Options in making use of pregnancy history in planning and analysing studies of reproductive failure

Jørn Olsen

Abstract

Background—Although more epidemiological studies of reproductive health have been published recently, important scientific achievements remain few. Among the reasons for this are existing methodological constraints. Fertility is low in most parts of the industrialised world, making the observation time short. Pregnancy itself is under extremely close medical surveillance with intensive health care interventions such as prenatal screening, induced abortions, induced labour, and prescribed changes in environmental exposures, but the epidemiologists are generally left with cause-effect experiences which escaped medical attention or study of diseases for which there is no cure. The main problem in non-experimental research of reproductive failure, however, is that most pregnancies are planned and past pregnancy experience is used in the planning. Self selection into different exposure categories according to past pregnancy experience should therefore be considered when studies are designed and analysed.

Study objective—To consider how to deal with past pregnancy experience as an indicator “background” risk of reproductive failure. The present standard is to take parity or gravidity into consideration but this may be of little value and could even be very misleading.

Main results—No simple and universal analytic strategy is available in reproductive epidemiology. On some occasions studies should be restricted to the first pregnancy only, in other situations the past reproductive experience may be used to form groups that are comparable in terms of pre-exposure risk.

Some reproductive failures are common and they often have major implications for the child, the family, and for society. Sometimes reproductive failure has preventable causes related to the parent’s lifestyle (smoking and drinking habits, for example), occupational hazards, or other environmental exposures. These exposures can be eliminated or reduced for the population at risk if they are identified, and much of the effort in reproductive health epidemiology has been devoted to identifying such risk factors. Except for some noteworthy exceptions, however, the results have been rather meagre. There are several reasons for this, some of which are related to difficulties in designing and analysing these studies. A short description of some of the methodological problems was published recently. Only problems related to the past pregnancy experience will be addressed in this paper.

Study results often state that the analyses were adjusted for parity, as if parity were an important confounder. It is unclear, however, which risk factors parity could be a proxy variable for, except for fertility, and (perhaps) fecundity. Even in this case, however, parity might be a poor proxy variable.

It is true that parity is associated (in the statistical a sense) with the frequency of most, if not all, reproductive failures but this need not indicate a causal association. If the frequency of reproductive failure (like perinatal mortality) is recorded over time for all women in a given region, the lowest frequency is usually seen in the second pregnancy and increasing failures are seen in subsequent pregnancies. However, if the frequency of perinatal mortality as a function of parity at the end of the reproductive age is examined, it can be seen that perinatal mortality declines with parity and that women with a larger final parity generally have higher perinatal mortality according to birth order compared with women who end up in a lower parity group.2–5

Any of these associations may be entirely the result of forces of self selection, and the latter way of presenting the data is no better than the former. Mantel6 has clearly illustrated how force of selection, in itself, may produce a declining perinatal mortality by birth order. A constant perinatal mortality (say of 20 per 1000 newborns) would, for example, produce a declining perinatal mortality by birth order if all losses are replaced by a new pregnancy but only 50% of those who gave birth to a child that survived try again. We would then have the results seen in a cohort of 10 000 pregnant women.

The example in the table is extreme but illustrates clearly the problem that probably operates in a less extreme form in real data. In short, most empirical data show an association between parity and reproductive failure. This is expected in populations in which planned pregnancy is the norm, and it need not reflect any causal relationship. Adjusting for parity need not lead to comparable groups in terms of similar a priori risk of reproductive failures.

If one believes that the causal link between parity and reproductive failures is correctly
stated in strata defined by the final parity (after the end of the reproductive "career"), one should wait until the reproductive career has finished and then adjust according to final sibship size and pregnancy order. Fortunately, that is not necessary. Any associations could in fact be artifacts related to selective pregnancy planning according to a desire for a given family size and the results of previous pregnancies.

It is well established that past experience does play a role in pregnancy planning. A pregnancy failure is often followed by a new pregnancy. Whereas a pregnancy that ends in the birth of a child with a congenital malformation may delay further childbearing. Even the gender of the children play a role in further family planning.

