

Reproductive Epidemiology

III.5

Jørn Olsen, Olga Basso

5.1	<i>Introduction</i>	1045
	Reproductive Epidemiology: Reading Instructions	1045
	Reproductive Health – Specific Epidemiologic Research Problems.....	1046
	Pregnancies as Repeated Events: Problems and Design Options.....	1049
5.2	<i>The Case-Parent-Triad Design</i>	1056
5.3	<i>Infertility and Subfecundity</i>	1057
	Measures Used to Describe Fertility	1057
	Design Options in Studies of TTP	1059
5.4	<i>Twins</i>	1062
5.5	<i>Measuring Reproductive Failures</i>	1063
	The Measures Used to Describe Mortality	1065
5.6	<i>Foetal and Infant Death</i>	1067
	Perinatal Mortality and Health Care.....	1071
5.7	<i>Foetal Growth and Birth Weight</i>	1072
	Optimal Birth Weight	1074
5.8	<i>Gestational Age: Pre- and Post Term Delivery</i>	1074
5.9	<i>Congenital Malformations</i>	1079
5.10	<i>Pregnancy Complications</i>	1082
	Operational Definition	1082
	Methodological Challenges.....	1083
	Gestational Diabetes.....	1088
	Pregnancy-Induced Hypertension and Pre-Eclampsia	1090
5.11	<i>Delivery Complications</i>	1095

5.12	<i>Foetal Origins of Adult Diseases</i>	1097
5.13	<i>Sources of Data</i>	1098
5.14	<i>Conclusions</i>	1099
	<i>References</i>	1100

Introduction

5.1

Reproductive Epidemiology: Reading Instructions

5.1.1

In writing this chapter we assumed that the reader is familiar with the basic concepts in epidemiology. You will not find any overview of different designs, measure of disease occurrence, standardisations, other ways of adjusting for confounders, or any general discussion on bias, confounding or on measuring effects. If you are not familiar with these topics you should start by reading other parts of the book or turn to one of the many fine available textbooks.

Our intent is to point out the problems and methods that are of particular interest in reproductive epidemiology and the areas that are perhaps peculiar to this field of epidemiology.

We focus upon methodology and, for the most part, we avoid reporting on any 'state of the art' overview on what is known about specific exposures or endpoints. Such reviews are soon outdated and, furthermore, space does not permit them.

By focusing upon some of the aspects that set reproductive epidemiology somewhat apart from other areas of epidemiology, we hope to alert readers to the problems and options that this field presents and to illustrate how important it is to develop a highly critical outlook. Research in reproductive epidemiology, like research in general, is not to prove or confirm anything, but rather to question and make critical appraisals.

We use references to illustrate specific problems of methodological interest. We use more of our own work than could ever be justified on the basis of our modest contribution to the field. Our excuse is that these references reflect our source of information for learning about the many problems inherent to reproductive epidemiology. We cannot rule out that they also reflect our inflated egos. Most of the references we have selected present information of methodological relevance. We also provide references to full textbooks in reproductive epidemiology.

We do not in any way claim to provide a complete list of problems, which you should be aware of as a student of reproductive epidemiology. We do not know all these problems ourselves: some are yet to be described and some have not yet caught our attention. Even describing all the specific problems that we are aware of would require more space than you, the reader, would like us to have. What we have tried to do is to describe the most important problems as we see them. Our choice is a subjective one, reflecting our experience.

Reproductive health was defined by the WHO at the Cairo conference in 1994 as:

... a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity, in all matters relating to the reproductive system and its functions and processes.

Reproductive Health, therefore, implies that people are able to have a satisfying and safe sex life and that they have the capability to reproduce and the

freedom to decide if, when, and how often to do so. It also includes sexual health, the purpose of which is the enhancement of life and personal relationships, and not merely counselling and care related to reproductive and sexually transmitted diseases.

We do not intend to cover all the possible topics related to this definition, nor do we intend to provide research methods for studying well-being or even happiness. We will limit our focus to the more traditional domain of epidemiology, namely to the studies of determinants of diseases directly related to reproduction, as long as these studies can be applied to human populations. Although many diseases will have an effect on procreation through biological, psychological or social mechanisms, we will restrict ourselves to studies dealing with fecundity, pregnancy, birth, and early markers of child health. Many of the diseases we describe only manifest themselves in the time period of reproduction; such as subfecundity, pre-eclampsia and, of course, all the diseases related to the child.

Our experience mainly stems from research in Europe and the USA, and we do not cover research problems of particular relevance to developing countries. Since reproductive health problems are usually larger in these countries, we are aware that this is a major shortcoming, and our only excuse is our limited experience in this field. We do want to stress, however, that research in developing parts of the world should also be based upon sound methods. We do not believe in low quality research anywhere, but we accept that circumstances may set limitations for what can be done.

Many diseases of the reproductive organs, like cancer or infections, may have an effect on reproduction if the diseases appear before or during reproductive age. In most cases, studying the determinants of these diseases will be similar to studying determinants of other diseases and, as such, they are not pertinent to the analysis in this chapter.

Our aim is not to provide a cookbook for research in reproductive health. Our aim is only to make the reader aware of the aspects he or she should be concerned about. Associations come in many shapes, and they are not always what they pretend to be. 'Be careful' is the main (and perhaps the only) message we want to convey, and if you are happy with that you could stop here – if you want to know why you should be careful, please go on reading. We expect you to disagree on several occasions. Remember, you may be right and we may be wrong.

Reproductive Health – Specific Epidemiologic Research Problems

5.1.2

Unlike epidemiologists studying cancer and chronic diseases, reproductive epidemiologists deal with an area that has been shaped by evolution. Selective forces operate before and during pregnancy, even in the highly medicalized industrialized world; keeping this feature in mind when dealing with reproductive epidemiology is important when trying to understand events that occur in pregnancy.

Most epidemiologists deal with identifying determinants of diseases, which might operate over time spans of varying length. The induction periods may be very short (e.g. is a migraine attack triggered by a given nutritional component?) or may span many years (e.g. does prenatal exposure to smoking affect semen quality?) In reproductive health we study determinants of factors that play a role for successful reproduction. Sometimes this involves studies that span generations, and many studies address couples, or families, rather than individuals. Unlike other fields of epidemiology, the individuals studied in reproductive epidemiology are often not ill in the common meaning of the word. Women who have repeated abortions or give birth to dead or severely malformed children are usually healthy themselves, and so are their partners. Similarly, even women who become severely ill during pregnancy, for instance with pre-eclampsia, often fully recover if the pregnancy has been interrupted in time. Their risk of experiencing pre-eclampsia in their next pregnancy is, however, relatively high.

In many cases, reproductive epidemiology deals with hidden phenomena that may represent serious disorders, which, however, may never come to light if a woman does not become pregnant. Furthermore, many women experience several pregnancies in their lifetime, so that a pregnancy rather than an individual will be the unit of analysis. All these elements intrinsic to reproduction provide interesting design options and challenging methodological problems, and they set reproductive epidemiology somewhat apart from other areas of epidemiology.

Successful reproduction in evolutionary terms means that the parents' genes are transferred to viable offspring, who will, eventually, produce offspring of their own. Part of this process is subject to epidemiologic observation. Although the maternal and paternal evolutionary interests are similar as far as gene transmission is concerned, the mother plays an additional role. She provides not only half of the foetal genes, but also the environment in which the foetus develops and is nourished over the duration of pregnancy and, in many cases, also for a period of time after birth (through breastfeeding). Pregnancy can be life threatening for the mother, and it is estimated that in Africa the lifetime risk for a woman to die of pregnancy-related causes is 1 in 16, while in more developed countries this figure is 1 in 1800 or even lower (AbouZahr 1998). Situations may thus occur when a pregnancy, which is too costly, is interrupted for the sake of the mother's chance of reproducing again (Haig 1993). The uterine environment's role is therefore of great importance, but it is often difficult to disentangle its effects from those that are genetically mediated by the mother. Sometimes it is the interaction between foetal and maternal genes that may trigger adverse outcomes, such as recurrent spontaneous abortions. Reproductive epidemiology has acquired a powerful tool with the new genetic technologies. Using these tools will, hopefully, provide new important clues on the complex events taking place from conception to birth and onwards.

Most of the time, we can only obtain data on deliveries or abortions (and usually only on abortions of clinically recognised pregnancies, thus missing the early ones) and most of the processes leading to these events are hidden or altogether absent

from the data we have access to. We thus have to take into consideration what potentials for errors this entails.

The time period of reproduction is short within a life span in most countries, and it is under intensive surveillance. What we are able to study is often the causal links that remain after health care providers have tried to prevent negative outcomes. Data on births and, sometimes, spontaneous abortions are routinely collected in several countries at national levels following standardised procedures, therefore allowing comparisons over time and place on some occasions. In some countries, at least part of the data are computerised and thus of relatively easy access to epidemiologists. Most countries produce reports of some indicators of reproductive health at regular intervals.

Reproductive health, at least the part associated with pregnancy and childbirth, covers events that are often frequent and serious, such as infertility, spontaneous abortions, preterm birth, and pre-eclampsia. Events that we would rather prevent than treat, and most are willing to accept that prevention must be based upon research of good quality. We not only need to identify determinants of reproductive failures connected to lifestyle, occupation, environment, diet, or medication, but also to find out how antenatal care (ANC) can be best organized. Antenatal care is, in many countries, the most expensive part of preventive care within the health care system. At present, only a limited part of ANC is evidence-based and much more research is still needed, especially in countries where resources are sparse. Much of this documentation has to be time- and place-specific. Evidence on health care technology cannot just be transferred from one country to another. But many of the findings related to causation will apply to most populations. Treatment effects may also be non-particularistic in some situations.

By developing more and more efficient contraceptive methods and, especially, by producing more and more sophisticated methods for treating infertility, we are increasingly interfering with the forces of evolution. We should make it one of the priorities of reproductive epidemiology to study whether and to what extent these factors will affect future generations. Reproduction is, in many ways, a playground for new technologies that are being introduced without much concern for the long-term consequences. Children of in vitro fertilization (IVF) are still too young for us to study whether they may suffer any long-term health consequence associated with the mode of their conception. Interventions like intracytoplasmic sperm injection (ICSI) may undermine the very core of the natural selective forces behind reproduction. At present we do not know if ICSI is a safe assisted reproductive technology (Ludwig and Diedrich 2002), although early reports are reassuring to some extent.

Trying to solve health problems for some may introduce health problems for others, especially in reproductive health related to treatment of infertility. Some of those health problems may never come to the attention of the treating physician and must be disentangled by research. In relation to this and many other technologies, epidemiologists have an important role to play in monitoring disease data over time in the population. Reproductive health should be an important part of any health surveillance system.

It is peculiar to reproduction that during pregnancy the mother provides most of the exposures that may influence the future health of the unborn child. The foetus is only partly protected by the placental barrier. Many toxic substances pass this barrier and the developing foetus will often be less able than the mother to metabolise these substances, especially early in pregnancy. This is e.g. the case with alcohol, caffeine, and several drugs.

Thus, another important feature of reproductive epidemiology is that not just one individual is involved in the process under study, but three: the mother, the father (through foetal genes and, possibly, substances carried by semen) (Savitz et al. 1994), and the foetus itself through its continuing interaction with the mother during the pregnancy period. When studying reproductive outcomes it is, in many cases, insufficient to focus only on the mother, although she usually represents the most accessible player.

It is now known that during pregnancy and delivery there is a two-way traffic of cells between the mother and the foetus, and that these cells can persist in the bloodstream for decades (Bianchi et al. 1996). This phenomenon, called 'microchimerism', is a quickly developing area of research, as it is suspected that these cells may interfere with the immune system and thus contribute to the aetiology of several diseases, especially those of autoimmune origin (Bianchi 2000; Whitacre et al. 1999). That pregnancy itself, or phenomena related to pregnancy, might have something to do with the aetiology of these diseases is suggested by the fact that females are more subject to autoimmune diseases than males, although these differences may also depend on immunological, hormonal, or other factors that are related to sex but not specifically to pregnancy (Whitacre et al. 1999; Whitacre 2001). Changes in physiological status during pregnancy may have health implications for the mother as well as for the child, and the growing foetus totally depends upon maternal supply of nutrition and oxygen. Changes in this supply may 'reprogramme' organ development with possible long-term health consequences (Barker 1994).

Pregnancies as Repeated Events: Problems and Design Options

5.1.3

Each pregnancy provides a new set of observations and, in populations with a high fertility, a woman may be subject to repeated studies of exposures of interest for the outcome of each pregnancy. A pregnancy is a new event that, at the same time, is correlated with the previous pregnancy/ies. Most reproductive failures (such as spontaneous abortions, preterm delivery, pre-eclampsia, etc.) have a tendency to repeat themselves because time-stable causes (e.g. genes or part of the maternal environment) are present during all events. However, it is not clear how this aspect should be taken into consideration when analysing data.

In animal studies, the variance within a litter and between litters is dealt with separately, probably even successfully, in the statistical analysis, but the situation is

different for humans. Human births are usually sequential, except for multifoetal pregnancies, which, however, are relatively rare and often not comparable to single pregnancies. Methods have been developed to take into account the statistical dependence resulting from women contributing more than one pregnancy (Hoffman et al. 2001), but we have limited confidence in these methods (Olsen and Andersen 1998).

Each new pregnancy has a number of features in common with the previous pregnancy, and a number of features that are unique to it. In some situations, each new pregnancy is an independent event and should be treated as such, but the contrary argument can be made as well. However, even if the dependency between pregnancies were to be modelled in the analysis, it is not clear whether doing this would produce a different type of bias, because the statistical model will not take selective change in behaviours related to pregnancy history into consideration. Pregnancies are, in many cases, planned as a function of the outcome of a previous pregnancy, and this aspect can hardly be addressed by methods aimed at accounting for the statistical dependency between pregnancies. An unwanted abortion or a stillbirth may be replaced very shortly with a new pregnancy in order to have the desired child. A surviving child with severe handicaps may delay or stop further childbearing, while parents of a child that dies with the same handicap may attempt a new pregnancy soon. Some exposures will be avoided as a result of how previous pregnancy ended. A mother who is a heavy smoker and had a child with a very low birth weight may well give up smoking when she becomes pregnant again, even in situations where smoking was not the only or even the major cause of the low birth weight. If smoking was not the only cause of low birth weight, she may then become a non-smoking pregnant woman with a high a-priori risk of getting a child with a low birth weight. How should that be addressed in the statistical analysis? Or how would we be able to design a proper counterfactual comparison? This type of confounding by the risk that triggers a change in exposure is very difficult to rule out, and it is a common and often neglected problem. A valid analytical solution may be to study first pregnancies only (Olsen 1994), but this approach may prove to be too conservative in the sense that it removes available information and will thus reduce statistical power. Furthermore, there is heterogeneity on outcomes depending on parity, some of which may be due to selection (not all women will be able – or want – to have a new pregnancy), others probably depend on physiology or on changes in the uterine environment brought about by the previous pregnancy/ies. A first pregnancy modifies the uterine arteries in such a way that placentation is facilitated in the successive pregnancies.

When studying the effect of a given determinant on an adverse reproductive outcome, one may be tempted to stratify on (or adjust for) a previous occurrence of that outcome (e.g. spontaneous abortion). Such a temptation should be resisted if the intent is to study aetiology, since adjustment for any factor that is caused in part by the exposure under study and is also correlated with the outcome will likely bias the estimate (Weinberg 1993). If the intent is purely descriptive, however, this approach is acceptable.

Beyond the difficulties and traps caused by the fact that women have more than one pregnancy, however, this feature also provides reproductive epidemiologists with a powerful tool that is unique to this discipline. Many reproductive outcomes tend to recur in different pregnancies of the same woman, and this provides the opportunity to examine whether some putative factors play a role in the aetiology of a given event by studying women who had the outcome in question in a pregnancy and estimating whether their recurrence risk changes accordingly as a function of a change in a given factor in between the two pregnancies (Olsen et al. 1997). This design, which we called 'the computerized square dance design', makes it possible to estimate whether the paternal genome plays a role by studying maternal half siblings. It can be applied to reproductive failures as well as disorders occurring in early childhood, such as febrile seizures (Vestergaard et al. 2002). One of the advantages of this approach is that a great deal of confounding is adjusted for by using the woman as her own control. One aspect that has to be taken into consideration is that time plays a role, and thus the interval between pregnancies may have to be taken into account, especially when studying the effect of changing partners, since women who change partner tend to have a much longer interpregnancy interval (Basso et al. 2001; Skjaerven et al. 2002).

The Problem of Incomplete Denominators

Ideally, we would like to be able to study the outcome of all conceptions that take place in a given population in a given time period, in order to be able to observe how many end in very early losses, how was the karyotype of the lost fetuses, their sex ratio, etc. Less ideally, we would like to be able to obtain information on all conceptions surviving the first 8 weeks of gestation, because missing them may lead to serious bias in studies examining specific exposures. In many instances, reproductive epidemiologists work with incomplete denominators.

Spontaneous Abortions

A frequent outcome of a pregnancy, spontaneous abortion, is probably to a large extent part of nature's own quality control system (Quenby et al. 2002). Some spontaneous abortions are, however, man-made and could in principle be avoided if their causes were identified. Some abortions occur in the pre-clinical phase, before the pregnancy is recognised, while others occur after the pregnancy is recognised. The timing of observation thus becomes of crucial importance.

The best approach would be to start observation before conception, but this requires access to women who plan their pregnancies (Wilcox et al. 1988). Pregnancy planners are likely to include an excess of subfecund women (women with a low probability per cycle of conceiving), as women who become pregnant as a result of contraceptive failures are less often subfecund. Participating women may thus be at higher risk of a number of adverse outcomes, since subfecundity is correlated with several reproductive failures (Basso et al. 2003).

Currently, few investigators have attempted to detect pregnancies by using biomarkers (usually hCG, human chorionic gonadotropin). These studies (Wilcox

et al. 1988; Bonde et al. 1998a, b) do not tell us how frequent abortions really are, but they suggest that at least 30% of conceptions end as spontaneous abortions and that a little more than half of these occur in the pre-clinical phase.

An exposure that only delays abortions without increasing their incidence would appear as a risk factor in a study based exclusively upon recognised pregnancies. Such an exposure may move an abortion from the preclinical phase to the detectable phase. Exposures that advance the time of abortions would appear to prevent the occurrence of abortions. Such an exposure would produce a low abortion rate because abortions now would occur in the pre-clinical phase. For this reason, the timing of pregnancy diagnosing is important. If women who take a long time to become pregnant seek earlier confirmation of pregnancy (compared to women who have not waited a long time), they will also be aware of early losses that would not be detected by women who were not aware that they were pregnant. This type of bias may be partly responsible for the association between subfecundity and spontaneous abortion, where good quality data on early conceptions are available (Baird et al. 1993). The use of early pregnancy tests may play a role in analysing and interpreting data from studies based upon pregnancy planners. Such studies are difficult and expensive to conduct and require a very cooperative study population.

If you study environmental determinants of abortions, you may like to exclude 'habitual aborters' – women who will abort any pregnancy no matter what. Such women will not provide information related to risk following the exposure (Gladen 1986; Weinberg et al. 1994b). The problem is that these habitual aborters can only be identified by their abortion history and thus cannot be identified at all if they have no pregnancy history at all. Stratifying results on pregnancy number can distort associations, since only some of the women with many abortions will be habitual aborters. Some will abort due to chance and some will abort due to the exposure under study.

Measuring Infertility

When we are interested in measuring the biological component of fertility, called fecundity, we are often able to ask pregnant couples how long it took them to become pregnant. While doing that we also need to identify planned pregnancies, which may be difficult since planning is often not a well defined concept. We also need to find a way to deal with couples who had not planned their pregnancy as well as with those who became pregnant despite using contraception. Their underlying fecundity may differ from that of couples who plan a pregnancy. We would also like to have information on all couples that conceived and had an early pregnancy loss (Jensen et al. 1998) or failed to conceive because they were sterile or gave up trying for any reason. When studying time to pregnancy in samples of pregnant women, one must be reasonably confident that giving up a pregnancy attempt is completely independent of the putative risk factors under study, a condition similar to what we encounter when working with censored data in general (Basso et al. 2001).

Congenital Malformations

Congenital malformations are relatively frequent and they constitute a major cause of infant mortality and morbidity. They are also a very stressful event for the involved families, making it difficult to get comparable information in a case-control study addressing determinants of congenital malformation.

