AOGS
Acta Obstetricia et Gynecologica Scandinavica

AOGS REVIEW

Cohort studies in the context of obstetric and gynecologic research: a methodologic overview

CARMEN MESSERLIAN[1] & OLGA BASSO[2,3]

[1]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [2]Department of Obstetrics & Gynecology, Royal Victoria Hospital, Research Institute of McGill University Health Center, and [3]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

Key words
Cohort studies, reproductive, perinatal, epidemiology, design

Correspondence
Carmen Messerlian, Harvard T.H. Chan School of Public Health 665 Huntington Avenue, FXB 102 A, Boston, 02115 MA, USA.
E-mail: cmesser@hsph.harvard.edu

Conflict of interest
The authors have stated explicitly that there are no conflicts of interest in connection with this article.

Please cite this article as: Messerlian C, Basso O. Cohort studies in the context of obstetric and gynecologic research: a methodologic overview. Acta Obstet Gynecol Scand 2018; 97:371–379.

Received: 29 October 2017
Accepted: 22 November 2017

DOI: 10.1111/aogs.13272

Abstract

Observational cohort studies represent one of the most powerful designs in epidemiology. They are also the basis of evidence in many areas of obstetric and gynecologic research, given that randomization of women, couples or pregnancies is often impossible or unethical. Indeed, well-conceived cohort studies have led to a better understanding of many important clinical and public health questions over time, including the impact of different exposures on perinatal and pediatric outcomes in pregnant women and their children. In this paper, we describe the main features, challenges, and limitations of cohort studies in the context of obstetric and gynecologic research. As with all epidemiologic studies, cohort studies present numerous challenges and are vulnerable to bias. However, as we describe throughout this review, careful design – from formulating the study question to planning statistical analysis – can reduce the potential for bias. When possible, we also provide examples from the gynecological and obstetrical literature to illustrate the epidemiological challenge and suggest specific readings.

Abbreviations: IVF, in vitro fertilization.

In epidemiology, the term cohort refers to a group of individuals from a defined population followed over time (1). Technically, a cohort is a "closed" population – an event defines entry into the cohort, and no individuals enter or leave the cohort except through death. Conversely, a dynamic or open population is characterized by a state, and membership ends when the defining state ends (for example students at a given school constitute a dynamic population, whereas students who graduate each year represent a closed population). While often based on a closed population, any study in which individuals sharing one or more common characteristics are followed over time is generally considered a cohort study. Examples of shared characteristics include geography [for example The Danish National Birth Cohort, (2) or The Aberdeen Maternity Hospital Cohort (3)], occupation (for example Nurses Health Study) (4), exposure (for example Seveso Women's Health Study of dioxin exposure) (5), risk profile [for example Sister Study of breast cancer (6) or the Black Women's Health Study] (7) or

Key Message

Cohort studies present numerous advantages as well as challenges. Herein, we describe the main features and limitations of cohort studies in the context of obstetric and gynecologic research, and how careful design can reduce the potential for bias.

disease (for example The British Childhood Cancer Survivor Study) (8).

In this paper, we describe the main features, challenges, and limitations of cohort studies. When possible, we provide examples from the gynecological and obstetrical literature and suggest specific readings.

Where do cohort studies fit in epidemiologic research?

Cohort studies are usually undertaken for etiologic (or analytic) purposes, that is, with the aim of evaluating whether a given exposure causes a specified outcome. Cohort studies can be experimental (investigators have a degree of control over the allocation of the exposure) or observational (investigators have no control over the exposure). Case-control studies, another common design in etiologic research, are also based on an underlying cohort. However, unlike in typical cohort studies, identification of the cohort is often secondary (i.e. cases are identified before defining the underlying cohort from which they originated). Both cohort and case-control studies are longitudinal in nature: the information gathered refers to different points in time, from exposure to outcome, even though in many case-control studies exposures are ascertained after the outcome has occurred. For exposures that are stable in time, a cross-sectional design – where exposure and outcome are measured concurrently with no attempt to reconstruct an exposure history – may also be appropriate for etiologic purposes, provided that the exposure is not associated with survival to the event of interest. Such an approach could be used, for example, with exposures such as genetic variants (9), parental age at conception (or birth) or daughter's probability of lifelong childlessness (10). However, as in case-control studies, there is greater potential for bias if prevalent (rather than incident) cases are recruited, as survival may be differential by exposure. Essentially, however, all three designs (cohort, case-control, and cross-sectional studies) emanate from the concept of measuring events in a "sea of person-time" (11). This represents the distribution of person-time where all events occur (with the time from exposure to outcome ranging from very short to very long) and is the underlying sampling frame for all such designs. Individual studies differ in how the population is defined and in how and when the exposed and unexposed members of the population or group are sampled.