Pregnancy history and confounding
Past experience is, however, an indicator of the couple's risk of reproductive failure through genetic factors or exposures in past pregnancies that had a long lasting effect. Past experience could indeed produce valuable information as a proxy variable for risk factors which should be controlled for in order to obtain good internal validity. Since we still know little of true potential confounders (true causes of reproductive failure) we need to make use of as good proxy variables as possible for these confounders. We aim at obtaining an unexposed cohort with the same disease risk as the exposed cohort had they not been exposed; or to make sure that both cohorts have the same a priori risk; or to end up with a similar frequency of reproductive failure in the two cohorts if the exposure has no effect. Should that not be the case for all in the selected samples we need to stratify the samples when analysing the data to obtain comparability within the strata. This stratification should take past pregnancy experience into consideration in a detailed manner. A simple stratification on parity or pregnancy order will not do. We would have to take the outcome and the sequence of the events into consideration, perhaps even including previous periods of subfecundity.

If we had to identify specific outcomes in studies performed late in the childbearing years, this could be rather complicated and lead to sparse strata in populations with large families. In some cases these problems may be avoided by making use of stratification according to a confounder score.

This suggested strategy has not been used much, perhaps because of Gladen's report in 1986 which illustrated how chance alone could produce a poor pregnancy history under a probabilistic model. So called "habitual abortion" - more than a given number of spontaneous abortions in a woman—would be expected to occur by chance alone for a number of women, in spite of a "normal" risk of abortion, due to bad luck. Even though this may be true for individuals, past pregnancy history in an indicator of the outcome in subsequent pregnancies at a group level. If an exposed cohort had a poorer pregnancy history before the exposure than the unexposed group, this could indicate confounding by susceptibility. At least for a large group it is unlikely that "chance" alone would account for significant differences.

<table>
<thead>
<tr>
<th>Pregnancy cohort</th>
<th>Live births</th>
<th>Deceased</th>
<th>Perinatal mortality by birth order given two pregnancies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>In 1 pregnancy</td>
</tr>
<tr>
<td>(1) 10 000</td>
<td>9800</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>(2) 4900</td>
<td>4802</td>
<td>98</td>
<td>39 per 1000</td>
</tr>
<tr>
<td></td>
<td>+ 219</td>
<td>+ 4</td>
<td>102</td>
</tr>
</tbody>
</table>

Limitations in adjusting for past pregnancy experience
Controlling for past pregnancy history, however, may not solve all the problems. A number of limitations is part of this way of using previous pregnancy histories as a way of estimating prior pregnancy risk unrelated to the exposure under study.

Firstly, the pregnancy history adjustment may overadjust and therefore mask an effect of the present risk. This would be the case if the exposure under study was also present in past pregnancies, which would be almost unavoidable if common habits like coffee consumption, diet, physical activity, etc are to be studied. Should these factors be causally linked to pregnancy outcomes, past exposures should also be considered when making the stratification or when modelling the confounder score. Data collection and data analyses become much more complicated. Fortunately, under the null hypothesis the problem does not exist, and in other situations masking of an effect is expected if the exposure remains unchanged over time, or is changed equally among the groups to be compared.

A problem which is even more troublesome relates to the fact that the past pregnancy experience could partly determine the present exposure situation. To some extent this will be the case when the couples are aware of the present exposure and consider the exposure potentially harmful to the pregnancy. It is probable that present lifestyle exposures are heavily influenced by past reproductive experience. It is to be hoped that drinking habits during pregnancy will be modified if the histories of past pregnancy gave reason for concern, and the same could be expected for some occupational exposures too. If past experience...
is used in planning subsequent pregnancies, and also in adjusting exposure settings, the exposed cohorts would be expected to have a lower a priori risk (as reflected by a successful pregnancy history in the past). High risk groups could be over-represented in the reference groups of those not exposed, if people with an unsuccessful pregnancy history take care to avoid hazardous exposures. One may argue then that the stratification according to past pregnancy history will adjust for that in a way, but there is still uncertainty about why some couples choose to modify exposures and others not. The decision process may well be influenced by other types of more subtle information and may be related to other potential risk factors. Self selection to different exposure levels is similar to the "confounding by indication" situation, which may be satisfactorily solved only in a randomised trial. Since randomised trials are usually prohibited in studies of causes of diseases the only acceptable solution is to restrict the study base to couples having their first pregnancy. By doing that no one has a past history to make use of when exposures are avoided or accepted. All will be equally ignorant on their background risk of reproductive failures. They have no warning signals to make use of in the planning.