It is now well accepted that congenital malformations are usually measured as a prevalence at the time of birth. The incidence of congenital malformations is normally not available for study, given that it is a function of new events since time of conception. Congenital malformations that occur in utero and end as abortions, some as very early abortions, are usually not detected. Some exposures may simultaneously increase the incidence of some congenital malformations while at the same time reducing the survival rate of the affected foetuses (or embryos) in utero, thus decreasing the prevalence at birth. Monitoring systems of congenital malformations with no data on spontaneous abortions could therefore miss important teratogens and even wrongly conclude that a given factor is protective. Many of the established monitoring systems on the possible teratogenic effects of medicines taken during pregnancy are based only on data on prevalence at birth. When dealing with spontaneous abortions and congenital malformations, missing early losses may lead us to biased effects measures. Many of the routine monitoring systems are furthermore of poor quality, partly because only some of the congenital malformations are visible at birth. Heart defects may e.g. only produce symptoms – and thus first surface to clinical detection – under extreme physical strain late in life (Knox et al. 1984).

Time Matters

If you were a student of mortality you would know that the question is not *if* people die, but *when* they die. Mortality rates (MR) reflect this time function, as $1/\text{MR}$ is the life expectancy in the same time unit as the rate is measured, given that a number of conditions are fulfilled. All estimates of risks come with a time tag. The estimate depends upon the length of time it represents. Time is underlying all occurrence research, even when it is not explicitly mentioned. All events happen in time. If you do not wake up in the morning (or later), it is because you are dead. You ran out of time.

In reproductive health, time is important from several points of view, not only as the time from exposure to the endpoint of interest. The timing of exposure itself is more important in this area than in most others. Specific windows of vulnerability open and close over the time of gestation. The time of organogenesis and organ development plays a crucial role. The time periods of interest may even date back generations. Not only may malformations of genital organs impair reproduction, but also the number of Sertoli cells is at least partly determined in foetal life and this has implications for sperm production lasting decades into adult life (Sharpe and Skakkebaek 1993; Wilcox et al. 1995). Organ development is under the influence of hormonal factors that operate at certain time periods. Fetotoxic exposures may have different outcomes as a function of the timing of exposure,

and growth-determining factors may only play a role during a time period of rapid foetal growth.

The time period of spermatogenesis and ovulation may be under the influence of external exposures and genetic factors. Factors that reduce the sexual libido affect quality of life at any time but have consequences for reproduction only if present during the time window of procreation.

Soon after conception, the risk of abortion is high, but we know very little about the determinants of early abortions. Avoidable abortions probably depend to a great extent on the timing of the exposure that causes the abortion, but this time period may be difficult to identify, partly because foetal death need not coincide with the expulsion of foetal tissue. Fever could e.g. appear to be a cause of spontaneous abortion (Andersen et al. 2002), because fever may be induced by dead foetal tissue. In that case, the association would be from foetal death to fever and not from fever to foetal death (reverse causation). However, Andersen et al. did not find that fever was associated with spontaneous abortions.

It is a point of fact that the period at risk for teratogenic actions (usually the second and third months of gestation) is almost over when antenatal care (ANC) begins. All legal actions taken to reduce e.g. occupational exposures during pregnancy are usually put into operation too late. Some types of prevention (like the use of folic acid to prevent neural tube defects) need to be activated even before the pregnancy is planned. If a given toxicant (such as lead) is stored in the maternal tissue over a long period of time, its mobilization during pregnancy may affect the mother or the embryo many years after the exposure took place (Rothenberg et al. 2002).

It is possible, perhaps even likely, that relevant exposure windows for specific effects are only open during short time periods. Exposure to influenza virus has e.g. been associated with schizophrenia in the offspring, but only for infections during the second trimester of pregnancy (Mednick et al. 1994). On the other hand, neurotoxic exposures may influence brain development, if they operate at the time of the vulnerable structural changes that take place throughout foetal life and in early childhood. It is possible that diseases like ADHD (Attention Deficit and Hyperactive Disorders), autism or Tourette's syndrome have a pre- or perinatal aetiology (Linnet et al. 2004).

The nutritional demand of the foetus is at its maximum in late gestation, when foetal growth is rapid. Smoking in this period is much more closely associated with low birth weight than smoking early in pregnancy (Smoking or Health 1977).

The Dutch famine study indicated that the timing of under-nutrition during gestation played an important role for later health outcomes (Susser and Stein 1994). Being exposed to undernutrition in the last trimester was associated with obesity later in life (Ravelli et al. 1976).

Most side effects of drugs taken during pregnancy depend heavily on the timing of the drug intake within the gestational period. This is true not only for teratogenic effects. Drugs are particularly difficult to study as an exposure, because the effect of the drugs may not always be easily discerned from the effect of the indication for taking the drug in the first place (what is usually termed 'confounding by

indication') (Olsen et al. 2002). If two or more drugs with the same indication but different active molecules are available and are prescribed to pregnant women independently of the severity of the indication, this may provide useful options for better characterizing the effects related to one of these drugs.

The foetal alcohol syndrome (FAS) is characterised by specific facial characteristics, low birth weight, and cognitive impairment. It is likely that these impairments are the result of time specific exposures and, if so, binge drinking may be of concern, even binge drinking among low or moderate alcohol drinkers. Still, no studies in humans have shown that binge drinking alone causes cognitive impairments or any other important FAS characteristics, perhaps because all studies have been too small. If the exposure window is short, then only a small fraction of binge drinkers will be at risk and small studies will thus have no power to detect an effect unless all the study subjects are exposed at the right gestational time. FAS is seen in children of some women, who drink large amounts of alcohol every day. A high daily intake leads to exposure during the time periods of vulnerability, regardless of their duration.

Exposure to mercury, determined on the basis of amounts found in umbilical cord blood at birth, has been associated with impaired cognitive functions (Grandjean et al. 1997). An umbilical cord measurement reflects an average exposure throughout gestation and, as such, it does not indicate whether certain time periods are more vulnerable than others.

Hormonal factors play a profound role in foetal life. Sexual hormones are e.g. needed for a foetus to develop a male phenotype. It is furthermore believed that hormonal factors influence the probability of developing some diseases in adult life. For example, it has been suggested that a high intrauterine oestrogen level may modify breast cancer risk (Trichopoulos 1990) and reduce the number of Sertoli cells, thus subsequently reducing semen production and increase the risk of undescended testis, hypospadias and testis cancer. This latter hypothesis is, however, not corroborated by epidemiologic findings in general and it probably has to be modified to fit existing data (Strohsnitter et al. 2001). There is, on the other hand, clear evidence of an association between some of these diseases. Patients with testis cancer have a higher frequency of undescended testis and apparently a lower fecundity before the cancer is diagnosed (Jacobsen et al. 2000) Macro-epidemiologic (ecologic) studies furthermore show that the geographical variation in e.g. testis cancer is often followed by similar geographical variations in low sperm counts and a high prevalence of malformations of the male genitalia. The link between testis cancer and undescended testis is considered established, although most studies have been based upon self-report of undescended testis, often reported retrospectively. Many undescended testis descend spontaneously shortly after birth or at puberty. More recent studies based upon recording at birth show a weaker association between testis cancer and undescended testis (Stang et al. 2001; Sabroe and Olsen 1998). It is possible that only persistent undescended testis correlate with the risk of testis cancer. It is unclear whether descent through treatment eliminates the risk or whether the risk is caused by the underlying reasons of the displacement itself. If the risk is a function of underlying hormonal

disturbances at the time period of organogenesis, treatment is not expected to affect cancer risk.

Single peak exposures to drugs are seen in pregnant women who try to commit suicide by taking an overdose of medicine. In principle, these studies provide unique data for studying the fetotoxic effects of these drugs, since they often bypass confounding by indication (unless the study concerns anti-depressant drugs). Many of these studies, however, have not taken abortions (induced or spontaneous) into consideration and have often been too small to detect teratogenic effects, since such studies must focus upon suicide attempts in specific and short time periods. Furthermore, many women terminate their pregnancy after a suicide attempt because of fear of a fetotoxic effect or because of their inability to cope with a pregnancy. Newborn children available for this type of study are thus few and, to some extent, selected. Most of the findings from these studies have been reassuring, indicating that healthy babies are often born after a single exposure to some specific drug (Flint et al. 2002).

Although it is generally accepted that timing is of crucial importance for most outcomes of pregnancy, many studies rely on exposures where the timing is not specified. Many studies that make use of retrospective data from pregnancies and exposures are thus difficult to place in time, when they took place months or years before reporting. Large cohort studies have been started or are being planned in order to provide researchers with prospective data on exposures (Olsen et al. 2001).

5.2

The Case-Parent-Triad Design

There are several design options for studying genetic risk factors. Most of these are described in Chaps. I.7, III.6 and III.7 of this handbook. One of these designs is, however, of particular interest in reproductive epidemiology, namely the case-parent-triad design. The idea is to use parents of affected children to serve as genetic controls. Case-control studies may easily be confounded by genetic factors when they are applied to populations with a mixture of different ethnic groups. The parents are by definition ethnically matched to the case, and they are furthermore usually motivated to take part in the study. The two non-transmitted parental alleles can be compared to the two transmitted alleles, and since this Mendelian transmission occurs at random the observed allele structure can be compared with the expected values. If the parents e.g. both have the allele structure Aa, the children will under Mendelian transmission be AA, Aa, Aa or aa with equal probabilities; 25% probability of AA and aa, respectively, and 50% of Aa.

If the genotype is associated with the disease under study, affected children will have an allele distribution that deviates from these expected values. All this may be analysed using well-known log-linear models (Weinberg et al. 1998). The main limitation in using this design for diseases with an adult onset is that the assumption of non-selective survival of the parents to the time of the study is a strong assumption, at least for severe diseases. For diseases with an early onset

the assumption will usually be fulfilled. Using this design in e.g. studying genetic causes of congenital malformations is an attractive option. The parents will be available, motivated, and present in a setting where a blood sample can easily be taken at any time.

Infertility and Subfecundity

5.3

Measures Used to Describe Fertility

5.3.1

The fertility rate is defined as the number of live births a woman has during her reproductive life. A fertility rate of 1.3 means that in a population of 100 women there will be, on average, 130 liveborn children – not enough to replace their parents. A fertility rate of more than 2.0 is needed to bring the population into a steady state. How much more depends upon the life expectancy and the sex ratio of the newborn children in the population. On average, one woman should produce one girl who survives until her reproduction age in order to bring the population to a steady state if life expectancy is stable over time.

Fertility is determined by both a biological capacity to reproduce (fecundity) and the desire for a given family size, which may be under the influence of social and cultural conditions broadly defined, and methods available for family planning play, of course, a crucial role. Fertility is a term used by demographers to describe the actual production of live children, whereas infertility is used by the medical profession to describe a reduced biological capacity to reproduce. This dual use of the same terminology is confusing. The best would be to let fertility (and infertility) be reserved for demographers to indicate *de facto* reproduction.

We suggest the term fecundity to be used as a broad term to describe the biological capacity to reproduce, and fecundability (the probability of conceiving within a given menstrual cycle) to be a quantitative estimate of fecundity. Unlike Cramer and Goldman (1994), we propose to let the more specific details of the terms be defined in the actual study in order not to have too many words describing rather similar conditions. Reproduction, in its historical form, requires sexual contact, fertilisation, implantation, and survival until the age of reproduction. Now alternative methods exist.

Most studies use a recognised pregnancy (or birth) as the endpoint, and recognition of pregnancy usually depends upon clinical or biochemical measures. Fecundity could therefore, depending on the specific study, describe the ability to obtain a biochemically detected pregnancy (by means of hCG), a clinically recognised pregnancy, or a pregnancy that led to a liveborn child. Fecundity can thus be used to describe the capacity to achieve any of these endpoints, which may be confusing. On the other hand, having a specific term for each of these situations would complicate communication to an even larger extent. The fact that we use the term fecundity to refer to a variety of situations has to be kept in mind.

Childlessness can be voluntary or involuntary, and the latter may be subject to epidemiologic research. Subfecundity is a frequent problem, and treatment for subfecundity is a rapidly growing sector of many health care systems. Research that can potentially lead to prevention of subfecundity is therefore receiving increasing attention, and epidemiology plays an important part in this research. Subfecundity is a measure for a couple and is defined on the basis of unsuccessful attempts for a given length of time (usually 6, 12, or 24 months). In many industrialised countries, about 15% of all couples that try to become pregnant will experience subfecundity, if this is defined as a waiting time of 12 months (Juul et al. 1999). The term infertility usually describes an unsuccessful waiting time of 12 or 24 months. Often the cut-point of 12 months is used in affluent societies and 24 months in countries with limited health care resources.

Sterility is defined as an absent capacity to reproduce – a fecundability of 0. Since most couples' probability of conceiving is very rarely 0, many women will eventually become pregnant, if they keep trying, even when on an ineffective treatment.

Time to Pregnancy

If a normal fecundability is 0.25, then 3% of normal couples will not succeed within 12 months of trying ($(1-0.25)^{12}$) just because of bad luck. These couples need no treatment, but may (and some will) be treated nonetheless, usually with an 'excellent prognosis' (25% success rate for each cycle). The remaining 12% would be the proper candidates for treatment. If all couples who wait unsuccessfully for 12 months or more to become pregnant receive treatment, then any treatment will to some degree be a success. Some couples will have normal fecundity, and some will have subnormal fecundity without being sterile. An effective treatment has to demonstrate better performance than chance alone would produce.

If a couple is defined as subfecund after an unsuccessful waiting time of 24 months, then only about 0.1% of 'normal' couples will meet the definition, and those targeted for treatment will include 99.9% of actually subfecund couples. For this reason, less affluent societies do not start infertility treatment until couples have tried for at least 24 months.

Since fecundability is related to the duration of a waiting time to pregnancy (TTP), TTP is a frequently used measure in subfecundity research. The measure was first used by demographers when they examined the time from marriage to the first liveborn child; in this context, TTP is a measure that only has biological relevance in societies where procreation starts with marriage and contraception is not practiced.

In societies with a higher degree of family planning, TTP (now defined as the number of cycles that a couple take from the moment they first start trying to when they actually conceive a clinically recognised pregnancy) becomes a useful tool in the study of determinants of subfecundity. It has been used in epidemiology since the early 1980s (Rachootin and Olsen 1982, 1983; Olsen et al. 1983; Baird et al. 1986). It appears that women are able to recall with sufficient accuracy how long

they took to conceive, even after several years (Joffe et al. 1993; 1995). Time to pregnancy is easy to use and, perhaps for this very reason, has undoubtedly been used too frequently without proper concern for its shortcomings and pitfalls.

Design Options in Studies of TTP

The best option is to study TTP in a longitudinal design that starts when couples stop using contraception in order to conceive (starting time). Exposures of interest can then be recorded independently of the length of TTP and at the relevant point in time. They can also be registered for attempts that do not lead to a pregnancy. Studies that rely on exposures recorded at the time of pregnancy rather than the starting time may produce biased results, especially for exposures suspected to cause infertility. If smokers stop smoking after having tried in vain to become pregnant for a certain time period, but not if they conceive quickly, smoking recorded during pregnancy will correlate with a short waiting time, as if smoking prevented subfecundity. In fact, almost all evidence points towards the opposite assessment: smoking impairs fecundity, especially in women (Rachootin and Olsen 1983; Baird and Wilcox 1985; Bolumar et al. 1996; Alderete et al. 1995).

The proportion of couples that become pregnant during the 1st cycle is an estimate of the fecundability rate in the population. The entire TTP distribution will, however, normally be used in the analysis, e.g. in a discrete Cox model. Such a study need not last 12 or 24 months. Even a study with a follow-up time of only one cycle could provide evidence that smoking women have, on average, a fecundability that is approximately 20% lower than that of non-smoking women. We would then e.g. expect 30 out of 100 non-smoking women to be pregnant within the first cycle versus 24 out of 100 smoking women. Exclusion criteria may be used to get rid of non-informative couples in the study of environmental causes of sub-fecundity. Sexual activity could be recorded and taken into consideration in the analysis as well as time of pregnancy detection. Such a study is straightforward to design and to analyse, but extremely difficult and expensive to carry out, unless it is based upon couples that are seeking treatment for infertility, like IVF patients. Results from infertility patients are, however, hampered by limited generalizability to the population at large, partly because the couples are highly selected, partly because of the forces of clinical selection that are being used. Spermatozoa, eggs, and fertilised eggs are selected by the health personnel according to criteria quite different from the 'natural' selection. A longitudinal study on pregnancy planners may also provide other related endpoints, such as information on early abortions or semen quality.

Epidemiologists have looked for less expensive designs than the concurrent follow-up of pregnancy planners. These designs have mainly been population based cross-sectional studies or designs based upon samples from pregnant women, both of which are prone to a number of problems.

Cross-sectional surveys rest upon the assumption that women recall their TTP with some accuracy even a long time after the event. In a survey, the selected women will try to record attempts to become pregnant, exposures of relevance – often from the same time period – infertility treatment, and TTP (or time waited

in unsuccessful pregnancy attempts). Population studies of this type are relatively inexpensive and they usually rest on random sampling principles. The main problem is accurate recall, especially of exposures at the relevant point in time (the starting time), and the ability to obtain response rates that do not introduce selection bias. Analysing data may be simple, although TTPs usually have to be recorded in quite broad categories, and often referring to months rather than to cycles. Women may be able to recall if they had to wait for 6, 12 or more months to become pregnant, but usually they will not be able to remember if they became pregnant after 5, 6 or 7 months of waiting time when that waiting time took place several years ago. (They will probably remember if they became pregnant in the first month of trying.) Digit preference also has to be taken into consideration, as clusters of reporting will be seen at specific waiting times (such as 6 or 12 months).

Using pregnant women in data collections has, for obvious reasons, been a convenient design. In most countries, pregnant women are easy to locate and to contact. They are usually more willing to take part in studies than other women, and they are usually able to accurately report their TTP if the pregnancy was planned. They usually remember when planning started, and may be able to recall the exposures around the starting time, at least if the planning did not start too long ago, or if the exposure refers to the occupation they held at the time, or whether they were smoking or not. Since pregnancy is a condition for participation in the study, this design is unable to pick up an all-or-none effect of an exposure. An exposure that causes sterility will not be identified in a sample of pregnant women. Fortunately, most exposures that we know of are not of this type, and the exposures that reduce fecundability will show a longer TTP in a pregnancy-based study. The qualitative effect measure that comes from a study of this kind is, however, not a measure of fecundability in itself (Olsen and Andersen 1999), although it will correlate with fecundability under a number of conditions. The most important, and often neglected, of these conditions is the persistency in trying to become pregnant. This persistency should not be related to the exposure under study, because, if so, the exposure may falsely appear to increase fecundity (Basso et al. 2000b). The timing of diagnosing a pregnancy also plays a role. If couples wait very long to have a pregnancy verified, then an early abortion may count as waiting time. Had they used a sensitive pregnancy test early, the event would have been recognized as a TTP of a given length followed by an abortion. The length of the waiting time may influence the timing of detecting, thus potentially producing bias (Baird et al. 1993). Since socio-cultural factors may influence the couples' behaviour, this could pose a problem, especially if the exposure of interest is an occupational one that correlates with social and cultural conditions.

Exposures that cause irregular or long cycles may also interfere with the timing of pregnancy recognition as well as with the ability to report the starting time on the TTP.

The use of family planning methods should be taken into consideration. Couples who use unsafe contraceptive methods without getting pregnant will, over time, include an increasing fraction of couples with a low fecundability. Couples who

know how to time sexual intercourse around the time of ovulation have shorter waiting times than other couples, all other factors being equal. Couples who, in the past, have experienced subfecundity may have modified their exposures or their sexual behaviour in a way that may severely distort our ability to find a proper reference group.

It has also been suggested to calculate a standardized fertility ratio (Starr and Levine 1983). The idea is to compare the observed fertility with the expected fertility for couples of the same age and from the same region. Such a standardized fertility ratio can then be calculated before and after an exposure of interest – if the exposure impairs fecundity the fertility ratio after start of exposure is less than one, other things being equal. The observed number of liveborns before exposure may be close to expected values based upon age and calendar time specific rates in the population at large. The exposure may then reduce this observed/expected ratio. Although this method was able to detect the fecundity reducing effect of dibromochloropropane (DBCP), it rests upon strong assumptions, and it only works when examining exposures that have a specific (and known) starting point in time.