Observational cohort studies are the main source of evidence in many areas of obstetric and gynecologic research, given that randomization of women, couples or pregnancies is often impossible or unethical. Indeed, well-conceived cohort studies have led to a better understanding of rare exposures and their sequelae, such as the long-term health of diethylstilbestrol (DES)-exposed women and their sons and daughters (12–14), and through well-established birth cohorts (15–20), to substantial evidence in perinatal and pediatric health.

Main features of cohort studies

In cohort studies, researchers aim to determine whether exposure status (at cohort entry or over time, for time-varying exposures) is associated with the incidence of the outcome(s) of interest. Regardless of the timing of the outcome relative to the start of the actual study, all etiologic studies aim to measure the exposure → outcome relation. However, in cohort studies the design explicitly reflects this temporal sequence (Figure 1). Cohort studies are sometimes classified as either prospective or historical (also sometimes called "retrospective") based on the time at which the exposure → outcome relation is observed relative to current time. The term historical cohort study is broadly used to describe longitudinal studies in which outcomes have already occurred when the cohort is first defined. In such instances, the investigator reconstructs the person-time experience and exposure history through secondary data (such as administrative registries or medical records) (21). However, the terms prospective and historical more appropriately refer to the timing of exposure assessment with respect to the outcome: prospective when measurement of exposure precedes the occurrence of the outcome, and historical when the outcome has already occurred when exposure data are collected (Figure 1). Often historical cohort studies with prospective assessment of exposure and covariates are more efficient (and less costly) than prospective (or concurrent) ones. If well-designed and appropriately analyzed, they may also be less vulnerable to bias from selective participation and retention. However, in such cases, researchers are limited by the quantity and quality of the available information. Furthermore, compared with concurrent cohort studies, there is a greater danger of error, particularly having to do with time, such as failing correctly to assign exposed and unexposed time (see section on immortal time bias) and inadvertently "peeking" into the future, which will nearly always result in invalid findings [see, for example Bakketeig & Hoffman (22) on perinatal mortality and birth order, and related commentaries (23,24)]. In prospective cohort studies, investigators measure exposures and covariates in eligible individuals at cohort entry and, often, at regular intervals during follow-up. Outcomes are documented, as they occur in real time, either by contacting study participants or passively, for example through population-based registries.

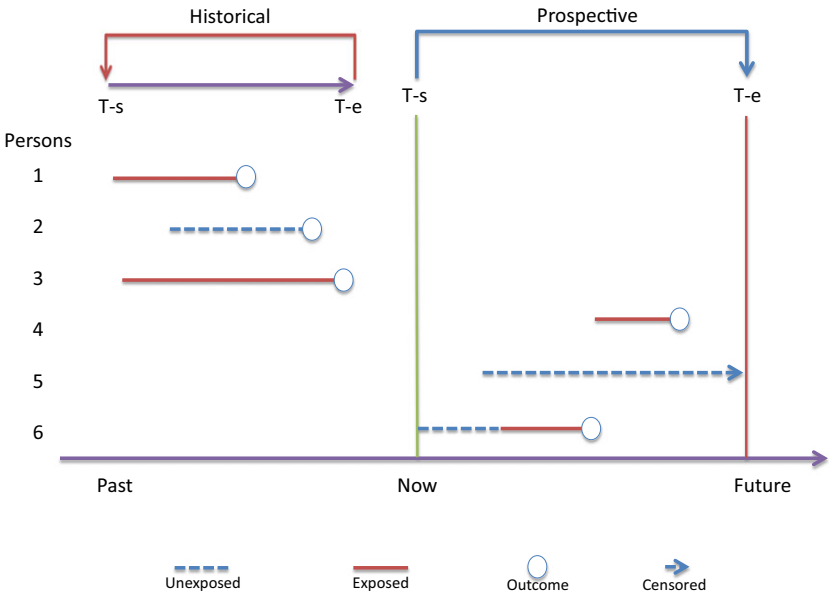

Figure 1. Schematic representation of historical and prospective designs, and exposure, outcome, and censored follow-up time in cohort studies. [Color figure can be viewed at wileyonlinelibrary.com]

