td>
<td>(90 X 25) / (75 X 10)= 3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds Ratio2</td>
<td>(90 X 98) / (10 X 2)= 10.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As seen in Table.3-A, the simplest way to ensure proper selection of controls is to select random sample of individuals from “source population” that gave rise to cases.[14] However, most of the studies done in India choose controls by selecting among patients from the hospital. In such cases, one can note that (as in Table.3-B), the estimation can be invalid and can be either underestimated or overestimated. There are several other problems such as Berksonian selection bias, confounding by indication induced by selection of hospital controls. For a detailed discussion, readers are strongly advised to refer to literature on this topic.[14-17]

**OTHER CHALLENGES INCLUDING LACK OF ENABLING SYSTEMS FOR RESEARCH**

In India, we do not have any established mechanism to monitor effects of drugs, agents or some chemicals taken during pregnancy and hence do not have any data to estimate correct prevalence of diseases resulted to these agents at birth. Let us say that there exposures which may simultaneously increase the incidence of LBW while at the same time reduces the survival rate of the affected fetuses (or embryos) in utero, it would lead to decreased prevalence of LBW at birth. Missing early losses such as spontaneous abortions can lead to severe bias and underestimation of congenital malformations and LBW’s at birth.
CONCLUSION

We have not discussed several methodological issues unique to reproductive epidemiology such as adjusting repeating tendencies of reproductive failures (such as spontaneous abortions, preterm delivery, preeclampsia) or how to account for time stable causes (eg, genes or part of the maternal environment are present during all events). We suggest readers to update knowledge on epidemiological methods by referring to detailed discussions such as by chapter by Jørn Olsen and Olga Basso(2005).[3]

References

Author Information

Giridhara R Babu, MBBS, MBA, MPH, PhD. Associate Professor
Public Health Foundation of India, IIPH-H, Bangalore campus, SIHFW premises, Beside leprosy hospital, 1st cross; Department of Epidemiology, UCLA Jonathan and Karin Fielding School of Public Health, University of California Los Angeles
Bangalore, India; Los Angeles, California, USA
giridhar@iiphh.org

Tanmay Mahapatra
Department of Epidemiology, UCLA Jonathan and Karin Fielding School of Public Health, University of California Los Angeles
Los Angeles, California, USA

Sanchita Mahapatra
Department of Epidemiology, UCLA Jonathan and Karin Fielding School of Public Health, University of California Los Angeles
Los Angeles, California, USA

Sanjay Kumar, MBBS, DNB (Obstetrics & Gynecology), Associate Professor
Department of OBG, Melaka Manipal Medical college, Karkala, Kasturba Medical College Manipal, India

Mohammad Akhtar Hussain, 4. MBBS, MD, Assistant Professor
Public Health Foundation of India, IIPH-Bhubaneswar
Bhubaneswar, India