In addition to the problems mentioned, there are other sources of bias in time to pregnancy studies (Weinberg et al. 1994a; Weinberg and Wilcox 1998; Baird et al. 1994). It is not an easy task to develop a monitoring design sensitive enough to pick up subtle changes in fecundity over time (Olsen and Rachootin 2003). Studies that aim at detecting determinants of fecundity are probably less vulnerable to bias.

Only about half or less of those who experience infertility seek medical help, making treated patients a poor measure of the problem of subfecundity in the population (Olsen et al. 1998a, b). Infertile couples in infertility treatment are therefore a selected part of all couples in the population, which should be considered when using infertility patients in epidemiology studies.

If the forces of selection are correlated with the exposure under study, then a case-control study with a population-based definition of the source population is not a valid option for an analytic study. We expect these forces of selection to treatment to be related to several factors, such as cultural background of the population, age, parity, education, social status, and availability of treatment facilities. Some of these conditions may well correlate with life-style factors, dietary factors, infections or other putative causes of subfecundity.

Given these conditions the source population has to be defined by the case series. The source population could be defined by all those who would come to our treatment centre if they had a similar infertility problem as our cases had in the time period of study. The source population is then defined by all potential cases (and, of course, by the enrolled cases as well). Had we known the source population, we could have sampled controls from this group at the time the cases came to be detected. Candidates for controls selection would then be couples who have planned a pregnancy and who would be cases if they had a waiting time of 12 months or longer. We could then compare exposures at starting time, or before, and estimate the relative risk of being infertile as a function of the exposures under study. The problem in this design is that we do not know the

source population and it may be impossible, even in principle, to identify it along with its exposure experience. A viable option may be to take advantage of the fact that infertility is a couple phenomenon. Infertility is sometimes caused by exposures mediated through female factors, sometimes through male factors, and sometimes through both male and female factors. If we e.g. take an interest in exposures to prenatal tobacco smoking as a cause of 'male infertility' we may then compare the frequency of this exposure for males in couples that had a male problem (e.g. poor sperm quality) with the exposure frequency in males for couples where the female was identified as having the problem. Both sets of couples sought help and therefore belong to the source population. Since couples where both members have a fecundity problem belong to both the case and the control group, they do not provide useful information for the question under study and can be excluded.

If the exposure is more frequent in cases than in controls, it suggests that the exposure is a risk factor for the disease, although it may be impossible to calculate proper quantitative effect estimate from this type of a case/non-case study (Olsen and Basso 2001). This type of design may furthermore produce estimates that are too conservative if the studied exposures affect both male and female fecundity, especially if the exposures cluster in couples, like some lifestyle and occupational factors will.

Alternative design options for monitoring fecundity have been proposed like using dizygotic twinning rates as a surrogate for fecundity (Tong et al. 1997), but one of the problems is to exclude twins that are a result of infertility treatment.

Studies on semen quality may also be used as a surrogate measure of fecundity. Much of the concern for a declining fecundity stems from studies on semen (Carlsen et al. 1992).

The main problems related to using measures of semen quality in epidemiologic studies relate to low participation rates in most studies and difficulties with obtaining comparable conditions of the analyses. Time since last ejaculation has to be taken into consideration, the conditions related to the ejaculation itself, time from ejaculation to analysis, the technical conditions for the analysis, the season of sampling (higher sperm counts in winter than summer periods in temperate climates) as well as recording of potential confounders. Specific diseases may interfere with sperm production, together with external exposures.

5.4 **Twins**

Since twinning can be the result of infertility treatment, it is difficult to say what the 'natural' incidence is. It also varies in different ethnic groups – but the 'natural' incidence in Caucasian populations is probably around 1 in 80 births. Dizygotic twins may be seen as a sign of a high male and female fecundity, since they require two eggs to be fertilised by two sperms within a short time period.

Twins may also be seen as an anomaly, since twinning is associated with a higher risk of perinatal morbidity, including possibly congenital malformations. Twins, however, may not only be seen as a gift to their parents, but to epidemiologists as well, since they provide a very interesting source of data concerning the nature-nurture discussion in disease occurrence. Twin studies have been the basic epidemiologic design to disentangle the effect of genes and the environment.

Genetic disorders are expected to have higher correlation (concordance) in monozygotic (MZ) twins than in dizygotic (DZ) twins. MZ twins are genetically identical (although some minor genetic differences have been found to exist and have even been used to identify a genetic cause of a congenital malfunction (Kondo et al. 2002)) and DZ twins share genes like ordinary sisters/brothers. Compared to the latter, however, DZ twins share the same uterine environment as well as more similar conditions in early childhood and these factors considerably reduce the confounding that would arise by comparing ordinary siblings. Although the twin model is definitely of interest, the situation is clearly more complicated. MZ twins have the same gene map but need not have the same functional genetic expression. Since females are mosaics where a random X chromosome is inactivated in each cell line, female twins are not functionally genetically identical, and genetic imprinting may differ. Furthermore, MZ twins' intrauterine conditions are not the same, as inferred by the – often large – variation in foetal growth between babies in a pair. Furthermore, both twins have to survive to be part of a twin study and twin mortality is higher than what we see for singletons, most likely from the time of conception.

It is well documented that some singletons started as twins but one foetus did not survive, and it has been suggested that the surviving twin is at increased risk of e.g. cerebral palsy (Pharoah and Adi 2000).

Twins are often excluded from epidemiologic studies on possible fetotoxic hazards, since they often cannot be grouped together with singletons in a meaningful analysis due to their higher risk of low birth weight, preterm birth, and congenital abnormalities. In most studies, there are too few twins to provide an informative subgroup for analysis, which is unfortunate. It would often be of interest to study the effect of fetotoxic exposures on twins only.

Twin pregnancies have been used as a model to study the effect of hormone exposure during pregnancy. The oestrogen level is high in all twin pregnancies and twins of different sex also offer unique intrauterine exposure conditions (Storgaard et al. 2002).

Measuring Reproductive Failures

Estimating the frequency of reproductive failures is, in many ways, similar to estimating the frequency of any disease in populations. Frequencies are measured by means of rates or proportions (and many proportions are unfortunately called

rates). Measurements address new events over time as incident events or existing states at a given point in time, prevalence.

Rates are used to present reproductive failures as a function of observation time, e.g. the incidence of cervix cancer in Danish women in 1997 was 11.42 per 100,000 years of observation in women.

The cumulative risk is estimated by the proportion of people who contract the disease over a given time span. The estimation can be done directly in a population where follow-up is complete, e.g. the cumulative risk of stillbirth can be estimated if a given number of pregnant women can be followed from their 24th week of gestation until they give birth (if 24 weeks define the separation of abortions and births for terminated pregnancies). If complete follow-up is not possible, as for longer follow-up periods, the risk may be estimated by connecting incidence rates (IR) to risk (CI) according to the formula $CI = 1 - e^{-IR}$ (cf. Chap. I.2 of this handbook). This calculation from rates to risk requires stable incidence rates over the time period during which they are measured. Rates describe occurrence of events per time unit, like new respiratory syncytial virus (RSV) cases per 1000 observation months in children less than 1 year in a given region and a given time period. Rates are thus for populations. Risks indicate the probability of a given event in a given time period and for an average member of the population in question. In a population with complete follow-up, the risk of e.g. 2nd trimester abortions in 500 pregnant women who are followed until the start of the 3rd trimester is estimated as 0.02 if 10 of the 500 women abort in these 3 months.

Incidence rates and cumulative risks are based upon data from a population at risk, that is, a population at risk of getting the disease but without having the disease at the time when the observation starts. Prevalence proportions are estimated as the number of people with the disease in question at a given point in time divided by all in the population at that time, regardless of their disease status. If 500 women are pregnant in a population of 50,000, the pregnancy prevalence proportion is 0.01 (500/50,000). Since prevalence (P) is a function of incidence (I) and duration (D), the incidence of new pregnancies in a steady state population would be $500/8$, if the average duration of a pregnancy is set at 8 months, taking abortions into consideration, that is 62.5 pregnant women per month in the population of 50,000, or 750 per year.

Measuring reproductive failures during pregnancy is often complicated by the fact that the time of conception is unknown. When the pregnancy is planned, the time from the start of the pregnancy planning to a recognised pregnancy and to the end of the pregnancy may be known. Estimating the incidence rate of spontaneous abortions requires registration of time from conception to the abortion in question or to gestational week 24. At best, the time of conception may be identified by means of biochemical measures at a very early stage, but in most cases a pregnancy diagnosis is not established until 3–4 weeks after conception at the earliest, and then it is retrospectively estimated by means of last menstrual period data (LMP) or – later – by using growth measures based upon ultrasound examination. Observation time for calculating rates of spontaneous abortions,

therefore, often starts at different time periods in gestation, and in studies we have to take this delayed entry into consideration to obtain meaningful results when we try e.g. to identify determinants of spontaneous abortions (Baird et al. 1993).

The Measures Used to Describe Mortality

5.5.1

The World Health Organization (WHO) defines live births, foetal deaths and induced abortions in the following way:

The definition of a *live birth* is the complete expulsion or extraction from the mother of a product of human conception, irrespective of the duration of pregnancy which, after such expulsion or extraction, breathes or shows any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles whether or not the umbilical cord has been cut or the placenta is attached.

Foetal death is defined as death prior to the complete expulsion or extraction from the mother of a product of human conception, foetus and placenta, irrespective of the duration of pregnancy: the death is indicated by the fact that, after such expulsion or extraction, the foetus does not breathe or show any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles. Heartbeats are to be distinguished from transient cardiac contractions; respiration is to be distinguished from fleeting respiratory efforts or gasps.

This definition excludes induced terminations of pregnancy.

Induced termination of pregnancy is defined as the purposeful interruption of an intrauterine pregnancy with the intention other than to produce a liveborn infant, and which does not result in a live birth. This definition excludes management of prolonged retention of productions of conception following foetal death.

Induced Termination of Pregnancy Rate (Conceptions). This measure uses live births, induced terminations of pregnancy, and foetal deaths in the denominator.

$$\text{Induced Termination of Pregnancy Rate (conceptions)} = \frac{\text{Number of induced terminations occurring during a specific time period}}{\text{Number of induced terminations + live births + reported foetal deaths during the same time period}} \times 1000$$

Induced Termination of Pregnancy Rate (Population). This is the probability that women of reproductive age will have an induced termination of pregnancy within a given time period.

$$\text{Induced Termination of Pregnancy Rate (population)} = \frac{\text{Number of induced terminations occurring during a specific time period}}{\text{Female population aged 15 through 44 years}} \times 1000$$

$$\text{Foetal Death Rate} = \frac{\text{Number of foetal deaths during a specific time period}}{\text{Number of foetal deaths + number of live births during the same time period}} \times 1000$$

$$\text{Foetal Death Ratio} = \frac{\text{Number of foetal deaths during a specific time period}}{\text{Number of live births during the same time period}} \times 1000$$

Maternal mortality ratio is the number of deaths attributed to maternal conditions in a given time period divided with a number of live births during the same time period.

WHO recommends including maternal deaths that occur within 42 days of the end of the pregnancy. Some countries use other time periods (i.e., within one year).

Although international comparisons are difficult to make because of variable reporting practices, we know that wide differences in maternal mortality exist worldwide (AbouZahr 1998).

The *perinatal mortality ratio* is the number of foetal deaths (> 24 weeks of gestation) and deaths during the first 7 days of life divided by the number of stillbirths and liveborn children in the same period. Stillbirths are births of foetus that show no sign of life at births that occur after 24 weeks of gestation.

Recent results indicate that this definition should be separated into stillbirths and death during the first weeks of life. Stillbirths should furthermore be divided into death before labour and death during labour. In the past, foetal death and early death after birth often had asphyxia as the common cause. Congenital mal-

formations are now a much more common cause of death around the time of birth (Kramer et al. 2002).

Infant mortality is computed as the number of deaths, usually during the first year of life, divided by the number of live births during the same time period. Instant mortality rates vary between 5‰ to 10% in the poorest countries in the world.

Many of these measures are difficult to record in a comparable way over time and between countries. Live births are well registered in many countries, but that is not the case with stillbirths, partly because the gestational age that separates abortions from births differs between countries, partly because stillbirths do not count in population statistics (Gourbin and Masuy-Stroobant 1995).

In some countries, the threshold of 24 weeks is used to distinguish birth from an abortion. Other cut offs have been 28 weeks or 27 weeks. The complicating issue is, however, that birth with a liveborn child is a birth, regardless of the time of delivery. This may also be applied for children, who did not survive but showed signs of life. Any study that makes use of routine registration systems to identify births and where both live- and stillbirths are of interest, should make sure that the time at risk for the outcome of interest is comparable (e.g. including in the analysis only babies born from the 28th week onward, if that is the threshold for defining a stillbirth). It is often very difficult to identify the cause of death for stillbirths (Winbo et al. 1997) or the time of death, which may be of importance in a monitoring system. It has e.g. been suggested that foetal death during labour is a better indicator of the quality of obstetric care than foetal death before labour (Kiely et al. 1985).

Foetal and Infant Death

When we study mortality we usually estimate mortality rates, the number of people who die as a function of the size of the underlying population, and the period of time during which this population was under observation. Since mortality rates strongly depend upon age (and sex), we usually calculate age (and sex) specific mortality rates. The mortality rate for people of 90 years of age will be the number of 90-year-olds who die within a year divided by the number of observation years we have for 90-year-olds in that population. We will count observation time from their 90th birthday until they turn 91, die or leave the population for other reasons.

In perinatal epidemiology, age is even more important, but the problem is that age may be counted from either the time of conception or the time of birth. We expect mortality to be high shortly after conception and we also expect mortality to be high when the foetus leaves the intrauterine environment. We would prefer to present mortality as a function of observation time in the population at risk from conception, which is difficult for early foetal deaths (abortions) because we

usually do not know how many have conceived in the source population. We have to start our observation when the pregnancy is diagnosed and that often varies. Sometimes the pregnancy is first diagnosed by the occurrence of a spontaneous abortion, and sometimes a conception ends in a spontaneous abortion without the woman ever realizing that she was pregnant. Moreover, since some women bleed early in pregnancy this may cause problems in the attribution of gestational age (Gjessing et al. 1999).

An additional problem is that the timing of foetal death is often not known, and only the time of expulsion is. In early gestation these two points in time may differ by several days or even weeks (missed abortions), and some dead foetuses may even be absorbed rather than aborted. Later in gestation we expect a stillbirth to be closer to the time of death. When signs of life disappear most women in affluent societies will seek medical assistance. The foetal death will be diagnosed and an abortion induced. For multiple pregnancies where only one twin dies the situation may be different.

Some studies on foetal death use survival methods that take delayed entry and gestational age into consideration, or they use the ratio of all abortions to births as the endpoint. In any case, caution is called for. If exposures cause very early (pre-clinical) spontaneous abortions these will not be detected. Exposures that move clinical abortions to the pre-clinical stage will appear as if they prevented abortion for both measures, regardless of whether they are based upon rates or cumulative risk.

The time in which a pregnancy is diagnosed depends upon a number of known and unknown factors. It is reasonable to expect a planned pregnancy to be detected earlier than an unplanned pregnancy. On average, women with regular menstrual cycles probably become aware of the pregnancy earlier than women with irregular cycles. A woman who has been pregnant before may detect symptoms of a pregnancy earlier than a woman who is pregnant for the first time. Availability and sensitivity of pregnancy tests will also play a role for the starting point of observation.

Assume that we base our study upon a cohort of pregnant women. Assume furthermore that we take an interest in an exposure that, for some reason, correlates with how early the pregnancy is recognized. The exposed women would then enter the cohort at a later (or earlier) gestational age than the unexposed women. Since the risk of abortion decreases with gestational age, their ratio of spontaneous abortions to birth would be lower (or higher) than that of the unexposed group and the results would thus be biased. Comparison of gestational age-specific abortion rates should, however, be unbiased – provided that the exposure does not modify the time period from foetal death to expulsion.

If an exposure changes only the timing of pregnancy detection, gestational week specific abortion rates remain valid. If the exposure changes the timing of abortions around the threshold for their detection, all the possible abortion measures may be biased.

Studying spontaneous abortions may also be complicated for other reasons. Since the risk of spontaneous abortions varies largely with gestational age, it is

important to use valid data for gestational age in the model that is left truncated at the time of entry. Unfortunately, precise data on gestational ages are difficult to obtain, even if ultrasound measures were done. A foetus with a poor survival destiny may show early growth patterns that deviate from standard values, possibly making ultrasound estimates invalid. Furthermore, the exposure of interest may correlate with the validity of estimates of gestational age, which may make it impossible to obtain unbiased comparisons even when doing the proper statistical analyses. Furthermore, the start of observation time need not coincide with the time of exposure. If the exposure causes abortion after a short exposure time, the susceptible pregnancies may be removed from our study for those who had been exposed before start of observation time, leaving a selected group available for study. This selection could attenuate or eliminate an effect of the exposure on the risk of spontaneous abortion.

Using a case-control approach to study causes of spontaneous abortions may be prone to bias in some situations. Controls should (using the principles of incidence density sampling, see Chap. I.6 of this handbook) be selected at the time of foetal death (which is often unknown) and not at the time of abortion. Furthermore, hormonal measures at the time of abortion that change over gestational time may be poor indicators of the cause of foetal death, even in situations where they are the cause of death, rather than a consequence of it.

Dietary habits (such as coffee consumption during pregnancy) may change when nausea disappears, and since a foetal death would reduce nausea, cases may then have a higher intake of coffee than controls, not because coffee killed the foetus, but because nausea and aversion against coffee disappeared when the foetus died. The exposure frequency is then high in the time from death to abortion, but was not so at the time of death. The high coffee intake is a consequence of foetal death, not a cause of death.

When the child is born, an entirely new time schedule starts. Preterm or very preterm children will start this time clock before their foetal maturation has come to its natural end. Babies born in week 35 will start their extrauterine life 5 weeks before a baby born at term. Diseases that originated in utero with a fixed induction time will have an onset that perhaps should have been counted from conception time rather than the time of birth. Childhood colic is perhaps a disease that peaks at a given time from conception, independently of the time of birth (Sondergaard et al. 2000).

Measures that use births as the denominator rather than the population at risk in the proper time intervals apply a practice that deviates from the practice that normal age specific rates usually rest upon. Foetal death rates would, in normal practice, be seen as death within a given gestational time period divided by the observation time for foetuses at that gestational age, just as infant mortality is estimated as death during the first year of life divided by observation time for children less than 1 year of age in that population.

Rather than using survival principles in studying determinants of abortion, many use the ratio of spontaneous abortions to birth, or spontaneous abortions/(induced abortions + spontaneous abortions + births). The latter presents a pro-

portion of spontaneous abortions among all who terminate their pregnancy with either a birth or an abortion. The first measure will overestimate the cumulative risk of spontaneous abortions, as it does not take induced abortions into consideration (some of which would have ended in a spontaneous abortion). The latter estimate will attenuate the cumulative risk since some of the induced abortions would have contributed to the numerator had they not been induced. The frequency of induced abortion and the timing of induced abortion thus becomes a source of bias in these studies (Olsen 1984), unless data are analysed by means of survival techniques.

The risk of ending a pregnancy with a spontaneous abortion is high, especially if the mother is more than 35 years of age (Andersen et al. 2000). How high abortion rates are from the time of conception is not known because we only have data on conception for very specific *in vitro* fertilization (IVF) conceptions, which do not represent the general population. Kline et al. (1989) estimated that 50% of all conceptions end as spontaneous abortions; 40% of pregnancies that could be detected by hCG measures and 10%–15% among clinically recognized pregnancies. These figures probably do not need much adjustment today, although a new cohort study among pregnancy planners found a slightly lower abortion rate in the pre-clinic, but detectable, phase of pregnancy of about 25%–30% (Hjollund et al. 2000), close to what Wilcox found in his study of pregnancy planners (Wilcox et al. 1988). No biological measure is at present able to pick up the very early foetal life.

Many abortions have chromosomal aberrations, especially among the very early losses (Macklon et al. 2002). Chromosomal aberrations, or more specific genetic defects, may be used to perform more detailed analyses of cause-specific mortality. If all spontaneous abortions were grouped together, the measure would represent a general mortality endpoint and, since most exposures are expected to be specific in their causal action, a general mortality measure may in many situations be too imprecise for meaningful research. The problem is, however, that obtaining and genotyping foetal tissue when an abortion occurs is neither easy nor inexpensive in an epidemiologic study that often includes large numbers of participants.