Cohort definition and eligibility

When designing an observational cohort study, it is essential to start with a well-defined study question (or questions). A crucial aspect of this process is accurately defining the exposed and referent categories. Although the exposure is generally given more attention, the referent category is no less important. For example, if the objective is to determine whether delivery induction from ≥37 weeks of gestation without medical indication (i.e. for social reasons) is associated with a higher risk of adverse perinatal and obstetric outcome, defining exposed pregnancies is straightforward: any pregnancy induced from week 37 for non-medical reasons. Reference (unexposed) pregnancies are not those delivered spontaneously at the same time as the exposed ones but ongoing pregnancies at the same gestational age (25). To see why, it is helpful to think about how this question would be addressed in a clinical trial: eligible women (i.e. pregnant, with no medical indication for delivery) would be recruited at, say, 36 weeks of gestation, and randomized to being induced (at 37, 38, or 39 weeks) or expectant management. At the end of the study, women randomized to being delivered at 37 weeks would have experienced overall fewer instances of fetal death, preeclampsia, and macrosomia than those randomized to expectant management (as these outcomes are "prevented" by delivery) but their babies will have more respiratory and other complications related to having been delivered early. If pregnancies induced at 37 weeks were to be compared with spontaneous deliveries at 37 weeks, such a study would not answer the relevant clinical question, as the decision to induce or not is relevant only for women who are still pregnant.

Appropriate definition of the research question also facilitates the next critical step, that of determining inclusion and exclusion criteria. Eligible participants should have an equal chance of enrolling and remaining in the study, irrespective of their exposure status or probability of the outcome. At enrolment, investigators should establish that eligible individuals are not only at risk of the outcome(s) of interest (for example women who have had a hysterectomy would not be included in a study where uterine cancer is the outcome), but also free from the outcome (women with prevalent uterine cancer would be excluded, as well as women diagnosed in the first months of follow-up, since disease may have been present at enrolment).

Exposure and outcome ascertainment

An advantage of cohort studies is that if the exposure of interest is rare, the investigator can choose the study population to maximize the likelihood that a sufficient proportion of cohort participants will be those exposed (for example, by studying nail salon technicians if exposure to solvents is of interest) and also ensure that there will be a broad variability (for continuous exposures, the broader the exposure range, the higher the power to detect an

effect). Exposure status (current and, if appropriate, past) is determined at baseline for all members of the cohort. Exposure should also be updated during follow-up time because an individual who is unexposed at baseline may subsequently become exposed (or vice versa), and exposed and unexposed time must be allocated accordingly. Depending on the study, an exposure may be unchangeable (for example diethylstilbestrol exposure in utero) or time-varying (for instance taking oral contraceptives, or contracting influenza during pregnancy). In the latter case, follow-up time should be divided into exposed and unexposed intervals and analyzed accordingly [for example, with Cox regression (26)]. The level of detail of exposure assessment should be appropriate for the study question and should aim at minimizing the potential for exposure misclassification. However, cost and time considerations – as well as the fact that many cohort studies are not focused on a single exposure–outcome relation – often preclude collecting information on all exposures of interest at the optimal level of detail.

Follow-up of all members of the cohort allows ascertainment of new cases (incident cases) of the outcomes as they occur. Typically, when analyzing a given endpoint, the end of follow-up for each cohort member is determined by the occurrence of the outcome under study, death, loss to follow-up, withdrawal, or study end, whichever comes first (see Figure 1).

The principle of "comparable accuracy" for data collection procedures dictates that the intensity of follow-up and accuracy of measurement (of outcomes and covariates) be equivalent across all exposure categories, as differential quality, quantity or accuracy of follow-up data can lead to misclassification or information bias (27).

Outcome measurement

Measuring the occurrence of certain outcomes, particularly in perinatal epidemiology, can be challenging, as the relevant denominators (and numerators) are often missing (28). For example, the number of conceptions and early losses is nearly always unknown. [Although extremely laborious outside an assisted reproductive technology treatment setting, it is possible to estimate the number of implantations and early losses in a cohort of women over a certain period (29–31)]. Even miscarriages occurring later in the first trimester of pregnancy are challenging to measure. Although the proportion of congenital malformations at birth can be estimated (with a degree of error if neonates are assessed only at birth), this quantity is a prevalence – not an incidence. Incidence cannot be estimated, as the proportion of fetuses with malformations that are lost from conception to birth is unknown. [The estimated difference between the proportion of fetuses diagnosed with Down's syndrome at prenatal screening and the proportion at birth clearly illustrates this problem (32)].