If the exposure under study is unknown to the couple (like the exposure to selenium, pesticides, etc may be), or is not considered to be of any harm, like coffee consumption in some countries, the anticipated comparability should be better, given proper comparability on past reproductive histories. Adjustment for post reproductive experience is sound if the exposure of relevance for post reproductive failures has no impact on the pregnancies under study.

When studies are not restricted to first time pregnancies one should also pay attention to the probability of collecting the same information on childbearing performance in the groups to be compared, and this depends partly on the desire for a given family size and partly on other issues such as cultural differences in risk taking behaviour, etc. If, for some reason, the exposed and unexposed cohorts differ in their desire for a given number of children, their opportunities for modifying exposure based upon past performance will also differ, and the problem is not only related to an imbalance of pregnancy information in the study base. Even after stratification according to past pregnancy histories, the groups need not be comparable. Those with a strong desire for a large family will probably continue to try to reproduce even after miscarriages whereas those who are satisfied with having only one child would not make such efforts. If the compared cohorts do not, on average, have the same desire for a given family size, the couples with high risk of poor reproductive performance will be more easily identified than those who are satisfied with few or no children. They will not repeat unsuccessful attempts, and stratification upon past reproductive histories need not lead to comparable a priori risk subgroups.

The desired size of family should be considered in studies that follow up couples during their reproductive career, but this need not be considered in studies done after the reproductive career when past exposures are linked to past pregnancies. In this case stratification by pregnancy history takes account of desired family size as well, under very simplified conditions, as shown by Gladen.11

It should be noted that adjusting for past pregnancy by including the past pregnancy history in a multivariate analysis may cause problems since a constant effect measure across all strata is hardly to be expected. If, for example, we accept the idea of the "habitual aborters", then some women have a high constant risk of abortion as a result, for example, of genetic factors. These women would be of little value in a study of environmental exposures. If, in fact, they have an a priori risk of 1-0, they produce no information at all. The more pregnancy experience we observe, the more we know of the prior risk of reproductive failure. We would not, however, expect to get a homogenous measure of association in the different strata, since we are studying the effect of an external exposure working on a very different a priori risk. One may consider an a priori risk of, say, 10% for spontaneous abortions among women without prior pregnancies. Should the exposure increase that to 20% there would be a relative risk (RR) of 2-0. In the strata of four prior spontaneous abortions, the prior risk could be 50% regardless of the exposure, and an increase to 60% produces a RR of 1-2. Such an additive effect would be expected if the environmental and the biological risk factors operated in different causal fields according to a deterministic causal model. The more strongly the prediction of poor reproducers is based on past reproductive history, the less informative one expects the subgroups to be. The opposite is true for a past history of several pregnancies without problems. The relative risk is then expected to be high if the exposure causes harm, due to the low prior risk among those not exposed.

Conclusions

Proper adjustment for past pregnancy history goes much further than just adjustment for parity and pregnancy order, and parity alone could be a very poor proxy of reproductive performance.

(a) When the exposures under study are considered potentially harmful to reproduction, it may be necessary to restrict the study base to first time pregnancies only. This would also obviate the statistical problem of lack of independence of two or more pregnancies.

(b) If past experience plays no role in the present exposure situation and the exposure was not present in the past, pregnancy history should be considered in the analyses, perhaps by making use of detailed stratification. In addition, one should design the study base in such a way that the compared cohorts are comparable in their desires for a given family size. Usually, one has to use proxy variables in
doing so, such as marital status, income, working hours, housing conditions, religious beliefs, etc.

If groups are compared with a different pregnancy experience, one should always use the subset of first time pregnancies to estimate the exposure-effect association in a separate analysis. Should that measure differ much from the overall measure bias related to forces of selection may well be the explanation.

The study was supported by the EEC (BIOMED: The European Study of Infertility and Subfecundity).