When studying a specific exposure that is not believed to cause chromosomal aberrations, it would be wise to restrict the outcome to abortions without such defects, if possible. Chromosomal analysis may, in some situations, be used to distinguish between consequences of foetal death from causes of the death itself. If e.g. coffee intake is a result of foetal death rather than the cause of it, one should expect to see coffee equally associated with abortion with and without chromosomal aberrations. If coffee drinking is causally related to the foetal death, it will probably operate either via chromosomal aberrations or through another fetotoxic mechanism independently of the chromosomal aberrations. These analytical principles were to some extent used in a study by Cnattingius et al. (2000).

Some reproductive failures correlate (Basso et al. 1998a) and they often have a tendency to repeat themselves (Basso et al. 1999). It may be necessary to take this into consideration when designing a study.

Perinatal Mortality and Health Care

It is obviously desirable to reduce foetal and early childhood deaths as much as possible, at best by removing or reducing the underlying causes of death, especially if these causes also lead to long-term health problems for those who survive. Still, the main investments in most affluent societies have been spent on improving treatment. In some situations effective treatment will not only save the life of the foetus but also lead to a better potential for a normal and healthy life. In other situations, treatment may increase the incidence of diseases (by keeping alive babies (often very preterm) that may be severely impaired).

A number of mortality measures, especially perinatal mortality rates, have been used to monitor how part of the health care system performs. Perinatal mortality has declined over time in most countries and is reaching very low values in many affluent societies.

It has been suggested to stratify perinatal mortality by birth weight in order to obtain a better monitoring instrument for the quality of treatment. The idea is that advanced treatment would especially show its effect on children born with a low birth weight; the babies having the highest risk of not surviving the early extra uterine life.

When the method of stratifying mortality rates according to birth weight (and sex) became widespread, a number of so-called paradoxes appeared. The offspring of smoking mothers had lower mortality at low birth weights than the offspring of non-smoking mothers with the same low birth weight. The same was seen for populations living at high altitudes and for African Americans compared with European Americans in the US (Adams et al. 1991). These paradoxes could be seen as a result of confounding by the underlying causes of impaired foetal growth. Smoking may, for example, be a less harmful way of reducing foetal growth than whatever caused the growth retardation among those not being exposed to intrauterine tobacco smoke, although smoking in general increases perinatal mortality.

Another explanation of the paradoxes was given by Russell and Wilcox (1991) and Wilcox (2001), who showed that the strongest predictor of perinatal mortality in a population is not the proportion of newborns with a low birth weight, but the proportion of newborns with a birth weight that falls outside the population specific Gaussian birth weight distribution (the residual). Birth weight usually follows a normal distribution with a small 'bump' in the left side tail, mainly representing pre-term births. The size of the residual (usually reported in percent of the total) is a stronger predictor of the perinatal mortality for the population than the proportion of newborns with a low birth weight (< 2500 g), which is the present indicator used in monitoring systems. If we follow Allen Wilcox's thinking, the aim should not be to change the birth weight distribution for the population, but to reduce the residual portion of the distribution, i.e. the proportion falling outside the Gaussian distribution. This translates mainly into preventing pre-term birth rather than increasing birth weight. Most reproductive epidemiologists would probably agree on this strategy. The present obesity epidemic in many countries is e.g. expected to increase birth weight in the com-

ing years, but not to decrease perinatal mortality; in fact, we may observe the opposite.

Stratifying perinatal mortality according to birth weight ranks (or a z-score) for the particular distribution rather than the absolute birth weight also eliminates the paradox of crossing mortality risk in most situations, indicating that birth weight in itself may be an inappropriate indicator of mortality risk.

Birth weight has been used because it is available in most countries. Data on gestational age are less often available and the quality may be poor. Still, uncritical use of birth weights as an 'exposure' or endpoint in reproductive epidemiology is not appropriate.

5.7

Foetal Growth and Birth Weight

If a woman is exposed to agents that reduce foetal growth, growth may be reduced proportionally or disproportionately. If mainly fat tissue is reduced, as seen in smoking women, the newborn baby may have normal height but a reduced birth weight. The ponderal index is a measure that attempts to distinguish between thin and normal body proportions. The index is calculated as the newborn's weight divided by height raised to the third power. Readers, who are familiar with the Body Mass Index (BMI), will know that the ponderal index deviates from BMI only by raising height to the power of 3 rather than to the power of 2. The only reason for this difference is to obtain a more symmetrical distribution of the ponderal index in newborns. How well this index actually reflects body composition among newborn children is, however, not well known.

There are of course other anthropometric measures of interest than birth weight or birth length. Head circumference is one such measure. Abdominal circumference may also be of interest. Most of these measures are probably measured with less precision than birth weight, at least in countries that use standard weighing conditions (like time since birth) and well-calibrated electronic weights. It is also more difficult to measure the length of a baby, or a circumference. It is e.g. likely that babies born vaginally will present a molding of the cranial plates that will modify their head circumference compared to babies born with a caesarean section. Also, it is possible that some of these additional measurements would not be taken if a baby were ill at birth, so that excluding babies because not all measures were taken may produce selection bias.

Since birth weight is a function of both pregnancy duration and foetal growth, pregnancy duration is usually taken into consideration when analysing birth weight. The simplest procedure is to stratify results on preterm and term birth, but this may not fully adjust for confounding related to gestational age. Another option is to estimate small for gestational age (SGA), which implies identifying the, say, 10% with the lowest birth weight among children born during each given gestational week. Since the birth weight distribution is population specific, it is wise to use an internal reference rather than an external reference if the study is

large enough. Using SGA measures has the disadvantage that they do not make use of the birth weight distribution. The SGA term should furthermore not be taken as a measure of intrauterine Growth Retardation (IUGR) since the term SGA is purely descriptive. The term IUGR should be restricted to situations where it is known that the foetus is growth retarded. Such a measure would in principle require a documented deviance from the foetal natural growth curve.

Gestational age could also be included in the statistical model to adjust for gestational age confounding. In order to account for non-linearity in confounding, it may be preferred to include gestational age plus gestational age squared, or to include several categories of gestational age as dummy variables (for an introduction to regression models see Chap. II.3 of this handbook).

The drawback of adjusting for gestational age one way or the other (or to use SGA measures) is mainly related to the often poor quality of data for gestational age. Birth weight is measured more accurately than gestational age and, by using an endpoint like small for gestational age (SGA), good data may be turned into bad data by making use of a composite measure that includes a variable that is imprecise at best and possibly even biased. An exposure with no effect on foetal growth that causes irregularities in the menstrual cycle could show a biased effect on SGA, if gestational age was based upon LMP data, or a biased effect if gestational age is measured by ultrasound and the exposure correlates with early foetal growth.

The term 'small for gestational age' is misleading, since it is a purely descriptive population concept: the baby is among the smallest in this particular population. Usually, we would like to know if the baby is small because it is growth retarded. A baby that has achieved its full genetic growth potential could be an SGA baby just because it is genetically small. We expect such a baby to be at low risk for all complications and diseases that may be related to poor foetal growth.

Our interest in birth weight from a health perspective should focus upon a deviation from the genetic growth potential rather than on the absolute birth weight. The soundness of this approach was elegantly demonstrated by our Norwegian colleagues, who compared observed birth weights with the expected birth weight estimated from their sibling's (and mother's) birth weight. A birth weight lower than expected was the most important risk factor for perinatal mortality all over the birth weight distribution – not only for newborn children with a low birth weight (Skjaerven et al. 2000).

Unfortunately, however, we do not know the genetic programming of the foetus and we often have to rely upon indirect estimates that could lead to severe misclassification.

Given the fact that birth weight has been used for convenience – too extensively and during too long a period – a group of scientists (Adams et al. 2003) met in June 2002 in Denmark to announce the so-called Sostrup Statement (named after the residency where the meeting took place). These statements concluded that:

1. In population studies, 'percent low birth weight' (LBW) is a poor research tool for detecting factors or conditions that damage perinatal health.
2. 'Percent LBW' is a poor index of population perinatal health.

3. Adjustment for absolute birth weight is rarely justifiable in looking for effects of specific exposures on infant or perinatal mortality.
4. Some exposures or conditions may compromise foetuses without causing pre-term delivery or impaired foetal growth.

It is, however, difficult to propose another indicator than LBW that is easy to obtain and subject to a small degree of error, especially if we want such an indicator to be applicable in less affluent countries. The best alternative for a single indicator is probably preterm birth, although that may be subject to a higher degree of misclassification.

5.7.1 **Optimal Birth Weight**

The concept of an 'optimal birth weight' has been used in the literature mainly to indicate the birth weight with the lowest perinatal mortality in specific populations. We do not recommend the concept to be used broadly since an optimal birth weight depends upon what the birth weight has to 'optimise'; mortality, immune defence, cognitive development, etc. Studies on 'optimal birth weight' in the context of mortality have, nonetheless, shown interesting results. The 'optimal birth weight' is higher than the average birth weight, which may reflect a trade-off between the mother's and the child's interests (Haig 1993). The foetus will try to take as much of the nutritional supply as possible, while the mother needs to reserve some for herself to continue her (reproductive) life. According to Haig, during the course of evolution genes have been shaped to balance these two aims. Data furthermore show that the 'optimal birth weight' is population specific but is correlated with the average birth weight in the specific population (Graafmans et al. 2002). According to the evolutionary theories, it is to be expected that the 'optimal birth weight' is, in fact, a given birth weight for a given individual.

Gestational Age: Pre- and Post Term Delivery

5.8

A pregnancy of course starts at conception but, since the time of conception is usually unknown, the starting point is often taken from the first day of the last menstrual period (LMP), which is usually around 2 weeks prior to conception. A pregnancy is expected to last 40 weeks or 280 days, according to this calculation.

In most countries, Naegele's rule is still applied to estimate gestational age from using LMP data. The expected day of birth is calculated starting from the first day of LMP, then 3 months are subtracted and one week is added.

This rule works well for women who remember their menstrual periods and whose periods are regular (and close to 28 days of duration). The Naegele rule also works best in non-leap years at the population level (Basso et al. 2000a). With

the easy access to electronic calendars, one would expect to see more electronic devices that simply add 280 days to LMP, taking leap years into account.

In affluent societies, ultrasounds (US) are more and more often used as a way of estimating gestational age, even in normal pregnancies with a certain LMP date. The idea behind the estimate is that certain structures, like the biparietal diameter (BPD), grows linearly and similarly for all in the beginning of the pregnancy.

A given diameter is compared with a growth standard and the gestational week is based upon the measure read from the standard growth curve – 16 weeks in the example above (cf. Fig. 5.1).

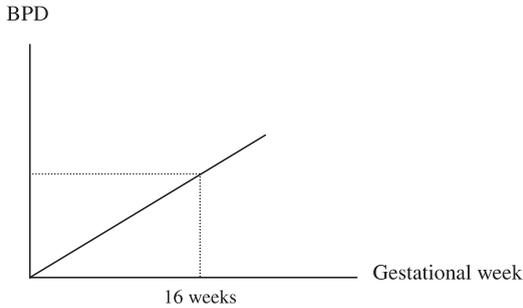


Figure 5.1. BPD according to time since conception

Experience has shown that ultrasound estimates are more precise than LMP measures and for this reason they are more precise in clinical predictions of the date of delivery. They need not always be better for research, however. If you study exposures that impair early foetal growth, then an ultrasound measure may cause bias. The bias is probably too small to be of relevance for clinical practice, but it may be of concern in research. There may be research projects that are better off with an unbiased estimate with a low precision (like LMP) than with a biased estimate with a high precision. It has been shown that smoking, for example, may impair even early foetal growth (Henriksen et al. 1995) and, if this is the case, then

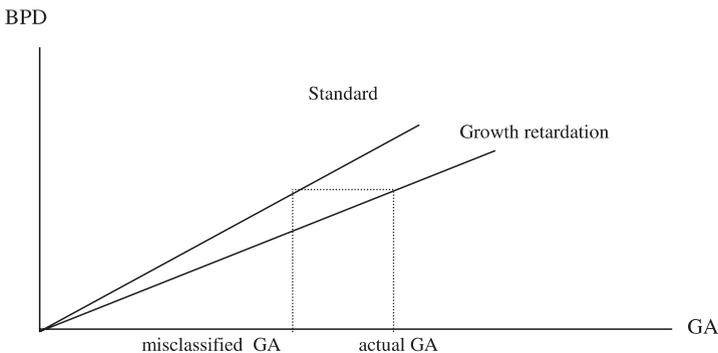


Figure 5.2. Actual and standard BPD growth curves

smokers would systematically be misclassified with an earlier gestation than their actual one because the growth function will be lower than the standard, as shown in Fig. 5.2. Assume that the difference between the actual and estimated GA is two days. If the woman gives birth shortly after 37 weeks of gestation, she would – in error – be defined as giving birth preterm.

If an inappropriate standard is used at the population level indicating a more rapid growth than for the population, gestational age will then be biased towards a low value, which will lead to a higher frequency of preterm birth. Using an inappropriate standard for the population under study will thus have impact on the estimated proportions of preterm and post-term birth for the population.

Gestational age is counted in days or in completed weeks. Preterm birth is a birth occurring before the woman has reached week 37, while post-term birth is defined as a birth taking place after completion of the 42nd week. The term prematurity was sometimes used in the past for newborns with a birth weight of less than 2500 g. This term should be avoided, because we do not know if newborns with a birth weight of less than 2500 g are premature; some will not be. Preterm is a better factual description, and sometimes it will be useful to study very preterm as well. Very preterm has been defined as birth before 34, 33, or 32 weeks (Berkowitz and Papiernik 1993). We prefer not to have a fixed definition for very preterm but to let it be defined in the study. The best definition may depend not only on the hypothesis you want to examine, but also on the available sample size.

The frequency of preterm and post-term births depends upon precision and validity of the estimates of gestational age. If the central tendency of two measures (ultrasound and LMP) are the same, but one is measured with larger measurement errors (like LMP), then the imprecise measure will lead to more pre- and post-term births, provided that the central tendency is the same. In a study of pre- or post-term delivery it is therefore important that gestational age is at least measured using the same methods in the groups to be compared. If studies are based upon routine registrations of gestational age, it may not be known how it was measured. If the exposure under study correlates with early foetal growth, LMP should be applied. If the exposure correlates with menstrual irregularities, then ultrasound is preferable. In order to detect a given difference in gestational age between two compared groups, a larger sample size is usually needed if gestational age is estimated by LMP compared with ultrasound estimates.

An example of the latter case may occur when studying whether a long time to pregnancy leads to preterm delivery. If only LMP measures are available, one may find that women with irregular or long menstrual periods have a longer TTP, as many women with subfecundity may have irregular cycles. The measure of time to pregnancy itself will be affected by this irregularity, since time to pregnancy, although sometimes reported in months, should – however – reflect the number of cycles that a couple takes to achieve a clinically detectable pregnancy. In such cases, not only imprecision but also bias can affect the effect measures, and the burden to estimate how much the observed effect can be ascribed to these problems falls on the researcher. If ultrasound measurements are not available, one way to assess whether the association may be due to bias caused by differential misclassification

of gestational age could be to use birth weight as a support measure to corroborate the finding. Since birth weight is less subject to error and its accuracy is less, if at all, dependent on correlates of the exposure, this may provide reassurance for the finding – or grounds for rejecting it. If there is no difference in the birth weight distribution according to time to pregnancy before adjustment for gestational age, this suggests that most of, or all, the observed effect is due to differential measurement error of gestational age. If, on the other hand, the birth weight of children born to women with a long time to pregnancy is lower than that of women with a short time to pregnancy, this is an indication that there is at least some effect of time to pregnancy on gestational duration, if there are no residual confounders that can reduce birth weight that are more common in women with long time to pregnancy. (This would be a problem if, for instance, short women – who have lighter babies – also had a longer time to pregnancy than tall women, but a gestation of the same duration). A long time to pregnancy may have an effect on foetal growth as well as on gestational age, and one could thus estimate this by adjusting for gestational age when examining birth weight as a function of time to pregnancy.

Another example of a factor affecting the accuracy of gestational age selectively in exposed and non-exposed women may occur when examining the effect of short interpregnancy intervals (the time between the previous birth and the next conception) on the gestational age of the subsequent pregnancy. Women who have short intervals are likely to have less accurate estimate of gestational age, because cycles resume some time after a pregnancy, and they may be irregular at first. Therefore, women with very short intervals may have systematically inaccurate measures of gestational age. Again, by checking with birth weight one can try to assess whether there is evidence of this phenomenon. Although short interpregnancy intervals have been associated with both preterm delivery and low birth weight, in a study among Danish women (Basso et al. 1998b) only the association with preterm delivery was observed, while the one with low birth weight disappeared entirely after adjustment for preterm delivery.

As stated previously, being born pre- or post-term may have stronger health impacts than deviates from a 'normal' birth weight. Using e.g. the proportion of preterm births as an indicator in a monitoring programme is therefore of interest. In the past, such a measure was based upon LMP data. Now it would often be partly LMP and partly ultrasound based, or based upon ultrasound measures only. From a monitoring point of view this raises issues of concern related to comparability over time. The standard used to estimate gestational age based upon BPD should be appropriate and that means both time- and population-specific. It should be a standard for the population it is applied to, and it should be changed over time if foetal growth in the population changes over time. Many countries face increasing obesity problems, which may influence not only birth weights but also early foetal growth. If so, BPD standards need frequent adjustments for an ultrasound-based monitoring system of preterm births to be unbiased.

No matter how gestational ages are calculated, you will sometimes find implausible gestational ages based upon the newborn children's birth weight or maturity. We prefer to code such data as missing rather than use some of the methods that

have been proposed to clean gestation age data (Parker and Schoendorf 2002). For estimating gestational age at abortion, the LMP method is to be recommended in most cases, even if ultrasound was used. A foetus that does not survive might have been severely growth-retarded.

The pregnancy period is, in many countries, under intense monitoring by health care personnel. The effects of any exposure under study reflect only the effect remaining after health care intervention. This limitation is always true, but it is especially important to keep in mind when studying pregnancy duration, especially post-term birth. A birth may be induced when the clinicians believe the child is better off outside the uterus, or perhaps just because they believe the child is sufficiently mature to be born, or if the mother's health is at risk if the pregnancy is continued. A pregnancy may then be terminated by a caesarean section, by medically induced labour, or by other means. In any case, a substantial number of pregnancies are not carried to a 'natural' end, and these observations are censored in the life table terminology. Since the censoring, in many cases, will be associated with the risk of pregnancy or birth complications, we cannot study e.g. the risk associated with post-term birth per se. The only option is to study what remains of risk after health care intervention. In like manner, one cannot study determinants of post-term delivery, only determinants for pregnancies that are allowed to continue after 42 weeks of gestation. Although these limitations are self-evident, they are often not mentioned in scientific reports – perhaps because they are self-evident. In our experience they may be self-evident but are often 'forgotten', even among experienced epidemiologists. In any case, the necessary precautionary warnings are often omitted from the discussion, and the results could thus mislead the public.

The risk associated with being born at an early or late gestational age may reflect a risk associated with gestational time itself or the underlying conditions leading to early or late birth. Only occasionally will it be possible to distinguish between these two possibilities.

Preterm birth is a frequent and strong risk factor for child health. Our ability to prevent preterm birth is limited by our sparse knowledge of avoidable causes of the condition. A number of social, environmental and dietary factors may play a role. Much of the present research effort is devoted to determining the role of infections, but it remains true that preterm birth is one of the important hazards where our preventive efforts have had limited success. In many countries, preterm birth remains a frequent determinant of perinatal morbidity and mortality. It is furthermore a frequent problem affecting 4%–6% of all births or even more. The best-known predictor of preterm birth is having had a previous preterm birth. Low social status and smoking, which are well recognized in predicting a poor birth weight outcome, are not strong determinants of preterm delivery.

There are, however, many types of preterm delivery, since births can be induced early in mothers at risk (pre-eclampsia is the most frequent cause of iatrogenic preterm birth, for example), so some authors prefer to consider only rupture of the membranes and disregard other types of preterm births, such as induced deliveries. This approach will provide a 'purer' set of cases, which might share a more homogeneous aetiology than the totality of preterm births.