Several endpoints – for example, fetal death and preeclampsia – are defined from around 20 weeks of gestation, which is when registration of pregnancies that have survived to that point is mandatory in many countries. Thus, the incidence rate of fetal death and other pregnancy events can be estimated, provided that the underlying cohort of fetuses (or pregnancies, if the outcome of interest is pregnancy complications or obstetric interventions) is identifiable, with competing events accounted for. [Competing events are those that prevent observation of the outcome of interest. For example, in the case of preeclampsia, competing events would be preterm births and fetal deaths for causes other than preeclampsia].

As in all areas of epidemiology, identifying the appropriate denominator is essential to calculate the relevant measure of occurrence. However, even this aspect is challenging. The gestational-age-specific risk of antepartum stillbirth (fetal death) used to be estimated using births at each week as the denominator, until Yudkin and colleagues pointed out that births are not at risk of antepartum stillbirth – fetuses are (33). Pregnancies are the appropriate denominator for estimating the risk of events such as preterm birth, pregnancy complications, and obstetric intervention – and time (i.e. gestational age) should be taken into consideration. [The subtle difference between pregnancies and fetuses depends on the outcome of interest and is due to multifetal pregnancies]. The appropriate denominator for estimating the gestational-age-specific risk of neonatal death remains controversial (34–39), with some researchers suggesting that fetuses be used instead of live births, despite the fact this approach can result in misleading estimates, given that fetuses can only enter the numerator after they are born alive and die as neonates (34,39). Other challenges in measuring outcomes stem from the fact that there are many endpoints that cannot be observed in fetuses, or that can be observed only with substantial uncertainty (for example fetal weight, although improvements in ultrasound imaging may reduce such errors). Although we generally speak of incidence when referring to a newly diagnosed condition, in many cases, they preexist diagnosis by an unknown amount of time. Think for example about ovarian cancer, the vast majority of which is diagnosed when the disease is already very advanced (40). Such uncertainty about the timing of onset often makes it difficult to correctly establish the temporality of the exposure-outcome relation. A well-known example is the difficulty of establishing whether drinking coffee during pregnancy causes miscarriage [see, for example (41)].

Strengths and limitations of cohort studies

The main advantage of cohort studies over other designs is that they can capture dynamic exposures and their relation to events across time (see Figure 1). Besides being appropriate to studying rare exposures, cohort studies allow investigation of multiple outcomes and track the transition, pattern, and change in exposure over time. In cohort studies, investigators can estimate incidence rate ratios (or differences) or risk ratios (or differences), depending on the outcome and type of population [see (42) for general measures in epidemiology and (28) for measures of specific reproductive endpoints]. They are also highly flexible in terms of exposure and outcome ascertainment. However, cohort studies also have disadvantages, most notably, the long duration of follow-up typically required for many outcomes and the associated costs [although duration is limited for studies of periconceptional exposures and pregnancy outcomes, see (30,31)]. Recruiting and retaining individuals and obtaining and organizing resources for the labor-intensive work involved in collecting and entering data is challenging. Recently, investigators have begun using online platforms for recruitment and data collection. For example, the Snart Gravid study in Denmark (2), and its North American complement, Pregnancy Study Online (PRESTO) (43), recruit couples attempting conception via the internet in order to study lifestyle factors in relation to time to pregnancy, thus reducing costs for recruitment and in-person data collection visits (2). Another important consideration in cohort studies is that they are highly inefficient for studying rare endpoints, such as specific congenital abnormalities.

Methodologic challenges in cohort studies

Selection bias

Selection bias occurs when the probability of being selected and retained in the study depends on exposure and/or disease status. A typical manifestation arises when the unexposed are not representative of the baseline rate of the outcome or disease (i.e. "sea of person-time" distribution) in the source population. The unexposed should be comparable to the exposed in all but exposure status (in observational studies, a portion of the inevitable differences across exposure categories is handled by statistical modeling). The purpose of the unexposed person-time experience is to act as a "counterfactual" for the exposed person-time experience. In other words, the unexposed should generate disease/outcome rates that approximate those that the exposed would have had, *had they not been exposed*.