Congenital Malformations

Congenital malformations have been subject to numerous epidemiologic studies and monitoring systems because they are very serious for the affected children, their families, and society in general. Researchers dealing with studies of teratogenic or fetotoxic effects prefer to include a wider range of abnormalities than structural malformations. Some prefer to use the broader term congenital anomalies (CA) that will include genetic disorders and some functional impairment as well. CA will be present in 2%–7% of all newborn children, depending upon the definition and upon the level of diagnostic procedures that have been applied. Many anomalies will not be diagnosed until childhood or even later, and a number of defects (such as some heart malformations) may go undetected for many years. Effectiveness of prenatal screening followed by induced abortion of affected foetuses also plays a role for the prevalence of congenital anomalies at birth.

If congenital anomalies are taken to be all structural or functional defects or deviations that are present at birth (whether or not they are detected at the time), their frequency could be defined to cover many more. Many functional defects may be present at birth in a form that is not yet detectable with the diagnostic means we have at present, like cognitive defects, other brain defects, mutations of importance for cancers like childhood leukaemia, or testis cancer. Foetal organ programming of organ functioning could, in principle, also be seen as congenital anomalies in the sense that programming may increase susceptibility to many diseases, like insulin resistance and all the diseases it may lead to. Clearly, using such a broad approach makes congenital abnormalities an impracticable or even impossible endpoint for studies or monitoring. Most studies and monitoring systems will restrict the endpoint to what is described in official disease classifications, such as Chap. 17 in the International Classification of Diseases (ICD10).

Monitoring systems of congenital malformations have been, and are still being, used in order to detect changes in the prevalence of malformations over time. Many of these systems have their root in the Thalidomide disaster. Thalidomide was released on the market in the 1950s to treat nausea and insomnia and was regularly used by pregnant women. The drug was teratogenic, as reported by Lenz and MacBride in 1961 (Diggle 2001), in about 40% of the pregnant women who used the drug during organogenesis (mainly in the 2nd and 3rd gestational months), and the most common defect produced at birth was phocomelia, a syndrome where the extremities were severely underdeveloped. The drug is not teratogenic in all experimental animals and, since pregnant women are never included in premarketing randomised trials, pregnant women in the population are often the first to experience potential side effects of new drugs. It is therefore important to set up programmes that fully utilise the information generated by pregnant women using new drugs. At present, we are short of such information systems (Olsen et al. 2002).

It is currently hoped, somewhat naïvely, that the monitoring systems will pick up new teratogenic drugs, even at an early phase and that such an effect will be detected by the reporting of side effects. Most drugs, especially new ones, are, however, only

used by few pregnant women, and should they cause only one (or a few) specific congenital malformations they will not be detectable in a general monitoring system or in a side effect register, since the person who prescribes the drug is usually not the same as the person who diagnoses a congenital malformation. A monitoring system of congenital malformations would, on the other hand, be of importance for setting up specific studies, because good quality data on congenital malformations are often lacking in routine medical records.

The technology available to detect congenital malformations at an early stage in gestation is continuously improving with the use of ultrasound or biochemical and genetic methods. Clearly, measuring prevalence at birth may become a questionable endpoint if prenatal diagnosis is not used in exactly the same way in the groups to be compared.

Since specific malformations are rare, most studies on determinants of congenital malformations will be based upon large routinely collected data sources covering many thousands, possibly hundreds of thousands, of pregnant women, or by using a case-control approach. The main advantage with the case-control is having the possibility to collect valid exposure information concerning the pregnancy (often very early pregnancy) by means of interviews. These data may, however, be difficult or even impossible to obtain for exposures that are difficult to recall. The recall easily leads to bias related to a lack in symmetry of the information obtained from a woman who had a child with a severe handicap when compared to the information provided by a woman who had a healthy child. Using another set of patient controls (e.g. another type of congenital malformation than the one under study) may be a possible solution, although it is not without problems. When using patient controls as a surrogate for representative source population samples, the "control" disease should neither be caused nor prevented by the exposure under study, and, since we know so little about the causes of most malformations, this might be a hazardous decision to make. However, interviewing women who all had a child with congenital malformations should render the quality of information more comparable, thus reducing the potential for bias (Swan et al. 1992). There is indication of a more accurate recall of medicine intake in mothers who had a child with a congenital malformation than in mothers, who had a healthy child (Rockenbauer et al. 2001). Furthermore, it has been shown that the way questions on medicine intake are phrased plays an important role (Mitchell et al. 1986).

Using a case-crossover design (Maclure 1991) might be an option to overcome biased reporting (and confounding by personal characteristics). Since only exposures at a given time period may be of relevance for the malformations under study, another time interval during gestation before or after the index exposure may be selected as a reference. The relative prevalence ratio for the case in question can be estimated by dividing mothers who were exposed in the exposure window but not in the reference window and vice versa, given that these windows have the same duration. The rest of the exposure combinations do not provide any information to the study concerning the associations between drug intake and the specific malformations.

A case-control approach will allow specific diagnostic classification of the cases. By setting up specific diagnostic standards it is possible to make sure that only

the individuals with the malformation in question enter as cases (high specificity) although some with the disease may not meet the criteria (sensitivity < 1). As long as specificity is high, however, a relative effect measure will not be biased by a low sensitivity, but the study power will be reduced. The following example illustrates why this is the case.

Imagine that the underlying source population has the following structure:

Exp	Cases	All births
+	2000	100,000
-	1000	100,000

$$RP = \frac{2000/100,000}{1000/100,000} = 2.0$$

Now assume that only half of the time cases will be diagnosed. The display of data would then be:

Exp	Cases	All births
+	1000	100,000
-	500	100,000

$$RP = \frac{1000/100,000}{500/100,000} = 2.0$$

The relative prevalence ratio (RP) is still 2, simply because there are still twice as many diagnosed children with malformations among the exposed compared with the non-exposed. The low diagnostic sensitivity did not change this. It is easy to design a case-control study to replicate these results. It is just a matter of proper sampling from the source population.

A case-control sampling using all cases and a sample from the base from any of the two above situations would produce an unbiased estimate of the RP. A 1 : 5 case-control sampling would, in the first case, give:

Exp	Cases	Controls
+	2000	7500
-	1000	7500
All	3000	15,000

$$OR = \frac{2000/1000}{7500/7500} = 2.0$$

and in the second case:

Exp	Cases	Controls
+	1000	3750
-	500	3750
All	1500	7500

$$\text{OR} = \frac{1000/500}{37500/3750} = 2.0$$

The power would be less in the second example, as is reflected by a larger variance. The variance of the OR is 0.0018 and 0.0035 in the first and second study, respectively.

5.10 Pregnancy Complications

5.10.1 Operational Definition

During pregnancy almost all maternal physiological systems are subjected to major changes. Cardiac output increases by 30%–50% and thus the kidneys have to filter a much higher amount of blood. The space taken up by the enlarging uterus changes the way the lungs and digestive systems work. The major changes are, however, hormonal, and the placenta produces a large amount of hormones that help maintain the pregnancy. The immune system is also affected, as pregnancy is a mildly immuno-depressed state. Also, there are indications that pregnancy requires a shift in the type of immune response from Th1 (pro-inflammatory) to Th2 (antibody-mediated), modifying the maternal type of immune response as well.

Because of these major changes, there are a number of pre-existing diseases (such as diabetes, kidney disease, affections of the thyroid, heart failure, autoimmune diseases) that may be exacerbated by pregnancy and harm the woman or the foetus. Some autoimmune diseases improve in pregnancy; others relapse or worsen, or present a cluster of onsets immediately following a pregnancy.

Pregnancy complications are defined in the Medical Subject Headings as the co-occurrence of pregnancy and a disease. The disease may have started before conception or after. This definition is rather general and not in line with what most people would think of as being a pregnancy complication. A puerperal depression is considered as triggered by the birth and not just a depression that happens to occur shortly after delivery, although it may be difficult to distinguish a depression triggered by birth from a depression that would have occurred in any case. In these paragraphs we will, therefore, deal with some aspects of complications that are only

seen during pregnancy, or which are only defined as such during a pregnancy. We will briefly deal with some common pregnancy complications (such as hyperemesis and placenta previa), reserving a particular emphasis to gestational diabetes and pre-eclampsia, the former because it is a clear example of some of the problems facing epidemiologists who deal with pregnancy complications. Pre-eclampsia is relatively common and a serious complication of pregnancy, and it is also one of the most fascinating mysteries of reproduction.

Other chronic diseases pose a risk to the mother and to the foetus. It is likely that most of them will be seen among pregnant women, but with a lower prevalence than in the general population, especially if they are severely debilitating for the woman in her reproductive years. The 'healthy pregnancy effect' is analogous to the better known 'healthy worker effect' and is due to the fact that a reasonably good health is required to conceive and carry a pregnancy to term. This "effect" need not cause bias if properly addressed in the design of the study. The 'healthy pregnancy effect' only underlines the fact that many diseases will be less frequent among parous women if these diseases interfere with actual fertility.

Methodological Challenges

5.10.2

Defining a Disease: Choosing a Cut-Point

A general problem in studying pregnancy complications is due to the fact that many such conditions are defined as an extreme of events that occur in the course of a normal pregnancy. Thus, a disease that represents an extreme value of the distribution of a given trait rather than a qualitatively different entity will be more problematic to study. In this case, the distinction between normal and pathological becomes relatively blurred, and very often the challenge faced by clinicians is that of defining a cut-point beyond which a condition is declared a disease (as in obesity, diabetes, hypertension, and pre-eclampsia) and below which the same condition is considered within the norm. It is immediately evident that definitions of this type are susceptible to many problems, as is the case with all diseases defined this way, because any arbitrary threshold will introduce some degree of misclassification, especially since pregnancy is a condition under intensive medical surveillance, which will then make pregnant women a population in which virtually the entirety of its members will be screened for severe diseases one way or the other and, in most cases, more than once in the course of pregnancy. It is well-known that even a test with a high specificity will produce a large number of false positives in a population with a low prevalence of a disease, and this leads to women being unnecessarily treated and subjected to the stress of being diagnosed with a disease they may not have. On the other hand, missing women with a given disease by moving the cut-point towards more extreme values will result in a low sensitivity that will lead to missing cases with consequences that may be very serious for the mother and the baby. The extent of these problems depends not only on the criteria for defining a disease but also on the approach adopted by the care providers, the access to prenatal care, and the frequency with which pregnant women are monitored. The more frequently women are seen and their blood glucose, blood

pressure, or proteinuria are measured, the more likely it will be that they can be wrongly classified as having a given disorder, especially if these measures fluctuate over time. Conversely, women that are not screened often or do not comply with prenatal care may be under-diagnosed in these circumstances. Factors related to monitoring, such as insurance coverage, distance from antenatal care centres, etc., can thus produce bias, especially – but not only – if risk factors for the disease are also part of the reasons why women comply less with prenatal care (smoking may, in some circumstances, fulfil these criteria). In some cases, modelling the number of ANC visits or other factors affecting the access to ANC (distance, social class, insurance plan when applicable) might provide a clue about whether a problem of this type has occurred, but this will not necessarily be sufficient to correct for the bias.

Even if women comply equally with prenatal care, problems may arise if the exposures under study correlate to some degree with the probability of being diagnosed, as health care personnel may differentially screen women according to their risk profiles.

Furthermore, the consensus on the cut-points usually changes in time and is often not even geographically homogeneous at the same point in time. Comparisons over time and between different areas thus become difficult to perform and of questionable value depending on the level of prenatal care and definition of the disorder. Often researchers do not have the crude values of what is actually measured but only the clinical diagnosis, which makes virtually every study susceptible to well founded criticism because the uncertainty and potential inaccuracy of the diagnosis may well depend on the putative risk factors under study. Random errors will also tend to dilute associations, often conspiring towards this end with the number of false positives that will be included in most case series. We are thus left with studying the phenotype that clinicians in that particular region and at a given point in time call a disease, which may not be the best classification from an etiological point of view.

ROC curves

Most readers will be familiar with the concepts of sensitivity and specificity. In the presence of a 'gold standard' diagnosis that is the 'truth' about whether a patient has a disease or not, the sensitivity of a given test is calculated as the number of subjects who test positive among the diseased divided by the total number of diseased. A test has a sensitivity of 1 when all with the disease are identified by the test. The specificity of a test, on the other hand, is calculated as the total number of subjects who test negative among the non-diseased divided by the total number of all non-diseased (cf. Chap. III.10 of this handbook). A test with a specificity of 1 will correctly identify as negatives all truly negative (no disease). There is probably no single test that will have both a sensitivity and a specificity of 1, so researchers have to live with a margin of error, even when several tests are used in a sequence or in parallel. The use of the concepts of sensitivity and specificity relies on a rather strong and often forgotten assumption: that the process

in question (the disease) is inherently dichotomous (that is, of the yes/no type). Although we often think of a disease as either present or absent and, for practical purposes, that is the way clinicians often treat them, this concept is a construct that depends on the definition of the disease, its severity and on the stage of the disease and does not necessarily reflect the underlying physiology – especially when we use a single marker to determine presence or absence of the disease. However, with this *caveat* in mind, the concepts of sensitivity and specificity can be useful. Associated with these are the concepts of positive and negative predictive values, which are the fractions that will be truly positive (or negative) given that they had tested positive (or negative). A test performed in a population with a high prevalence will, in general, perform fairly well in terms of negative and positive predictive values as long as sensitivity and specificity are high. In a large population with a low prevalence (say 1/1000), however, even a test with unrealistically high values of sensitivity and specificity for a clinical test (say, a sensitivity of 0.99 and a specificity of 0.98) will yield a large fraction of false positives among all those who test positive (with the above values, of the 21 that would be classified as positive in a population of 1000, 20 would be false positives). This misclassification will pose a serious problem to any epidemiologic study by diluting the case series.

In a situation where a given marker is distributed differently between diseased and non diseased subjects and where a cut-point is chosen to screen the population, a situation analogous to the one depicted in Fig. 5.3 will appear.

In the fictitious example given in Fig. 5.3, the population represented by the distribution on the left represents people without the disease under study, and the blood marker that is measured has a Gaussian distribution with a mean of 1.25 and a standard deviation of 0.12. The population on the right has the disease, and the same marker has a mean value of 1.55, with a standard deviation of 0.13. The two populations do not need to have the same absolute size, as what is represented in the figure are the relative proportions (in %) of each category of values. It appears from the figure that a large fraction of the diseased and of the non-diseased overlap as far as the marker value is concerned. A test that uses such a marker to discriminate between healthy and diseased subjects will either miss a large number of the diseased or include many false positives. In other situations there will be less (or more) overlap, but the same argument holds.

To establish a cut-point for a test (serological, or a blood pressure measurement) in order to define diseased and non-diseased subjects, 'Receiver Operating Curves' (ROC) has been used as a method to evaluate a test (Metz 1978). Each chosen cut-point would result in a certain proportion of subjects, who fail to be classified as diseased, being diseased (false negatives) and in a certain proportion of subjects, who would test positive, being in truth negative (false positives). To build a ROC curve the test is performed by progressively moving the cut-point from a lax one (in which a great number of false positives would be included) to a strict one (in which a high proportion of false negatives would be included; see also Chap. III.10 of this handbook). In the figure, a cut-off point of about 1.19 would produce

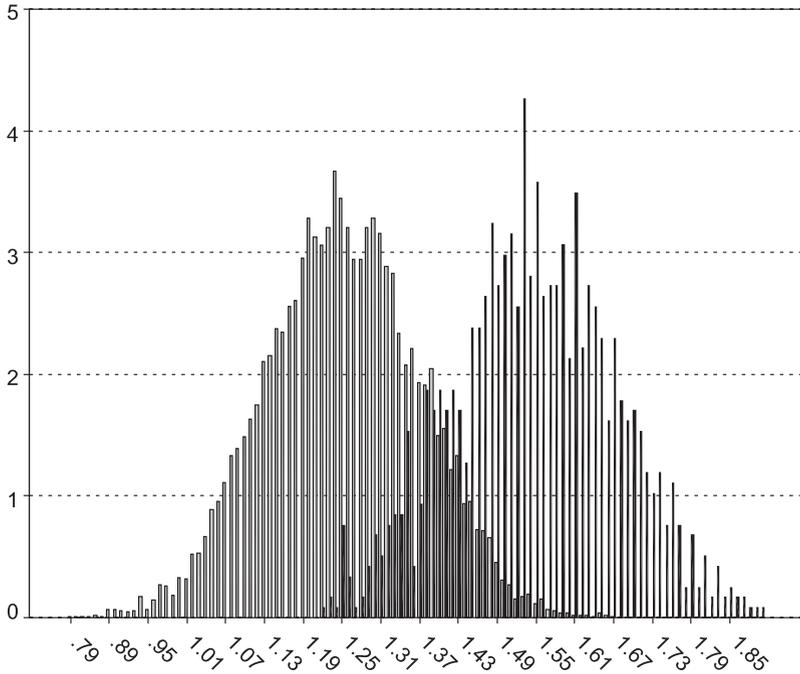


Figure 5.3. Hypothetical distribution of a blood marker in two populations, with (left) and without (right) a given illness. The area of overlap reflects the misclassification encountered when a given threshold is used to determine the disease status

a sensitivity close to 1, and a cut-off at 1.67 a specificity close to 1. When the measurement is not quantitative but is based – for instance – on radiographic results, the observer is asked to classify several times the same subject using a different rating for classifying disease: *very likely positive*|*likely positive*|*unlikely positive*|*very unlikely positive*. Of course, the ‘true status’ of the subjects has to be known with a degree of ‘certainty’ from a different source when building ROC curves. ROC curves are used to find the cut-point that maximizes benefits and minimizes side effects (taking into consideration the cost of missing a true positive and that of having a number of false positives).

The curve is built by plotting on the *x*-axis the proportion of FP (False Positives; 1-specificity) and on the *y*-axis the fraction of TP (True Positives; sensitivity).

When selecting the subjects for the experiment conducted in order to build a ROC curve, an adequate spectrum of the manifestations of the disease should be selected for the test to be applicable to the target population.

If prevalence is low, then the FP fraction must be kept low, unless it is of vital importance that all positives are identified. A combination of two tests could then be used (one very sensitive on all subjects, followed by a very specific test on the positive ones: this is applied, for example, in screening for HIV, where an Elisa test

(very sensitive) is followed by a Western Blot (highly specific)). The higher and the more to the left a curve is, the better is, in general, its performance.

In case there is no overlapping between diseased and non-diseased for the studied variable, the area underneath the curve is 1 and the curve passes through $y = 1$ and is parallel to the x -axis. When the assessment is totally random, the curve is the diagonal for the first quadrant, and the area is 0.5. An area of 0.8 means that a random diseased subject has a higher value of the test than a random non-diseased subject 80% of the time (Zweig and Campbell 1993).

Design Strategies

In general, i.e. if the exposure prevalence is low, relative effect measures, such as the relative risk or the odds ratio, are more profoundly affected by low specificity than by low sensitivity. A possible strategy when dealing with diseases that have been classified on the basis of a definition determined by exceeding a given cut-point is to limit the number of false positives by restricting the study to the more severe cases (which will, most likely, be identified with a lesser degree of error), if the possibility to do so exists. However, this will often have the consequence of substantially reducing the power of the study when dealing with disorders that, in general, are quite rare to start with.

Ad hoc studies (made for the specific purpose of investigating a given disease in a defined population) with access to medical records are an important option, which, however, will be more costly than studies based on routine information and will, once again, raise the issue of how many cases will be available for the study.

In some cases, combining the clinical diagnosis of the disease of interest with a feature that should be concomitantly present if the disease is actually present may increase the quality of the data for the study.

For example, placenta previa is a relatively rare pregnancy complication, where the placenta is wrongly positioned to partly cover the opening to the birth canal. In early pregnancy, however, the placenta will relatively often appear to be wrongly positioned but will in many cases spontaneously reposition itself during the course of pregnancy (Dashe et al. 2002). If a study is based on all diagnoses, even those made in early pregnancy, cases may include a number of these early cases, which represent a rather harmless condition. However, if placenta previa persists, the common practice is that of performing a caesarean section to deliver the baby and thus, by including only the women diagnosed with placenta previa who were later delivered by caesarean, one will probably be limiting the analysis to the more severe or, at least, the more persistent cases. This has been done, for example, in studies investigating whether pregnancies with placenta previa presented a higher male to female ratio (Wen et al. 2000; Ananth et al. 2001). Whether this type of approach should be used or not has to be evaluated depending on the specific aim of the study.

The significance of the same symptom may also change across the duration of pregnancy, and it may be a cause or a consequence of another condition also depending on the timing. This is probably the case, for example, of bleeding during

pregnancy, which, in the last trimester, is probably a consequence of placenta previa or placental abruption, while it may be an entirely different entity in the first two trimesters (such as threatened abortion). Hyperemesis, severe vomiting in pregnancy, is another example of a normal condition becoming pathological when presenting itself in an extreme fashion. Several studies investigated foetal outcome among women who had hyperemesis, with conflicting results (Depue et al. 1987; Kallen 1987; Godsey and Newman 1991; Hallak et al. 1996; Gross et al. 1989). Yet, a recent review states that severe vomiting does not have a negative effect on perinatal outcome (Eliakim et al. 2000). However, if the definition of the case series is based on a hospital diagnosis without taking severity into consideration, then the cases may include a number of women with a relatively mild disease that is not really distinct from the 'normal' nausea and vomiting of pregnancy. Restricting cases to those where some objective biomarkers, such as severe ketosis or serum electrolyte disturbance, can be measured may then provide a more purely defined case series, which might be one of the reasons for the inconsistent findings.

Timing of the disease may also be of relevance, as in two studies investigating the association between hyperemesis and female sex of the baby it was noted that the association was strong only for hospitalised cases of hyperemesis occurring in the first trimester of pregnancy (Askling et al. 1999; Basso et al. 2001). Hyperemesis occurs more frequently in women carrying multiple fetuses, as well as in women (carrying singletons) who will later be diagnosed with pre-eclampsia (Zhang and Cai 1991). These observations suggest that there might be multiple causal paths leading to hyperemesis and that, in some women, a large placenta (with a high hormone production) may be the cause of the disease, while in others hyperemesis may be a sign of some other pathological process under way. In any circumstance, whenever etiologic heterogeneity exists in a process, and signs and symptoms are the same as that of another process (so that they are considered the same disorder), it is always very difficult to disentangle what is being studied. If only a small fraction of the case population represents a different etiologic entity, even a moderately strong predictor will likely appear not to be associated with the mixed entity and it will be repeatedly missed. If the fraction is large, then a predictor will be called a risk factor for both entities, and perhaps this uncertainty contributes to some of the failures of epidemiology to encourage changes in people's habits and in policies. When these conditions occur in pregnancy, it may, in some cases, be possible to discriminate to some extent between different entities by paying particular attention to the timing in which events occur.

5.10.3 Gestational Diabetes

As many other metabolic functions, the metabolism of carbohydrates is altered in normal pregnancies. The fasting blood glucose level decreases early in pregnancy until the 12th week, and it usually remains at this lower level until the end of pregnancy. Insulin, by contrast, remains stable during the 1st and 2nd trimesters, but increases during the 3rd. Outside pregnancy, on the other hand, the blood glucose level returns rapidly to fasting levels after a meal, and in pregnancy both glucose

and insulin levels reach higher peaks than they would after a similar meal in the non-pregnant state. This level is furthermore maintained for a longer time. Human placental lactogen (hPL) is the hormone that is believed to be responsible for these changes in the metabolism (Haig 1993). In general, pregnancy is a state of mild insulin resistance, and some women develop gestational diabetes. As an adaptation, insulin production is increased at the same time that the mother is becoming insulin-resistant. Haig (1993) proposes an interesting hypothesis about why this change may occur within his evolutionary theory about the genetic conflicts of pregnancy. Glucose is an important nutrient for the growing foetus, but it is also important for the mother's survival that not too many of her resources are depleted by the foetus, which may happen if foetal demands went unopposed: the decline in blood glucose early in pregnancy could thus represent a maternal attempt to limit foetal uptake. In addition, after every meal there will be a competition between the mother and the foetus over the respective share. The longer it takes the mother to reduce her blood sugar, the higher the share taken by the foetus, hence the insulin resistance, according to Haig. At the beginning of pregnancy, the foetus has limited growing demands and limited 'power', which is why the mother succeeds in 'hiding' her blood glucose. During the third trimester, however, the foetus has very high growing demands and it is strong enough to take the upper hand in the competition with the mother. When seen in the light of how to optimise survival probabilities, this is an attractive hypothesis, although hard to test.

Barker (1995, 1998) has a different hypothesis, suggesting that insulin resistance is a consequence of limited nutrient supply. Clearly, however, glucose metabolism is crucial in pregnancy, and a mechanism has evolved that creates a delicate balance between the maternal and foetal needs. Diabetes during pregnancy is a complex problem and requires a very careful management to prevent damages to the foetus and to the mother. The consequences of gestational diabetes may be dire for the foetus; as with gestational diabetes there is an increased risk of stillbirth and macrosomia (Schmidt et al. 2001; Johnstone et al. 1990). Macrosomia is the most frequent outcome in diabetic mothers, and this can complicate delivery to the point that a caesarean section is required. Women who have had gestational diabetes are at increased risk of having it again in a subsequent pregnancy and are also at risk of developing diabetes (Dornhorst and Beard 1993), especially type 2, later in life. Gestational diabetes, as well as other types of diabetes, is also a risk factor for pre-eclampsia (Schmidt et al. 2001), another potentially severe pregnancy complication. Obesity is a predisposing factor (with insulin resistance the underlying condition), as is advanced maternal age or having a family history of diabetes.

The definition of gestational diabetes is problematic and the subject of many a controversy (Dornhorst and Beard 1993; Martin et al. 1995; Gabbe 1998; Bonomo et al. 1998; Schmidt et al. 2000). The discussion extends to whether there should be universal screening for all pregnant women. In 1996, the American Diabetes Association concluded that universal screening should be done, but these recommendations were then revised in 1997 and 1998 to selective screening of women satisfying at least one criterion among (1) age above 25, (2) age below 25 and a body mass index above 27, (3) a family history of diabetes, and (4) belonging to ethnic

groups with a known higher predisposition to diabetes. There is no agreement over the 'gold standard' test for women who are screened positive. The oral glucose tolerance test is the norm, but the load (the dose of administered glucose) varies geographically. There is also discussion about which cut-off limits should be used, as the prevalence of gestational diabetes would change depending upon the threshold, and the risk would be to either define too many women who are not diabetic as such or to miss too many women who are diabetic and at risk of having an adverse pregnancy or health outcome. There is no clear distinction between normal and pathological values, which hampers any diagnostic criterion. Often, epidemiologists do not deal with the actual values of glucose level but with the clinical diagnosis collected over several hospitals, without guarantee of uniformity in the criteria used for screening and diagnosis. The diagnosis will also depend on whether glucose is measured or not and, since gestational diabetes is mostly asymptomatic, this is an added complication. Researchers planning studies requiring an accurate diagnosis of gestational diabetes should thus be aware of the medical attitude towards screening in pregnancy in the locations where they plan to collect their data and of the tools and cut-off levels in use, as well as of the criteria that govern which women are screened and which are not. At best, studies are based upon follow-up of cohorts that are all subject to testing within the same protocol.

Geographical variations in the definition and incidence, as well as changes in definition and screening attitudes over time are also to be taken into consideration when making comparisons between places and periods. Long-term as well as short term consequences for the foetus have been identified, and in some countries the focus of the diagnosis of gestational diabetes has now shifted from the likelihood of progression to later chronic diabetes in the mother to the outcome of pregnancy; a shift that also has consequences on the diagnostic criteria. Since obesity is associated with a highly increased risk of type-2 diabetes as well as gestational diabetes, a number of women would only be diagnosed during pregnancy and the two types of diabetes would then be confused. However, since both types of diabetes increase the risk of pregnancy complications (such as pre-eclampsia) and of adverse foetal outcome, this may not be a major problem, depending on the specific purpose of the study. When studying pre-eclampsia, for instance, there are situations where it would be advisable to exclude women with pre-existing diabetes but not those with gestational diabetes (which shares with pre-eclampsia obesity as a risk factor and, possibly, other predictors) and this may prove to be difficult to do accurately. If and how much of an impact this could have on the estimates will once again depend on the specific situation and will in many cases be hard to evaluate.

5.10.4 **Pregnancy-Induced Hypertension and Pre-Eclampsia**

Definition and Diagnosis

In the first trimester of pregnancy, blood pressure is usually reduced from normal values. In many women, however, blood pressure increases around mid-pregnancy

to above normal values. A modest degree of hypertension is thus a rather common condition of pregnancy and has not been consistently associated with unfavourable outcomes.

Pre-eclampsia is, on the other hand, one of the most common and potentially severe complications of pregnancy. In pre-eclampsia, the maternal blood pressure can increase dramatically and heart, brain, and kidneys may be severely damaged. If the mother survives, the affected organs usually return to normality shortly after delivery, but pre-eclampsia can also be fatal. If seizures occur, the disease is called eclampsia (a very rare occurrence in countries with well-functioning health care systems) and the risk for both the mother and the foetus is then much higher. While eclampsia is a dramatic event that is probably rarely misclassified, pre-eclampsia is, by definition, much more elusive. In most countries, it is currently defined as the concurrent presence of hypertension and proteinuria. Gestational hypertension is defined as either a persistent rise of 25 mm Hg in systolic blood pressure during pregnancy compared to pre-pregnant values (if the pre-pregnant values are not known, a systolic blood pressure of 140 mm Hg or higher), or as a rise of 15 in diastolic blood pressure (DBP) (or as a DBP above 90). The definitions of gestational hypertension do, however, vary geographically. For the disorder to be called pre-eclampsia hypertension must be accompanied by a certain minimum level of proteinuria. The degree of severity depends on the values of blood pressure and the amount of protein loss, as well as additional signs and symptoms, often including oedema. Previously, pre-eclampsia was defined by the concomitant presence of two out of three symptoms (hypertension, proteinuria, and oedema), but the definition has changed to be restricted to cases where both hypertension and proteinuria are present at the same time, as oedema appeared to be too unspecific. The problem with this definition is, however, that a mild state of hypertension is common in pregnancy and, often, the pre-pregnant values are not known. What is termed moderate pre-eclampsia may thus, in some cases, be nothing more than a change in values of blood pressure and proteinuria within the norm. Sometimes, changes in these values may be severe enough to qualify for the diagnosis, but they will escape detection. Furthermore, some women become nervous when their blood pressure is taken in a clinical setting, and thus they would be classified as hypertensive while they are not ('white coat' hypertension). On the other hand, a number of cases may be missed by applying this definition, either because of ignorance of the baseline pre-pregnancy values, or because the women do not have their blood pressure measured at the moment of the increase and, if there are no severe symptoms, the diagnosis will never be made. In severe cases, women become very sick and there is little doubt about the diagnosis, but these cases are the minority.

The reported incidence of pre-eclampsia appears to vary widely across places, from an estimated 2% to approximately 8%. This variation may reflect real variations in susceptibility and determinants across populations, but it almost certainly also depends on the sources of information for the diagnosis as well as on the access to prenatal care and the problems mentioned above.

The only known 'cure' for pre-eclampsia is to end the pregnancy, as the placenta appears to be the organ that causes the disease, and pre-eclampsia is therefore the major cause of iatrogenic preterm delivery.

Pathophysiology of Pre-Eclampsia

In normal pregnancy, the maternal spiral arteries are modified and penetrate deeply into the decidua (first trophoblastic invasion) and, around the 16th to 18th week, into the myometrium (second trophoblastic invasion). The invasive trophoblast enlarges the vessels from within, and a fibrin substance that renders them flaccid and unresponsive to maternal vasoconstriction replaces the vessels' internal lining. In pre-eclampsia often, but not always, the second trophoblastic invasion does not occur, or occurs only to a very modest degree (Salas 1999; Roberts and Lain 2002), resulting in placental perfusion being severely compromised because the arteries are narrow and with a high resistance instead of being wide, low-resistance vessels, as they would be if the invasion had proceeded normally.

Haig (1993) expresses the view that hypertension in pregnancy is a foetal adaptive mechanism. Because of the structure of the spiral arteries and the fact that sympathetic nerves disappear from the placental site during pregnancy, the maternal control of the blood flow to the placenta is limited and the placental site is characterized by low resistance to blood flow. Due to these characteristics, for any given resistance of the placental unit, a compensatory rise in the maternal peripheral blood pressure will increase the blood flow to the placenta. According to Haig, this may be a sign of a feto-maternal conflict, where the growing foetus is able – by some unknown mechanism – to increase maternal blood pressure and thus increase the placental blood flow. Drug-induced reduction of mean arterial pressure may be associated with a reduction in foetal growth (von Dadelszen et al. 2000)

Pre-eclampsia is not, however, always accompanied by defective placentation, and is, most likely, a common syndrome resulting from heterogeneous causes (Ness and Roberts 1996). It is believed that large placental mass (as seen in multiple pregnancies) and endothelial disease (as seen in diabetics) are mechanisms that can also produce placental hypoperfusion and start the cascade of events that leads to pre-eclampsia (Salas 1999).

Known Predictors of Pre-Eclampsia

The aetiology of pre-eclampsia is mostly unknown, and this disorder is one of the most tantalizing mysteries of reproductive epidemiology. The best-known predictors are nulliparity, obesity, and multi-foetal pregnancies, while smoking is protective for reasons unknown, although several hypotheses have been raised to explain this association (Condé-Agudelo et al. 1999). Africans and African-Americans appear to be at a higher risk, possibly because susceptibility to pre-eclampsia is related to susceptibility to cardiovascular disease (Roberts and Lain 2002). Recent evidence suggests that only women giving birth with pre-eclampsia preterm are at a higher risk of later death for cardiovascular disease, while women

giving birth with pre-eclampsia at term have no increased risk compared to non pre-eclamptic women (Irgens et al. 2001).

Several trials have addressed the association between pre-eclampsia and dietary factors, mostly calcium, magnesium, antioxidants, and fish oil. Unfortunately, no clear answer has emerged from these studies, except for the finding that calcium appears to be protective among women with a very low baseline intake or for women with a very high risk of pre-eclampsia (Villar and Belizan 2000). In general, however, the attempts to prevent pre-eclampsia through dietary supplements or aspirin have been overall disappointing (Sibai 1998; Dekker and Sibai 2001).

It is well accepted that a genetic component to pre-eclampsia exists, since children born of pre-eclamptic pregnancies are themselves at a higher risk of having children born of pregnancies with pre-eclampsia (Esplin et al. 2001). Also, males whose partner had pre-eclampsia have almost twice the risk of having their subsequent partner developing pre-eclampsia compared to males whose partner had not developed it (Lie et al. 1998).

A large number of biomarkers and genetic factors have also been explored as predisposing to pre-eclampsia (Broughton Pipkin 1999; Roberts and Cooper 2001). Genetic studies on pre-eclampsia have not consistently revealed a specific genotype associated with pre-eclampsia, although women with pre-eclampsia are more likely to have a heterozygous factor V Leiden mutation and other thrombophilic mutations (Alfirevic et al. 2002). Not all researchers agree on the role of thrombophilic mutations (Livingston et al. 2001), however.

Methodological Challenges in Studies of Pre-Eclampsia

Although pre-eclampsia is probably the most studied among all pregnancy complications and keeps fascinating researchers from many areas of medicine, several difficulties face the investigators, mainly because of the difficulty of accurately identifying cases in sufficient numbers, without incurring selection problems. Research based upon nationwide hospital registries can provide population-based data that may, however, be of limited quality if the only available information is the code according to the International Classification of Diseases. The advantages of these studies are that women are most likely unselected and that the numbers will be large enough to allow studying even relatively rare predictors or outcomes. In some cases, these studies might be the only viable option. If, to improve the quality of the data, researchers restrict their case series to severe pre-eclampsia only, then the numbers will be dramatically reduced but, possibly, fewer false positives will be included.

On the other hand, studies of the case-control type where medical charts are available would provide a much better case series if proper diagnostic procedures can be applied to document the disease, whereas problems may exist in recruiting a sufficient number of controls retaining a sufficient confidence that self-selection will not bias the study. If the women who accept to enter the study as controls do so according to the exposure under study, this will produce biased estimates to an extent that is often impossible to judge. Since pregnant women are invited to lead

a healthy lifestyle for the sake of the baby if not their own, it is likely that some women whose pregnancy went well but whose habits were not beyond reproach would be relatively unwilling to take part in a study where such habits would be questioned. On the other hand, women who had a negative experience may be less reluctant to having their behaviour under scrutiny, because they want to know what went wrong. However, problems in studying pre-eclampsia go well beyond the objective difficulties of appropriately defining cases or of finding unselected study populations.

Pre-eclampsia is a cause of preterm delivery, mostly iatrogenic. This fact complicates the interpretation of studies attempting to evaluate whether pre-eclampsia is associated with conditions that are more common in babies that are born preterm, such as e.g. cerebral palsy. Some studies have reported that babies of pre-eclamptic pregnancies were protected from cerebral palsy when the risk was examined by gestational week at birth (Gray et al. 1998; Murphy et al. 1995). Is this a protection conferred by pre-eclampsia, or is it an artefact due to the fact that the causes of preterm delivery in pre-eclamptic pregnancies differ from those of other preterm deliveries, where the causal factors may include those of cerebral palsy? Because babies born preterm for causes other than pre-eclampsia have other pathological mechanisms that advanced birth, disentangling the effects of preterm birth from its causes is a major challenge, and so is examining the various causal paths leading to preterm birth that may very well be implicated in the diseases 'associated' with preterm birth. Also, many cases of pre-eclampsia are delivered by emergency caesarean section, and the delivery complications due to caesarean section are generally different from those arising from vaginal deliveries, as many preterm deliveries will be. If complications that can arise from vaginal delivery are associated with cerebral palsy (e.g. anoxia), then comparing pre-eclamptic women with non pre-eclamptic will result in a biased comparison.

Another problem is that of studying pre-eclampsia in connection with other conditions or factors that are associated with preterm delivery. If a given factor causes preterm delivery, it may also appear to be protective of pre-eclampsia simply because women with a shortened pregnancy have had less opportunity of developing it, since pre-eclampsia often occurs after the 36th week of gestation, but being pregnant (or just delivered) is a necessary condition for being diagnosed. If the date when pre-eclampsia was first diagnosed is known, data may be analysed through Cox regression or survival methods to overcome this problem (for a general introduction to survival analysis see Chap. II.4 of this handbook).

If an important confounder is systematically omitted when studying a given disease, this will lead to the potential establishment of a wrong conclusion (cf. Chap. I.9 of this handbook). For example, a currently accepted hypothesis about the aetiology of pre-eclampsia proposes that a maternal immune reaction to paternal antigens may be a cause of the failed trophoblastic invasion. This hypothesis was mainly based on the observation that pre-eclampsia is more frequent in first pregnancies. Among multiparous, women who had changed their partner from the previous pregnancy had an increased risk (Dekker et al. 1998; Dekker 1999; Trupin et al. 1996; Lie et al. 1998; Li and Wi 2000). Also, women with a long

period of sexual cohabitation prior to a pregnancy and women using oral contraceptives appeared to be at a lower risk of the disease than women with a short cohabitation period or those using barrier contraceptive methods (Robillard et al. 1994; Dekker et al. 1998). This suggested that a prolonged exposure to the partner's sperm may have constituted a protection that would reduce the risk of pre-eclampsia, thus leading to the expression that 'primipaternity' was a risk factor for pre-eclampsia. For some researchers (including the authors of this chapter), however, this hypothesis has lost some attraction since three studies (two based on the Norwegian Birth Registry (Skjaerven et al. 2002; Trogstad et al. 2001), and one based on a sample from the Danish Birth Registry (Basso et al. 2001)) independently reported that this increased risk with change of partner disappeared if the interval between births was adjusted for. Women who change partner have, on average, a much longer interval between births: if any factor correlated with time has an impact on the risk of pre-eclampsia, then women waiting a longer time will have an increased risk, regardless of whether they change partner. This was found to be true in the above-mentioned studies, even after maternal age was controlled for. This finding prompted a further study where time to pregnancy was investigated in association with pre-eclampsia, as a fraction of the women waiting a long time between pregnancies may be subfecund. Time to pregnancy, as previously mentioned, is a proxy for the couple's fecundity and thus a relatively crude marker, since it reflects a multitude of disorders. It is, however, interesting that an association between long time to pregnancy and pre-eclampsia could be found despite these limitations (Basso et al. 2003), and this evidence may lead to further research for identifying a subgroup of infertile women with a specific disorder that relates to pre-eclampsia.

Pre-eclampsia is most likely the result of an interaction between the maternal and the foetal systems, but its diagnosis relies exclusively on symptoms that are observed in the mother. It is perhaps for this reason that, so far, pre-eclampsia has eluded most attempts to clarify its aetiology.