Another common source of selection bias stems from differential attrition (loss to follow-up and withdrawals) between exposed and unexposed participants. Differential losses based on outcome probability can be a serious threat to the internal validity of a study (44). For example, if pregnant women who consume high levels of alcohol are more likely to be lost to follow-up, and they have worse outcomes, the association between alcohol and pregnancy outcome will be underestimated. Bias due to loss of follow-up becomes a concern if the final study sample differs in its exposure–outcome probability (is more or less at risk) compared with the originally enrolled population.

"Healthy user bias" has been discussed in the context of observational studies of postmenopausal hormone therapy and cardiovascular disease (45–47). Users of hormone therapy were thinner, more physically active, and had a higher socioeconomic status compared with nonusers. Thus, self-selected users of hormone therapy recruited to participate in these observational studies had a lower probability of heart disease compared with the baseline risk of the outcome in the source population. The inclusion of healthy hormone therapy users among the exposed led to overestimating its cardioprotective effects. Some of these studies were also affected by immortal time bias (48) (see section on misclassification bias).

Another source of selection bias is non-response or non-inclusion of participants based on exposure status. Most time-to-pregnancy studies include only women who plan a pregnancy. If an exposure such as smoking is associated with less consistent use of birth control (leading to more accidental pregnancies), smokers who plan their pregnancies will tend to be less fertile than smokers who had an unplanned pregnancy (who are ineligible to enter the study). Such a mechanism will result in overestimating the deleterious effect of smoking on fecundity (49,50).

In cohorts restricted to preterm births (such as neonatal networks), only certain etiologic questions can be explored without incurring selection bias. A cohort of very preterm infants is appropriate to study whether, for example, administration of probiotics reduces the risk of necrotizing enterocolitis. However, if the research question is whether preeclampsia results in a lower risk of neonatal death, then the study will be affected by selection bias – specifically, "collider-stratification bias" (51–53). The underlying problem is that the exposed infants (those born following preeclampsia) are being compared with infants who have several pathologies, other than preeclampsia, that have resulted in preterm birth. If these factors themselves increase mortality risk, preeclampsia

may appear protective only by virtue of being associated with a lower risk of neonatal death than the totality of causes in the "unexposed" infants. It is worth noting that restricting to preterm births is akin to adjusting for gestational age, as mentioned in the section about confounding. For a non-technical demonstration of collider-stratification bias in the aforementioned context, see Snowden and Basso (54).

Misclassification bias in cohort studies

All epidemiologic studies suffer from a degree of measurement error; if random, this generally results in bias towards the null (although not consistently when the exposure is not binary). The main concern is differential accuracy of measurement by exposure (or outcome) status, which will bias effect estimates, with direction and magnitude depending on the pattern and extent of differential accuracy. For example, at birth, physicians may examine more carefully infants born to women who conceived by in vitro fertilization (IVF) than those who conceived naturally. If a researcher were to use information in birth records to assess whether babies conceived by IVF are more likely to have congenital malformations, the risks associated with IVF would be overestimated. (Note that this example ignores the possibility that detection of malformations may have been differential also during pregnancy, as well as the possibility that couples who conceived through IVF may be more reluctant to terminate an affected pregnancy). If the outcome is a condition that would be diagnosed in any case, such as fetal death, then such a bias is less likely to occur – but many endpoints are not as clearcut. It should also be noted that, for relatively rare endpoints, a specificity less than one, even if non-differential, can substantially bias relative measures of association (for example the relative risk) towards the null.

An insidious type of information bias in cohort studies is "immortal time bias". First described in studies of heart transplant, and a common problem in pharmacoepidemiology (55), this bias occurs when there is a portion of follow-up time during which, by design, the outcome of interest cannot occur. Hutcheon et al. (2013) described this in the context of studies of stillbirth as a function of gestational diabetes (56). The authors show how the association of gestational diabetes with stillbirth changes from apparently protective when all births from 20 weeks are examined, to harmful when only births from 28 weeks and onward are included in the analysis. As gestational diabetes is usually diagnosed around 24–28 weeks, only pregnancies that survive to diagnosis can be screened. However, all stillbirths occurring before screening will, by definition, be considered unexposed. There are other examples of circumstances in which this type of bias can affect studies of pregnancy outcomes (57).