Delivery Complications

Delivery complications may arise before or during delivery and present a risk for the mother and/or the baby. Some have to do with the foetus's presentation or its inability to pass through the birth canal. Until less than a century ago impacted births were the major cause of foetal and maternal morbidity and mortality. Because of malnutrition, many women had under-developed pelvises and the baby's head would remain trapped in the birth canal. Nowadays, foetal or maternal death because of this is a very rare event in industrialized countries but still a major problem in developing countries, where most babies are delivered at home and hospitals are far away and are perhaps lacking adequate resources. The three major causes of maternal mortality in developing countries are haemorrhage and sepsis,

as well as hypertensive disorders, and the first two usually result from delivery complications.

Beyond foetal presentation (and position), and foeto-pelvic disproportion, delivery complications also include weak contractions, prolonged labour (of any of the three stages), prolapse of the umbilical cord, perineal or vaginal tears, foetal asphyxia, retention of the placenta, haemorrhage, etc. Caesarean sections, which account for between 15% and 30% of all births, depending on countries, constitute perhaps the major difficulties when studying delivery complications. Caesarean sections can be planned or acute, and the latter could be started before delivery or during delivery. Emergency caesarean sections, themselves considered a 'delivery complication', are triggered by complications arising in the mother or the foetus.

In the case of delivery complications even more than in other cases, researchers have to study what is left after physicians have acted. Therefore, only babies being born vaginally will be at risk of getting the umbilical cord wrapped around their neck, or of having any other accident during their descent through the birth canal that may affect the supply of oxygen to the brain. This would not be a problem if the decision of delivering a woman by caesarean section were independent of any factors that may put the baby at higher risk of encountering such mishaps, but – usually – this is not the case. Since the relative size of the mother and the baby or signs of foetal distress may well trigger the decision of performing a caesarean section, it is likely that babies born vaginally and those born by caesarean are not comparable before delivery, which will complicate any interpretation of findings associated with a given delivery complication. This will also complicate any study trying to evaluate the 'effects' of any given intervention during delivery, as it will be difficult to separate the effects of the intervention from the causes that provoked it, which may also be the causes of the outcome of interest. Even restricting to planned caesareans may not be sufficient to solve the problem, as caesareans are planned for a reason, and one likely reason is that a complication is foreseen and a caesarean section may prevent it from occurring. Experience in previous pregnancies will also affect the mode of delivery. A woman who has previously had a caesarean section will, in many cases, have one also for her next delivery, especially if the two pregnancies are close in time. Since many events tend to repeat themselves in one woman's reproductive life, it will be hard to decide what to do with such a woman, especially if the cause for her previous caesarean is not known.

A caesarean section is the preferred choice of delivery mode for an increasing number of women and immediate risks appear to be few. Only little is, however, known about long-term effects and recent results suggest that the risk of asthma may be increased (Kero et al. 2002), although these findings have to be examined cautiously.

Obstetric complications have been associated with schizophrenia (Geddes and Lawrie 1995; Verdoux et al. 1997), and hypoxia correlates with cerebral palsy (Blair and Stanley 1993). Anoxia or hypoxia will most likely cause cerebral damage and it is reasonable to assume a causal link. But it may also be argued that a baby who had brain damage (which will later cause cerebral palsy) prior to birth will be more likely to have a complicated delivery and suffer from anoxia.

In general, if one wishes to study a delivery complication that may, in some cases, lead to a caesarean section (or to induction of delivery), it will be necessary to have information on why the caesarean section was performed. Practice of caesarean section, induction of delivery, instrumental birth, etc., change between geographical areas and in time, and dealing with these variations may prove a daunting and perhaps impossible task. Usually, many elements are used in the decision to treat, and it may just be impossible to identify all these elements and control for them in the analysis. Confounding by indication is one of the strongest arguments for evaluating treatments in randomised trials, which, however, will often be difficult to carry out in this context.

Since the practice of inducing birth (also by other means than a caesarean section) is now widespread, with criteria for induction that often differ from one hospital to the other, it will always be complicated to study either the induction itself or phenomena such as post-term delivery or macrosomia, even when good information about the causes of the induction are present.

If preterm babies are at a higher risk of incurring delivery complications, it may be this latter fact rather than the timing of birth that makes them at higher risk of several diseases. On the other hand, if some babies who are born preterm are born preterm because of some damage that will later cause the disease and makes them at higher risk of delivery complications, then delivery complications will spuriously appear to cause the disease.

Foetal Origins of Adult Diseases

Most reproductive epidemiology has been related to the time period from pregnancy planning to the early time period of a new life. In the future, many diseases will be seen as trajectories that start at the time of conception during pregnancy or in early childhood. Obviously, studying exposures with an induction and latency time of causation spanning several decades raises severe problems of being able to control for intervening factors. Without longitudinal readings of the occurrence of possible confounding factors, such studies may often provide confounded results.

It is not unexpected (or not even questionable) news that exposures during foetal life may have lasting effects. What is new is that prenatal exposures may cause diseases that manifest themselves long after birth, perhaps even as late as in following generations. Foetal programming is the name that has been used to describe what could happen if a stimulus or insult at a critical period of organ development interferes with cell division and thus with the function of the affected organ. Permanent changes of organ function could, in principle, lead to many diseases (Olsen 2000), but best documented are the associations between origins of disproportional foetal growth and cardiovascular diseases, perhaps through insulin resistance (Barker 1994, 1995, 1998).

Although the brain growth seems to be less vulnerable to undernutrition, specific nutritional factors, stress, medicine, etc., may influence brain function. Study-

ing determinants of brain function and brain pathology in foetal life should be a high priority research area.

Hormonal factors during foetal life may not only affect the reproductive organs but could also be associated with other diseases, such as cancer of the breast, prostate, and testis. As early as 1990, Trichopoulos suggested that breast cancer might originate in utero. It was suggested that oestrogen could play a role and, since oestrogen correlates with foetal growth, it is expected that rapid foetal growth could be associated with a higher risk of breast cancer five or more decades later. The hypothesis has, to some extent, been corroborated (McCormack et al. 2003).

5.13

Sources of Data

As in other subgroups of epidemiology, the data come from different sources; secondary routine data or ad hoc data based upon self-reported information, information from clinical measures, or information extracted from biological samples; blood, urine, placenta tissue, etc. (Longnecker et al. 2001).

More secondary data are available in reproductive health than in most other epidemiologic areas. Most pregnant women and most newborn children are carefully monitored and data are stored in medical records or even in computerized birth registers that may include not only birth data, but also exposure data such as smoking, medical treatment, etc. (Ericson et al. 1999).

The data that usually have to be collected for research are data that describe putative causes of reproductive failures. Many of these exposures have to be collected prospectively, since they are often forgotten and cannot be reconstructed in an unbiased way back in time, once the outcome of the pregnancy is known. This is often true for dietary factors, medical treatments, occupational exposures, etc. Data on infections, occupational exposures, life-style factors, dietary factors etc., cannot be recalled for more than a few weeks or perhaps months. Usually, the mother is, not unexpectedly, a better source for data on pregnancy and birth than the father (Coughlin et al. 1998).

For these and other reasons, it seems justified to set up large cohort studies, starting shortly after conception and with the aim of collecting exposure information during pregnancy. It is also necessary to establish cohorts that can be followed over long time periods, the best case scenario being from conception to death, including information on their offspring. These cohorts need to be large to provide sufficient information for rare outcomes. Large cohorts of this type were set up in the past, and the best known is probably the National Collaborative Perinatal Project from the USA, started in the late 1950s, where more than 50,000 pregnant women were enrolled. The cohort aimed at studying obstetrical complications and the risk of cerebral palsy and other neurological disorders, although the cohort has served many other research purposes since then.

The Danish National Birth Cohort (DNBC) enrolled 100,000 pregnant women from 1996 to 2002 (Olsen et al. 2001) and included data from interviews, registers,

and self-administered questionnaires together with blood from the mother and child stored in a biobank (www.bsmb.dk). Similar studies are in progress in Norway, the USA, and several other countries.

Conclusions

In this chapter we have tried to highlight some of the main features of reproductive epidemiology as we see them. In particular, the fact that reproduction, even in our medicalised world, is a direct result of selective processes, which are still active. In addition, most of the events that we study in this area are what is left after main selection has taken place: selection of couples where a conception takes place, and further selection of those conceptions that will progress to clinical recognition and medical intervention, which may lead to anticipated delivery or a termination of pregnancy, or to treatment of a disorder. For this reason, denominators are usually unknown. Furthermore, the processes that occur during a pregnancy that ends in a birth are usually also hidden, and therefore we do not really know what has happened to the foetus during the most delicate phases of development.

Any event in reproduction generally concerns three individuals rather than one. In many instances pregnancies are voluntary events and many women have several, although the decision to have further pregnancies (and their outcomes) often depends to some extent on the outcome of the previous ones. Time is also of crucial importance when dealing with reproductive epidemiology, but its dimension is generally different from the time involved in the development of, say, cancer after exposure to a given mutagenic substance. The types of bias that can occur in this area are, in many cases, peculiar to this discipline, and they have to be taken into consideration.

We expect genetic and functional genetic studies to be important in future research. Although we may not be able to answer the big questions like how the entire process of organ development is coordinated, finding answers to less broad questions will also be of interest. We need to know much more about genetic and gene-environment interaction, not only in the development of congenital malformations and childhood cancers, but also for long-term organ programming (for an introduction to genetic epidemiology see Chap. III.7 of this handbook). Using information on genetic factors in e.g. metabolism of environmental exposures, like alcohol, may even be of help in examining confounding. How much of the association between e.g. alcohol and reproductive failures is due to confounding cannot be examined in a randomized trial, but the genetic factors that modify alcohol metabolism may follow 'Mendelian randomization' and thus provide a design for comparison that bypasses some of the problems associated with the intercorrelation between lifestyle factors (Smith and Ebrahim 2003).

The peculiarities of reproductive epidemiology offer a number of opportunities to researchers willing to exploit them. We have tried to introduce readers to

a number of the features that make this area of epidemiology exciting, vibrating, and fairly unique.

References

- AbouZahr C (1998) Maternal mortality overview. In: Murray CJL, Lopez AD (eds) *Health dimensions of sex and reproduction*. Harvard University Press, Boston, USA, p 144
- Adams M, Nybo Andersen AM, Andersen PK, Haig D, Henriksen TB, Hertz-Picciotto I, Lie RT, Olsen J, Skjaerven R, Wilcox A (2003) Sostrup statement on low birth weight (LBW). *Int J Epidemiol* 32(5):884–885. (Letter to the editor)
- Adams MM, Berg CJ, Rhodes PH, McCarthy BJ (1991) Another look at the black-white gap in gestation-specific perinatal mortality. *Int J Epidemiol* 20(4):950–957
- Alderete E, Eskenazi B, Sholtz R (1995) Effect of cigarette smoking and coffee drinking on time to conception. *Epidemiology* 6(4):403–408
- Alfirevic Z, Roberts D, Martlew V (2002) How strong is the association between maternal thrombophilia and adverse pregnancy outcome? A systematic review. *Eur J Obstet Gynecol Reprod Biol* 101(1):6–14
- Ananth CV, Demissie K, Smulian JC, Vintzileos AM (2001) Relationship among placenta previa, fetal growth restriction, and preterm delivery: a population-based study. *Obstet Gynecol* 98(2):299–306
- Andersen AM, Vastrup P, Wohlfahrt J, Andersen PK, Olsen J, Melbye M (2002) Fever in pregnancy and risk of fetal death: a cohort study. *Lancet* 360(9345):1552–1556
- Askling J, Erlandsson G, Kaijser M, Akre O, Ekblom A (1999) Sickness in pregnancy and sex of child. *Lancet* 354(9195):2053
- Baird DD, Ragan BN, Wilcox AJ, Weinberg CR (1993) The relationship between reduced fecundability and subsequent foetal loss. In: Gray R, Leridon H, Spira A (eds) *Biomedical and demographic determinants of reproduction*. Clarendon Press, Oxford, pp 329–341
- Baird DD, Weinberg CR, Schwingl P, Wilcox AJ (1994) Selection bias associated with contraceptive practice in time-to-pregnancy studies. *Ann N Y Acad Sci* 709:156–164
- Baird DD, Wilcox AJ (1985) Cigarette smoking associated with delayed conception. *JAMA* 253(20):2979–2983
- Baird DD, Wilcox AJ, Weinberg CR (1986) Use of time to pregnancy to study environmental exposures. *Am J Epidemiol* 124(3):470–480
- Barker DJ (1994) *Mothers, babies and disease in later life*. BMJ publishing Group, London
- Barker DJ (1995) Fetal origins of coronary heart disease. *BMJ* 311(6998):171–174
- Barker DJ (1998) *Mothers, babies and health later in life*, 2nd edn. Churchill Livingstone, Edinburgh

- Basso O, Christensen K, Olsen J (2001) Higher risk of pre-eclampsia after change of partner. An effect of longer interpregnancy intervals? *Epidemiology* 12(6):624-629
- Basso O, Fonager K, Olsen J (2000a) Are pregnancies shorter in leap years? *Epidemiology* 11(6):736-737
- Basso O, Juul S, Olsen J (2000b) Time to pregnancy as a correlate of fecundity: differential persistence in trying to become pregnant as a source of bias. *Int J Epidemiol* 29(5):856-861
- Basso O, Olsen J, Christensen K (1998a) Risk of preterm delivery, low birthweight and growth retardation following spontaneous abortion: a registry-based study in Denmark. *Int J Epidemiol* 27(4):642-646
- Basso O, Olsen J, Christensen K (1999) Low birthweight and prematurity in relation to paternal factors: a study of recurrence. *Int J Epidemiol* 28(4):695-700
- Basso O, Olsen J, Knudsen LB, Christensen K (1998b) Low birth weight and preterm birth after short interpregnancy intervals. *Am J Obstet Gynecol* 178(2):259-263
- Basso O, Weinberg CR, Baird DD, Wilcox AJ, Olsen J (2003) Subfecundity as a correlate of preeclampsia: a study within the Danish National Birth Cohort. *Am J Epidemiol* 157(3):195-202
- Berkowitz GS, Papiernik E (1993): Epidemiology of preterm birth. *Epidemiol Rev* 15(2):414-443
- Bianchi DW (2000) Fetomaternal cell trafficking: a new cause of disease? *Am J Med Genet* 91(1):22-28
- Bianchi DW, Zickwolf GK, Weil GJ, Sylvester S, DeMaria MA (1996) Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. *Proc Natl Acad Sci USA* 93(2):705-708
- Blair E, Stanley F (1993) When can cerebral palsy be prevented? The generation of causal hypotheses by multivariate analysis of a case-control study. *Paediatr Perinat Epidemiol* 7(3):272-301
- Bolumar F, Olsen J, Boldsen J (1996) Smoking reduces fecundity: a European multicenter study on infertility and subfecundity. The European Study Group on Infertility and Subfecundity. *Am J Epidemiol* 143(6):578-587
- Bonde JP, Ernst E, Jensen TK, Hjollund NHI, Kolstad H, Henriksen TB, Scheike T, Givercman A, Olsen J, Skakkebaek NE (1998a) Relation between semen quality and fertility, a population-based study of 430 first-pregnancy planners. *Lancet* 352(9135):1172-1177
- Bonde JP, Hjollund NHI, Jensen TK, Ernst E, Kolstad H, Henriksen TB, Givercman A, Skakkebaek NE, Andersen AM Olsen J (1998b) A follow-up study of environmental and biologic determinants of fertility among 430 Danish first-pregnancy planners: design and methods. *Reprod Toxicol* 12(1):19-27
- Bonomo M, Gandini ML, Mastropasqua A, Begher C, Valentini U, Faden D, Morabito A (1998) Which cutoff level should be used in screening for glucose intolerance in pregnancy? Definition of Screening Methods for Gestational Diabetes Study Group of the Lombardy Section of the Italian Society of Diabetology. *Am J Obstet Gynecol* 179(1):179-185

- Broughton Pipkin F (1999) What is the place of genetics in the pathogenesis of pre-eclampsia? *Biol Neonate* 76(6):325–330
- Carlsen E, Giwercman A, Keiding N, Skakkebaek NE (1992) Evidence for decreasing quality of semen during past 50 years. *BMJ* 305(6854):609–613
- Cnattingius S, Signorello LB, Anneren G, Clausson B, Ekblom A, Ljunger E, Blot WJ, McLaughlin JK, Petersson G, Rane A, Granath F (2000) Caffeine intake and the risk of first-trimester spontaneous abortion. *N Engl J Med* 343(25):1839–1845
- Condé-Agudelo A, Althabe F, Belizan JM, Kafury-Goeta AC (1999) Cigarette smoking during pregnancy and risk of preeclampsia: a systematic review. *Am J Obstet Gynecol* 181(4):1026–1035
- Coughlin MT, LaPorte RE, O'Leary LA, Lee PA (1998) How accurate is male recall of reproductive information? *Am J Epidemiol* 148(8):806–809
- Cramer DW, Goldman MB (guest eds) (1994) Infertility and reproductive medicine. Study designs and statistics for infertility research. *Clinics of North America* vol 5(2). W. B. Saunders Company, Philadelphia, USA
- Dashe JS, McIntire DD, Ramus RM, Santos-Ramos R, Twickler DM (2002) Persistence of placenta previa according to gestational age at ultrasound detection. *Obstet Gynecol* 99(5 Pt 1):692–697
- Dekker G, Sibai B (2001) Primary, secondary, and tertiary prevention of pre-eclampsia. *Lancet* 357(9251):209–215
- Dekker GA (1999) Risk factors for pre-eclampsia. *Clin Obstet Gynecol* 42(3):422–435
- Dekker GA, Robillard PY, Hulsey TC (1998) Immune maladaptation in the etiology of preeclampsia: a review of corroborative epidemiologic studies. *Obstet Gynecol Surv* 53(6):377–382
- Diggle GE (2001) Thalidomide: 40 years on. *Int J Clin Pract* 55(9):627–631
- Depue RH, Bernstein L, Ross RK, Judd HL, Henderson BE (1987) Hyperemesis gravidarum in relation to estradiol levels, pregnancy outcome, and other maternal factors: a seroepidemiologic study. *Am J Obstet Gynecol* 156(5):1137–1141
- Dornhorst A, Beard RW (1993) Gestational diabetes: a challenge for the future. *Diabet Med* 10(10):897–905
- Eliakim R, Abulafia O, Sherer DM (2000) Hyperemesis gravidarum: a current review. *Am J Perinatol* 17(4):207–218
- Ericson A, Kallen B, Wiholm B (1999) Delivery outcome after the use of antidepressants in early pregnancy. *Eur J Clin Pharmacol* 55(7):503–508
- Esplin MS, Fausett MB, Fraser A (2001) Paternal and maternal components of the predisposition to preeclampsia. *N Engl J Med* 344(12):867–872
- Flint C, Larsen H, Nielsen GL, Olsen J, Sorensen HT (2002) Pregnancy outcome after suicide attempt by drug use: a Danish population-based study. *Acta Obstet Gynecol Scand* 81(6):516–522
- Gabbe SG (1998) The gestational diabetes mellitus conferences. Three are history: focus on the fourth. *Diabetes Care* 21 Suppl 2:B1–2
- Geddes JR, Lawrie SM (1995) Obstetric complications and schizophrenia: a meta-analysis. *Br J Psychiatry* 167(6):786–793

- Gjessing HK, Skjaerven R, Wilcox AJ (1999) Errors in gestational age: evidence of bleeding early in pregnancy. *Am J Public Health* 89(2):213–218
- Gladden BC (1986) On the role of “habitual aborters” in the analysis of spontaneous abortion. *Stat Med* 5(6):557–564
- Godsey RK, Newman RB (1991) Hyperemesis gravidarum. A comparison of single and multiple admissions. *J Reprod Med* 36(4):287–290
- Gourbin G, Masuy-Stroobant G (1995) Registration of vital data: are live births and stillbirths comparable all over Europe? *Bull World Health Organ* 73(4):449–460
- Graafmans WC, Richardus JH, Borsboom GJ, Bakketeig L, Langhoff-Roos J, Bergsjø P, Macfarlane A, Verloove-Vanhorick SP, Mackenbach JP, EuroNatal working group (2002) Birth weight and perinatal mortality, a comparison of “optimal” birth weight in seven Western European countries. *Epidemiology* 13(5):569–574
- Grandjean P, Weihe P, White RF, Debes F, Araki S, Yokoyama K, Murata K, Sorensen N, Dahl R, Jorgensen PJ (1997) Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicol Teratol* 19(6):417–428
- Gray PH, O’Callaghan MJ, Mohay HA, Burns YR, King JF (1998) Maternal hypertension and neurodevelopmental outcome in very preterm infants. *Arch Dis Child Fetal Neonatal Ed* 79(2):F88–93
- Gross S, Librach C, Cecutti A (1989) Maternal weight loss associated with hyperemesis gravidarum: a predictor of fetal outcome. *Am J Obstet Gynecol* 160(4):906–909
- Haig D (1993) Genetic conflicts in human pregnancy. *Q Rev Biol* 68(4):495–532
- Hallak M, Tsalamandris K, Dombrowski MP, Isada NB, Pryde PG, Evans MI (1996) Hyperemesis gravidarum. Effects on fetal outcome. *J Reprod Med* 41(11):871–874
- Henriksen TB, Wilcox AJ, Hedegaard M, Secher NJ (1995) Bias in studies of preterm and postterm delivery due to ultrasound assessment of gestational age. *Epidemiology* 6(5):533–537
- Hjollund NH, Jensen TK, Bonde JP, Henriksen TB, Andersson AM, Kolstad HA, Ernst E, Giwercman A, Skakkebaek NE, Olsen J (2000) Spontaneous abortion and physical strain around implantation, a follow-up study of first-pregnancy planners. *Epidemiology* 11(1):18–23
- Hoffman EB, Sen PK, Weinberg CR (2001) Within-cluster resampling. *Biometrika* 88(4):1121–1134
- Irgens HU, Reisaeter L, Irgens LM, Lie RT (2001) Long term mortality of mothers and fathers after pre-eclampsia: population based cohort study. *BMJ* 323(7232):1213–1217
- Jacobsen R, Bostofte E, Engholm G, Hansen J, Olsen JH, Skakkebaek NE, Møller H. (2000) Risk of testicular cancer in men with abnormal semen characteristics: cohort study. *BMJ* 321(7264):789–792
- Jensen TK, Henriksen TB, Hjollund NHI, Scheike T, Kolstad H, Giwercman A, Ernst E, Bonde JP, Skakkebaek NE, Olsen J (1998) Caffeine intake and fecundability. A follow-up study among 430 Danish couples planning their first pregnancy. *Reprod Toxicol* 12(3):289–295

- Joffe M, Villard L, Li Z, Plowman R, Vessey M (1993) Long-term recall of time-to-pregnancy. *Fertil Steril* 60(1):99–104
- Joffe M, Villard L, Li Z, Plowman R, Vessey M (1995) A time to pregnancy questionnaire designed for long term recall: validity in Oxford, England. *J Epidemiol Community Health* 49(3):314–319
- Johnstone FD, Nasrat AA, Prescott RJ (1990) The effect of established and gestational diabetes on pregnancy outcome. *Br J Obstet Gynaecol* 97(11):1009–1015
- Juul S, Karmaus W, Olsen J (1999) Regional differences in waiting time to pregnancy: pregnancy-based surveys from Denmark, France, Germany, Italy and Sweden. The European Infertility and Subfecundity Study Group. *Hum Reprod* 14(5):1250–1254
- Kallen B (1987) Hyperemesis during pregnancy and delivery outcome: a registry study. *Eur J Obstet Gynecol Reprod Biol* 26(4):291–302
- Kero J, Gissler M, Gronlund MM, Kero P, Koskinen P, Hemminki E, Isolauri E (2002) Mode of delivery and asthma – is there a connection? *Pediatr Res* 52(1):6–11
- Kiely JL, Paneth N, Susser M (1985) Fetal death during labor: an epidemiologic indicator of level of obstetric care. *Am J Obst Gynecol* 153(7):721–727
- Kline J, Stein Z, Susser M (1989) Conception to birth – epidemiology of prenatal development. *Monographs in Epidemiology and Biostatistics* vol 14. Oxford University Press, New York
- Knox EG, Armstrong EH, Lancashire R (1984) The quality of notification of congenital malformations. *J Epidemiol Community Health* 38(4):296–305
- Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC (2002) Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. *Nat Genet* 32(2):285–289
- Kramer MS, Liu S, Luo Z, Yuan H, Platt RW, Joseph KS (2002) Analysis of perinatal mortality and its components: time for a change? *Am J Epidemiol* 156(6):493–497
- Li DK, Wi S (2000) Changing paternity and the risk of preeclampsia/eclampsia in the subsequent pregnancy. *Am J Epidemiol* 151(1):57–62
- Lie RT, Rasmussen S, Brunborg H, Gjessing HK, Lie-Nielsen E, Irgens LM (1998) Fetal and maternal contributions to risk of pre-eclampsia: population based study. *BMJ* 316(7141):1343–1347
- Linnet KM, Dalsgaard S, Obel C, Wisborg K, Henriksen TB, Rodriguez A, Kotimaa A, Moilanen I, Tomsen PH, Olsen J, Jarvelin MR (2003) Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: review of the current evidence. *Am J Psychiatry* 160(6):1028–1040
- Livingston JC, Barton JR, Park V, Haddad B, Phillips O, Sibai BM (2001) Maternal and fetal inherited thrombophilias are not related to the development of severe preeclampsia. *Am J Obstet Gynecol* 185(1):153–157

- Longnecker MP, Klebanoff MA, Zhou H, Brock JW (2001) Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth. *Lancet* 358(9276):110–114
- Ludwig M, Diedrich K (2002) Follow-up of children born after assisted reproductive technologies. *Reprod Biomed Online* 5(3):317–322
- Macklon NS, Geraedts JP, Fauser BC (2002) Conception to ongoing pregnancy: the 'black box' of early pregnancy loss. *Hum Reprod Update* 8(4):333–343
- Maclure M (1991) The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 133(2):144–153
- Martin FI, Ratnaik S, Wootton A, Condos P, Suter PE (1995) The 75 g oral glucose tolerance in pregnancy. *Diabetes Res Clin Pract* 27(2):147–151
- McCormack VA, dos Santos Silva I, De Stavola BL, Mohsen R, Leon DA, Lithell HO (2003) Fetal growth and subsequent risk of breast cancer: results from long term follow up of Swedish cohort. *BMJ* 326(7383):248
- Mednick SA, Huttunen MO, Machon RA (1994) Prenatal influenza infections and adult schizophrenia. *Schizophr Bull* 20(2):263–267
- Metz CE (1978) Basic principles of ROC analysis. *Semin Nucl Med* 8(4):283–298
- Mitchell AA, Cottler LB, Shapiro S (1986) Effect of questionnaire design on recall of drug exposure in pregnancy. *Am J Epidemiol* 123(4):670–676
- Murphy DJ, Sellers S, MacKenzie IZ, Yudkin PL, Johnson AM (1995) Case-control study of antenatal and intrapartum risk factors for cerebral palsy in very preterm singleton babies. *Lancet* 346(8988):1449–1454
- Ness RB, Roberts JM (1996) Heterogeneous causes constituting the single syndrome of preeclampsia: a hypothesis and its implications. *Am J Obstet Gynecol* 175(5):1365–1370
- Nybo Andersen A-M, Wohlfahrt J, Christens P, Olsen J, Melbye M (2000) Maternal age and fetal loss: population based register linkage study. *BMJ* 320(7251):1708–1712
- Olsen J (1984) Calculating risk ratios for spontaneous abortions: the problem of induced abortions. *Int J Epidemiol* 13(3):347–350
- Olsen J (1994) Options in making use of pregnancy history in planning and analysing studies of reproductive failure. *J Epidemiol Community Health* 48(2):171–174
- Olsen J (2000) Prenatal exposures and long-term health effects. *Epidemiol Rev* 22(1):76–81
- Olsen J, Andersen PK (1998) Accounting for pregnancy dependence in epidemiologic studies of pregnancy outcomes. *Epidemiology* 9(3):363–364
- Olsen J, Andersen PK (1999) We should monitor human fecundity, but how? A suggestion for a new method that may also be used to identify determinants of low fecundity. *Epidemiology* 10(4):419–421
- Olsen J, Basso O (2001) Study design. In: Olsen J, Saracci R, Trichopoulos D (eds) *Teaching epidemiology. A guide for teachers in epidemiology, public health and clinical medicine*, 2 edn. Oxford University Press, New York, pp 41–52

- Olsen J, Basso O, Spinelli A, Küppers-Chinnow M and the European Study Group on Infertility and Subfecundity (1998a) Correlates of care seeking for infertility treatment in Europe. *Eur J Public Health* 8(1):15–20
- Olsen J, Czeizel A, Sorensen HT, Nielsen GL, de Jong van den Berg LT, Irgens LM, Olesen C, Pedersen L, Larsen H, Lie RT, de Vries CS, Bergman U (2002) How do we best detect toxic effects of drugs taken during pregnancy? A EuroMap paper. *Drug Saf* 25(1):21–32
- Olsen J, Juul S, Basso O (1998b) Measuring time to pregnancy. Methodological issues to consider. *Hum Reprod* 13(7):1751–1753
- Olsen J, Melbye M, Olsen SF, Sorensen TI, Aaby P, Andersen AM, Taxbol D, Hansen KD, Juhl M, Schow TB, Sorensen HT, Andresen J, Mortensen EL, Olesen AW, Sondergaard C (2001) The Danish National Birth Cohort – its background, structure and aim. *Scand J Public Health* 29(4):300–307
- Olsen J, Rachootin P, Schiødt AV, Damsbo N (1983) Tobacco use, alcohol consumption and infertility. *Int J Epidemiol* 12(2):179–184
- Olsen J, Rachootin P (2003) Invited commentary: monitoring fecundity over time – if we do it, then let's do it right. *Am J Epidemiol* 157(2):94–97
- Olsen J, Schmidt MM, Christensen K (1997) Evaluation of nature-nurture impact on reproductive health using half-siblings. *Epidemiol* 8(1):6–11
- Parker JD, Schoendorf KC (2002) Implications of cleaning gestational age data. *Paediatr Perinat Epidemiol* 16(2):181–187
- Pharoah PO, Adi Y (2000) Consequences of in-utero death in a twin pregnancy. *Lancet* 355(9215):1597–1602
- Quenby S, Vince G, Farquharson R, Aplin J (2002) Opinion. Recurrent miscarriage: a defect in nature's quality control? *Human Reprod* 17(8):1959–1963
- Rachootin P, Olsen J (1982) Prevalence and socioeconomic correlates of subfecundity and spontaneous abortion in Denmark. *Int J Epidemiol* 11(3):245–249
- Rachootin P, Olsen J (1983) The risk of infertility and delayed conception associated with exposure in the Danish workplace. *J Occup Med* 25(5):394–402
- Ravelli GP, Stein ZA, Susser MW (1976) Obesity in young men after famine exposure in utero and early infancy. *N Engl J Med* 295(7):349–353
- Roberts JM, Cooper DW (2001) Pathogenesis and genetics of pre-eclampsia. *Lancet* 357(9249):53–56
- Roberts JM, Lain KY (2002) Recent insights into the pathogenesis of pre-eclampsia. *Placenta* 23(5):359–372
- Robillard PY, Hulsey TC, Perianin J, Janky E, Miri EH, Papiernik E (1994) Association of pregnancy-induced hypertension with duration of sexual cohabitation before conception. *Lancet* 344(8928):973–975
- Rockenbauer M, Olsen J, Czeizel AE, Pedersen L, Sorensen HT (2001) Recall bias in a case-control surveillance system on the use of medicine during pregnancy. *Epidemiology* 12(4):461–466
- Rothenberg SJ, Kondrashov V, Manalo M, Jiang J, Cuellar R, Garcia M, Reynoso B, Reyes S, Diaz M, Todd AC (2002) Increases in hypertension and blood pressure during pregnancy with increased bone lead levels. *Am J Epidemiol* 156(12):1079–1087

- Russell IT, Wilcox AJ (1991) A criterion for low birthweight. *Int J Epidemiol* 20(4):1145
- Sabroe S, Olsen J (1998) Perinatal correlates of specific histological types of testicular cancer in patients below 35 years of age: a case-cohort study based on midwives' records in Denmark. *Int J Cancer* 78(2):140-143
- Salas SP (1999) What causes pre-eclampsia? *Baillieres Best Pract Res Clin Obstet Gynaecol* 13(1):41-57
- Savitz DA, Sonnenfeld NL, Olshan AF (1994) Review of epidemiologic studies of paternal occupational exposure and spontaneous abortion. *Am J Ind Med* 25(3):361-383
- Schmidt MI, Duncan BB, Reichelt AJ, Branchtein L, Matos MC, Costa e Forti A, Spichler ER, Pousada JM, Teixeira MM, Yamashita T (2001) Gestational diabetes mellitus diagnosed with a 2-h 75-g oral glucose tolerance test and adverse pregnancy outcomes. *Diabetes Care* 24(7):1151-1155
- Schmidt MI, Matos MC, Reichelt AJ, Forti AC, de Lima L, Duncan BB (2000) Prevalence of gestational diabetes mellitus - do the new WHO criteria make a difference? Brazilian Gestational Diabetes Study Group. *Diabet Med* 17(5):376-380
- Sharpe RM, Skakkebaek NE (1993) Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract? *Lancet* 341(8857):1392-1395
- Sibai BM (1998) Prevention of preeclampsia: a big disappointment. *Am J Obstet Gynecol* 179(5):1275-1278
- Skjaerven R, Gjessing HK, Bakketeig LS (2000) New standards for birth weight by gestational age using family data. *Am J Obstet Gynecol* 183(3):689-696
- Skjaerven R, Wilcox AJ, Lie RT (2002) The interval between pregnancies and the risk of preeclampsia. *N Engl J Med* 346(1):33-38
- Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease. 30th Thomas Francis Jr. Memorial Lecture, delivered by George Davey Smith at the University of Michigan, School of Public Health, 6 March 2003. *Int J Epidemiol* 32:1-22
- Smoking or Health - the third Report from the Royal College of Physicians of London (1977) Pitman Medical Publishing, London
- Sondergaard C, Skajaa E, Henriksen TB (2000) Fetal growth and infantile colic. *Arch Dis Child Fetal Neonatal Ed* 83(1):F44-47
- Stang A, Ahrens W, Bromen K, Baumgardt-Elms C, Jahn I, Stegmaier C, Krega S, Jöckel KH (2001) Undescended testis and the risk of testicular cancer, importance of source and classification of exposure information. *Int J Epidemiol* 30(5):1050-1056
- Starr T, Levine RJ (1983) Assessing effects of occupational exposure on fertility with indirect standardization. *Am J Epidemiol* 118(6):897-904
- Storgaard L, Bonde JP, Ernst E, Andersen CY, Kyvik KO, Olsen J (2002) Effect of prenatal exposure to oestrogen on quality of semen: comparison of twins and singleton brothers. *BMJ* 325(7358):252-253

- Strohsnitter WC, Noller KL, Hoover RN, Robboy SJ, Palmer JR, Titus-Ernstoff L, Kaufman RH, Adam E, Herbst AL, Hatch EE (2001) Cancer risk in men exposed in utero to diethylstilbestrol. *J Natl Cancer Inst* 93(7):545-551
- Susser M, Stein Z (1994) Timing in prenatal nutrition: a reprise of the Dutch Famine Study. *Nutr Rev* 52(3):84-94
- Swan SH, Shaw GM, Schulman J (1992) Reporting and selection bias in case-control studies of congenital malformations. *Epidemiology* 3(4):356-363
- Tong S, Caddy D, Short RV (1997) Use of dizygotic to monozygotic twinning ratio as a measure of fertility. *Lancet* 349(9055):843-845
- Trichopoulos D (1990) Hypothesis: does breast cancer originate in utero? *Lancet* 335(8695):939-940
- Trogstad LI, Eskild A, Magnus P, Samuelsen SO, Nesheim BI (2001) Changing paternity and time since last pregnancy; the impact on pre-eclampsia risk. A study of 547,238 women with and without previous pre-eclampsia. *Int J Epidemiol* 30(6):1317-1322
- Trupin LS, Simon LP, Eskenazi B (1996) Change in paternity: a risk factor for preeclampsia in multiparas. *Epidemiology* 7(3):240-244
- Verdoux H, Geddes JR, Takei N, Lawrie SM, Bovet P, Eagles JM, Heun R, McCreddie RG, McNeil TF, O'Callaghan E, Stober G, Willinger MU, Wright P, Murray RM (1997) Obstetric complications and age at onset in schizophrenia: an international collaborative meta-analysis of individual patient data. *Am J Psychiatry* 154(9):1220-1227
- Vestergaard M, Basso O, Henriksen TB, Oestergaard JR, Olsen J (2002) Risk factors for febrile convulsions. *Epidemiol* 13(3):282-287
- Villar J, Belizan JM (2000) Same nutrient, different hypotheses: disparities in trials of calcium supplementation during pregnancy. *Am J Clin Nutr* 71(5 Suppl):1375S-1379S
- von Dadelszen P, Ornstein MP, Bull SB, Logan AG, Koren G, Magee LA (2000) Fall in mean arterial pressure and fetal growth restriction in pregnancy hypertension: a meta-analysis. *Lancet* 355(9198):87-92
- Weinberg CR (1993) Toward a clearer definition of confounding. *Am J Epidemiol* 137(1):1-8
- Weinberg CR, Baird DD, Wilcox AJ (1994a) Sources of bias in studies of time to pregnancy. *Stat Med* 13(5-7):671-681
- Weinberg CR, Baird DD, Wilcox AJ (1994b) Bias in retrospective studies of spontaneous abortion based on the outcome of the most recent pregnancy. *Ann N Y Acad Sci* 18(709):280-286
- Weinberg CR, Wilcox AJ (1998) Reproductive epidemiology. In: Rothman KJ, Greenland S (eds.) *Modern epidemiology*, 2nd edn. Lippincott-Raven, Philadelphia, USA
- Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 62(4):969-978

- Wen SW, Demissie K, Liu S, Marcoux S, Kramer MS (2000) Placenta praevia and male sex at birth: results from a population-based study. *Paediatr Perinat Epidemiol* 14(4):300–304
- Whitacre CC (2001) Sex differences in autoimmune disease. *Nat Immunol* 2(9):777–780
- Whitacre CC, Reingold SC, O’Looney PA (1999) A gender gap in autoimmunity. *Science* 283(5406):1277–1278
- Wilcox AJ (2001) On the importance – and the unimportance – of birthweight. *Int J Epidemiology* 30(6):1233–1241
- Wilcox AJ, Baird DD, Weinberg CR, Hornsby PP, Herbst AL (1995) Fertility in men exposed prenatally to diethylstilbestrol. *N Engl J Med* 332(21):1411–1416
- Wilcox AJ, Weinberg CR, O’Connor JF, Baird DD, Schlatterer JP, Canfield RE, Armstrong EG, Nisula BC (1988) Incidence of early loss of pregnancy. *N Engl J Med* 319(4):189–194
- Winbo IG, Serenius FH, Dahlquist GG, Kallen BA (1997) A computer-based method for cause of death classification in stillbirths and neonatal deaths. *Int J Epidemiol* 26(6):1298–1306
- Zhang J, Cai WW (1991) Severe vomiting during pregnancy: antenatal correlates and fetal outcomes. *Epidemiol* 2(6):454–457
- Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 39(4):561–577

Textbooks in Reproductive Epidemiology

- Bracken MB (ed) (1984) *Perinatal epidemiology*. Oxford University Press, New York
- Kallen B (1988) *Epidemiology of human reproduction*. CRC Press, Boca Raton
- Keith LG, Papiernik E, Keith DM, Luke B (eds) (1995) *Multiple pregnancy: epidemiology, gestation and perinatal outcome*, 1st edn. Parthenon Publishing Group, New York
- Kiely M (ed) (1991) *Reproductive and perinatal epidemiology*. CRC Press, Boca Raton
- Kline J, Stein Z, Susser M (1989) *Conception to birth – epidemiology of prenatal development*. Monographs in Epidemiology and Biostatistics vol 14. Oxford University Press, New York
- McDowall ME (1985) *Occupational reproductive epidemiology: the use of routinely collected statistics in England and Wales, 1980–82*. Her Majesty’s Stationery Office, London
- Murray CJL, Lopez AD (eds) (1988) *Health dimensions of sex and reproduction*. In: *Global Burden of Disease and Injury Series, Vol III, 1998*. WHO, Geneva