Confounding

Confounding occurs when the effect of the exposure is mixed with that of another factor that is associated with the exposure (but is not caused by it) and is a cause (or a proxy of a cause) of the outcome. For example, in the association between twinning and neonatal mortality, confounders will be factors that predict twinning and are risk factors of neonatal death. Thus, age, parity, conception by assisted reproduction, maternal race/ethnicity are confounders. However, gestational age at birth is *not* a confounder; rather it is an intermediate between twinning and mortality and should *not* be adjusted for (51–53). Confounders should not be determined based on a statistically significant difference between exposed and unexposed. As stated above, potential confounders should be identified based on the knowledge of their causal relation with exposure and outcome. [Sometimes, the decision as to which among these factors should be retained in the final model is based on whether they change the estimated measure of association by a certain amount (typically, 10% or more)]. Even when most important confounders are available, residual confounding (i.e. what confounding is left after attempts to control for it) remains a concern. Residual confounding generally results from relying on a variable that does not adequately represent the confounder (for example education as the sole indicator of socioeconomic status) and from inappropriate modeling (for example if a factor is modeled linearly but, in fact, has a J- or U-shaped relation with the outcome, or when the categories of the confounder are too broad).

Generalizability and participation rates

Generalizability refers to the extent to which findings of a given study can be applied to other populations with similar characteristics. For example, the Environment and Reproductive Health (EARTH) Study is a prospective preconception cohort of subfertile couples attempting conception, recruited at a large fertility clinic in Boston, MA (USA) (31). The study was designed to evaluate the impact of environmental exposure and diet on fertility and pregnancy outcomes. If subfertile couples were more sensitive to the adverse effects of environmental exposures, then findings from this study would be less generalizable to men and women from the overall population. However, they would still be generalizable to subfertile couples, provided that patients attending the clinic adequately represent all subfertile couples, not all of whom

seek help if they cannot conceive (58). Lack of generalizability, although worthy of note in the discussion, is not nearly as problematic as lack of internal validity.

A declining interest in taking part in studies has resulted in many cohort studies, especially those with heavy participant burden, being afflicted by low participation. Rates have declined from 80% to 30–40% in the last several decades (59). Low participation rates do not necessarily imply that the association identified from the study will be biased (59,60). However, this should be evaluated for every study, and generalizability may be affected.

Conclusion

Observational cohort studies represent one of the most powerful designs in epidemiology. As all epidemiologic studies, they present numerous challenges and are vulnerable to bias. Careful design, from formulating the study question to planning statistical analysis, will substantially reduce – but not eliminate – the potential for bias. We hope that our overview of cohort studies in the context of obstetric and gynecologic research will prove useful when evaluating the published evidence from this type of study.

References

1. Porta MS, Greenland S, Hernán M, Silva IdS, Last JM; International Epidemiological Association. A dictionary of epidemiology, 6th edn. Oxford: Oxford University Press, 2014. ii, 343 pages.
2. Huybrechts KF, Mikkelsen EM, Christensen T, Riis AH, Hatch EE, Wise LA, et al. A successful implementation of e-epidemiology: the Danish pregnancy planning study "Snart-Gravid". Eur J Epidemiol. 2010;25:297–304.
3. Ayorinde AA, Wilde K, Lemon J, Campbell D, Bhattacharya S. Data resource profile: the Aberdeen Maternity and Neonatal Databank (AMND). Int J Epidemiol. 2016;45:389–94.
4. Colditz GA. The nurses' health study: a cohort of US women followed since 1976. J Am Med Womens Assoc (1972). 1995;50:40–4.
5. Eskenazi B, Mocarelli P, Warner M, Samuels S, Vercellini P, Olive D, et al. Seveso Women's Health Study: a study of the effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on reproductive health. Chemosphere. 2000;40:1247–53.
6. Weinberg CR, Shore DL, Umbach DM, Sandler DP. Using risk-based sampling to enrich cohorts for endpoints, genes, and exposures. Am J Epidemiol. 2007;166:447–55.
7. Rosenberg L, Adams-Campbell L, Palmer JR. The Black Women's Health Study: a follow-up study for causes and preventions of illness. J Am Med Womens Assoc (1972). 1995;50:56–8.
8. Hawkins MM, Lancashire ER, Winter DL, Frobisher C, Reulen RC, Taylor AJ, et al. The British Childhood Cancer Survivor Study: objectives, methods, population structure, response rates and initial descriptive information. Pediatr Blood Cancer. 2008;50:1018–25.
9. Peter I, Shearman AM, Zucker DR, Schmid CH, Demissie S, Cupples LA, et al. Variation in estrogen-related genes and cross-sectional and longitudinal blood pressure in the Framingham Heart Study. J Hypertens. 2005;23:2193–200.
10. Basso ODAA, Weinberg CR, Sandler DP. Maternal age at birth and daughters' subsequent childlessness. Hum Reprod. 2017; https://doi.org/10.1093/humrep/dex350. [Epub ahead of print]
11. Morgenstern H, Kleinbaum DG, Kupper LL. Department of Epidemiology and Public Health, School of Medicine, Yale University, New Haven, CT. Int J Epidemiol. 1980;9:97–104.
12. Hadjimichael OC, Meigs JW, Falcier FW, Thompson WD, Flannery JT. Cancer risk among women exposed to exogenous estrogens during pregnancy. J Natl Cancer Inst. 1984;73:831–4.
13. Leary FJ, Resseguie LJ, Kurland LT, O'Brien PC, Emslander RF, Noller KL. Males exposed in utero to diethylstilbestrol. JAMA. 1984;252:2984–9.
14. Labarthe D, Adam E, Noller KL, O'Brien PC, Robboy SJ, Tilley BC, et al. Design and preliminary observations of National Cooperative Diethylstilbestrol Adenosis (DESAD) Project. Obstet Gynecol. 1978;51:453–8.
15. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C, et al. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol. 2006;35:1146–50.
16. Basso O, Wilcox AJ, Weinberg CR, Baird DD, Olsen J. Height and risk of severe pre-eclampsia. A study within the Danish National Birth Cohort. Int J Epidemiol. 2004;33:858–63.
17. Wadsworth M, Kuh D, Richards M, Hardy R. Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). Int J Epidemiol. 2006;35:49–54.
18. Elliott J, Shepherd P. Cohort profile: 1970 British Birth Cohort (BCS70). Int J Epidemiol. 2006;35:836–43.
19. Jaddoe VW, Mackenbach JP, Moll HA, Steegers EA, Tiemeier H, Verhulst FC, et al. The Generation R Study: design and cohort profile. Eur J Epidemiol. 2006;21:475–84.
20. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42:97–110.
21. Syddall HE, Aihie Sayer A, Dennison EM, Martin HJ, Barker DJ, Cooper C. Cohort profile: the Hertfordshire cohort study. Int J Epidemiol. 2005;34:1234–42.
22. Bakketeig LS, Hoffman HJ. Perinatal mortality by birth order within cohorts based on sibship size. Br Med J. 1979;2:693–6.

23. Golding J. Perinatal epidemiology in Wonderland. Br Med J. 1979;2:1436.
24. Wilcox AJ. Fertility and pregnancy: an epidemiologic perspective. Oxford: Oxford University Press, 2010. xii, 324 p.
25. Caughey AB, Sundaram V, Kaimal AJ, Gienger A, Cheng YW, McDonald KM, et al. Systematic review: elective induction of labor versus expectant management of pregnancy. Ann Intern Med. 2009;151:252–63, W53-63.
26. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health. 1999;20:145–57.
27. Rothman KGS, Lash TL. Modern epidemiology. Philadelphia: Lippincott, Williams & Wilkins, 2012.
28. Olsen J, Basso O. Reproductive epidemiology. In: Ahrens W, Pigeot I (eds). Handbook of epidemiology. Berlin: Springer, 2005. pp. 1043–109.
29. Wilcox AJ, Weinberg CR, O'Connor JF, Baird DD, Schlatterer JP, Canfield RE, et al. Incidence of early loss of pregnancy. N Engl J Med. 1988;319:189–94.
30. Hauser R, Gaskins AJ, Souter I, Smith KW, Dodge LE, Ehrlich S, et al. Urinary phthalate metabolite concentrations and reproductive outcomes among women undergoing in vitro fertilization: results from the EARTH Study. Environ Health Perspect. 2016;124:831–9.
31. Messerlian C, Wylie B, Minguez-Alarcon L, Williams P, Ford J, Souter I, et al. Urinary Concentrations of phthalate metabolites and pregnancy loss among women conceiving with medically assisted reproduction. Epidemiology. 2016;27:879–88.
32. Morris JK, Alberman E. Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register. BMJ. 2009;339:b3794.
33. Yudkin PL, Wood L, Redman CW. Risk of unexplained stillbirth at different gestational ages. Lancet. 1987;1:1192–4.
34. Basso O. Implications of using a fetuses-at-risk approach when fetuses are not at risk. Paediatr Perinat Epidemiol. 2016;30:3–10.
35. Caughey AB, Snowden JM. Measuring Perinatal complications: different approaches depending on who is at risk. Paediatr Perinat Epidemiol. 2016;30:23–4.
36. Joseph KS. A Consilience of inductions supports the extended fetuses-at-risk model. Paediatr Perinat Epidemiol. 2016;30:11–7.
37. Smith GC. Quantifying the risk of different types of perinatal death in relation to gestational age: researchers at risk of causing confusion. Paediatr Perinat Epidemiol. 2016;30:18–9.
38. Schisterman EF, Sjaarda LA. No right answers without knowing your question. Paediatr Perinat Epidemiol. 2016;30:20–2.
39. Basso O. No rates were harmed in the making of this paper: response to critiques. Paediatr Perinat Epidemiol. 2016;30:25–7.
40. Gilbert L, Basso O, Sampalis J, Karp I, Martins C, Feng J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. Lancet Oncol. 2012;13:285–91.
41. Savitz DA, Chan RL, Herring AH, Howards PP, Hartmann KE. Caffeine and miscarriage risk. Epidemiology. 2008;19:55–62.
42. Rothman KJ. Epidemiology: an introduction, 2nd edn. New York: Oxford University Press, 2012. viii, 268 p.
43. Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, et al. Design and conduct of an internet-based preconception cohort study in North America: pregnancy study online. Paediatr Perinat Epidemiol. 2015;29:360–71.
44. Hernan MA. Invited Commentary: selection bias without colliders. Am J Epidemiol. 2017;185:1048–50.
45. Michels KB, Manson JE. Postmenopausal hormone therapy: a reversal of fortune. Circulation. 2003;107:1830–3.
46. Petitti DB, Perlman JA, Sidney S. Noncontraceptive estrogens and mortality: long-term follow-up of women in the Walnut Creek Study. Obstet Gynecol. 1987;70(3 Pt 1):289–93.
47. Petitti DB. Coronary heart disease and estrogen replacement therapy. Can compliance bias explain the results of observational studies? Ann Epidemiol. 1994;4:115–8.
48. Suissa S. Randomized trials built on sand: examples from COPD, hormone therapy, and cancer. Rambam Maimonides Med J. 2012;3:e0014.
49. Wilcox AJ, Baird DD, Weinberg CR. Do women with childhood exposure to cigarette smoking have increased fecundability? Am J Epidemiol. 1989;129:1079–83.
50. Weinberg CR, Wilcox AJ, Baird DD. Reduced fecundability in women with prenatal exposure to cigarette smoking. Am J Epidemiol. 1989;129:1072–8.
51. Basso O, Wilcox AJ. Might rare factors account for most of the mortality of preterm babies? Epidemiology. 2011;22:320–7.
52. Ananth CV, Schisterman EF. Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics. Am J Obstet Gynecol. 2017;217:167–75.
53. Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. Am J Epidemiol. 2011;174:1062–8.
54. Snowden JM, Basso O. Causal inference in studies of preterm babies: a simulation study. BJOG. 2017; https://doi.org/10.1111/1471-0528.14942. [Epub ahead of print]
55. Suissa S. Immortal time bias in pharmaco-epidemiology. Am J Epidemiol. 2008;167:492–9.
56. Hutcheon JA, Kuret V, Joseph KS, Sabr Y, Lim K. Immortal time bias in the study of stillbirth risk factors:

the example of gestational diabetes. Epidemiology. 2013;24:787–90.
57. Hutcheon JA, Savitz DA. Invited Commentary: influenza, influenza immunization, and pregnancy – it's about time. Am J Epidemiol. 2016;184:187–91.
58. Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. Hum Reprod. 2007;22:1506–12.
59. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? Epidemiology. 2006;17:413–8.
60. Oerbeck B, Overgaard KR, Aspenes ST, Pripp AH, Mordre M, Aase H, et al. ADHD, comorbid disorders and psychosocial functioning: how representative is a child cohort study? Findings from a national patient registry. BMC Psychiatry. 2017;17:23.