



HAL
open science

A Systematic Comparison of Designs to Study Human Fecundity

Marinus J C Eijkemans, Henri Leridon, Niels Keiding, Rémy Slama

► **To cite this version:**

Marinus J C Eijkemans, Henri Leridon, Niels Keiding, Rémy Slama. A Systematic Comparison of Designs to Study Human Fecundity. *Epidemiology*, 2019, 30 (1), pp.120 - 129. <10.1097/ede.0000000000000916>. <hal-04652782>

HAL Id: hal-04652782

<https://hal.science/hal-04652782v1>

Submitted on 18 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

A Systematic Comparison of Designs to Study Human Fecundity

Marinus J. C. Eijkemans,^{a,b} Henri Leridon,^c Niels Keiding,^d and Rémy Slama^e

Background: Several epidemiologic designs allow studying fecundability, the monthly probability of pregnancy occurrence in noncontracepting couples in the general population. These designs may, to varying extents, suffer from attenuation bias and other biases. We aimed to compare the main designs: incident and prevalent cohorts, pregnancy-based, and current duration approaches.

Methods: A realistic simulation model produced individual reproductive lives of a fictitious population. We drew random population samples according to each study design, from which the cumulative probability of pregnancy was estimated. We compared the abilities of the designs to highlight the impact of an environmental factor influencing fecundability, relying on the Cox model with censoring after 12 or 6 months.

Results: Regarding the estimation of the cumulative probability of pregnancy, the pregnancy-based approach was the most prone to bias. When we considered a hypothetical factor associated with a hazard ratio (HR) of pregnancy of 0.7, the estimated HR was in the 0.78–0.85 range, according to designs. This attenuation bias was largest for the prevalent cohort and smallest for the current duration approach, which had the largest variance. The bias could be limited in all designs by censoring durations at 6 months.

Conclusion: Attenuation bias in HRs cannot be ignored in fecundability studies. Focusing on the effect of exposures during the first 6 months of unprotected intercourse through censoring removes part of this bias. For risk factors that can accurately be assessed retrospectively, retrospective fecundity designs, although biased, are not much more strongly so than logistically more intensive designs entailing follow-up.

Keywords: Cohort; current duration; fecundability; fecundity; pregnancy; prevalent cohort; simulation; time to pregnancy.

(*Epidemiology* 2019;30: 120–129)

Fecundity, the biologic ability to obtain a live birth, is markedly less efficient in humans than in most other mammalian species.^{1,2} Studies reported a temporal decrease in sperm parameters in some areas of industrialized countries,^{3–6} to an extent that may impact couples' fecundability, the probability for a pregnancy to start during a menstrual cycle with intercourse.⁷ Toxicologic and epidemiologic studies suggest that specific environmental^{8–11} lifestyle and behavioral factors can influence fecundity,^{12–14} but evidence from human studies is missing for many environmental factors. Therefore, there is a need for efficient approaches allowing monitoring of time trends in human fecundity and characterization of the influence of environmental and sociodemographic factors on fecundity.

The assessment of couples' fecundity can rely on several epidemiologic designs,^{15–18} and to our knowledge, no quantitative comparison of their efficiency is available. The case-control design has strong limitations when it comes to studying fecundability,^{19,20} and will not be considered further here. The remaining designs can be distinguished according to the time when couples are sampled with respect to the period of unprotected intercourse (eFigure 1; <http://links.lww.com/EDE/B404> and Table 1): if recruitment takes place after the end of this period, then the design usually corresponds to a pregnancy-based approach (if only periods of unprotected intercourse followed by a pregnancy are identified and recruited, as would be done in a study conducted in a maternity clinic).²¹ If recruitment takes place before the start of the period of unprotected intercourse and if couples are followed up, then the design corresponds to an incident cohort.²² From a cross-sectional sample of couples recruited during the period of unprotected intercourse, one can collect the time elapsed between the start of the period of unprotected intercourse and inclusion, from which the distribution of the underlying total duration of unprotected intercourse and the effect of covariates on fecundity can be estimated, which corresponds to the current duration approach, an approach that does not require follow-up.^{17,18,23} From the same population, if couples are

Submitted December 25, 2017; accepted September 3, 2018.

From the ^aJulius Center for Health Sciences and Primary Care, Department of Biostatistics and Research Support, University Medical Center, Utrecht, The Netherlands; ^bDepartment of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands; ^cINED (French Institute for Demographic Studies) and French Academy of Sciences, Paris, France; ^dDepartment of Biostatistics, University of Copenhagen, Copenhagen, Denmark; and ^eTeam of Environmental Epidemiology Applied to Reproduction and Respiratory Health, U1209, Inserm, CNRS and University Grenoble-Alpes Joint Research Center (IAB), Grenoble, France.

The authors report no conflicts of interest.

The computing code can be obtained by request to the first author.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Rémy Slama, Inserm, IAB Research Center, Team of Environmental Epidemiology, Allée des Alpes, Site Santé, 38700 La Tronche, France. E-mail: Remy.slama@univ-grenoble-alpes.fr.

Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/19/3001-0120

DOI: 10.1097/EDE.0000000000000916

TABLE 1. Characteristics of the Main Study Designs Considered

Study Design	Timing of Inclusion	Conditioning of Inclusion	Follow-up	Outcome of Interest	Censoring	Handling of Infertility Treatments
1) Incident cohort	Before the start of the PUI	Couple starts a PUI	Yes	Time until pregnancy	min(C, treatment, time gives up)	Censoring of duration at the treatment start
2) Prevalent cohort	During PUI	Couple is currently during a PUI	Yes	Time until pregnancy	min(C, treatment, time gives up)	Censoring of duration at the treatment start
3) Prevalent cohort with delayed entry limited to 6 months	During PUI	2) + PUI started <6 months ago	Yes	Time until pregnancy	min(C, treatment, time gives up)	Censoring of duration at the treatment start
4) Current duration	During PUI	2)	No	Time until end of PUI	At time C	
5) Current duration, treatments excluded	During PUI	4) + Couple has not started treatment at inclusion	No	Time until end of PUI	min(C, time gives up)	Couples with treatment before inclusion are excluded
6) Pregnancy-based design	After pregnancy detection	A pregnancy started and was detected	No	Time until pregnancy in fecund couples	min(C, treatment)	Censoring of duration at the treatment start
7) Pregnancy-based design, treatments excluded	After pregnancy detection	A pregnancy started without treatment and was detected	No	Time until pregnancy in fecund couples not resorting to medical treatments	min(C)	Exclusion of couples with treatment

C was set to 12 months in the main analysis and to 6 months in the secondary analysis.
C indicates censoring duration; PUI, period of unprotected intercourse.

followed up to monitor the time from inclusion until a possible pregnancy, then this corresponds to a prevalent cohort design, which is for fecundability studies generally restricted to couples who have initiated the period of unprotected intercourse a few months before inclusion.^{24–26}

These designs differ in many aspects. First, regarding the targets of inference, the pregnancy-based design can be considered as being biased in terms of estimation of fecundability, in that infertile couples are not included. The current duration approach targets the duration of the period of unprotected intercourse, corresponding to the minimum of time to pregnancy (TTP) and time to stop the period of unprotected intercourse. Second, the designs also strongly differ in terms of eligibility rates.^{18,22,23} Third, because the sampling takes place at different times with respect to the start of the at-risk period for various designs, the proportions of subjects with long durations (e.g., 1 year or more) differ between designs. This may have consequences in terms of attenuation bias if no censoring is applied.^{27,28} In our context, attenuation bias relates to the fact that, in a population that is heterogeneous in terms of fecundity, at each time, the most fecund couples are more likely to conceive a pregnancy, and hence to exit the population at risk of pregnancy, than less fecund couples. Consequently, in a study of the effect of a factor decreasing fecundability, as times goes by, the group not or least exposed to the factor (where fecundability is initially higher than in the exposed group) is more strongly depleted than the most exposed group. This leads to an attenuation over time of the time-specific ratio of the probability of pregnancy between the exposed and the unexposed groups.²⁷ Because of these and other differences between these designs, they may differ in terms of bias.

So far, comparisons between designs only considered the pregnancy-based and the incident cohort designs.^{29,30} They indicate that the pregnancy-based and the incident cohort designs differ in terms of ability to highlight the impact of a factor on fecundity.³⁰ To our knowledge, there is no systematic comparison taking into account more recently proposed approaches such as the prevalent cohort^{25,31} and current duration^{17,18,32,33} designs.

Our aim was to systematically investigate the amount of bias in the four designs mentioned, in terms of estimation of the cumulative probability of pregnancy and also in terms of ability to highlight the impact of an exposure factor influencing fecundity. We used a realistic simulation of the general population with age dependency and heterogeneity in fecundability.

METHODS

Study Population

We adapted a simulation model previously developed by Leridon.³⁴

The model simulates life histories of 1,000,000 women by randomly generating events during their life following realistic distributions. For each woman, age at onset of permanent sterility is drawn, and fecundability is assumed to vary with age (eFigure 2; <http://links.lww.com/EDE/B404>).³⁴

Starting from the age of 18 years, the model draws for each subject the emergence of a stable relationship, the start of attempts to become pregnant among those in a relationship, conception following the attempts and the pregnancy outcome, which can be a live birth or a miscarriage. Monthly

chances of these events, of the number of desired children and the distribution of birth spacing intervals, were drawn from distributions based on demographic data from France. Intervening events are also modeled: becoming widowed or becoming divorced.³⁴ Fertility treatments were also incorporated into the simulations. We assumed that couples start treatment in a deterministic way when they were not pregnant after a duration depending on the woman's age: 48 months before age 30 (i.e., an assumption that 100% of couples not pregnant within 4 years resort to infertility treatment), declining to 36 months at age 35, 24 months at age 40, and 12 months at 45. Irrespective of age, the duration before treatment start was 12 months for women sterile when trying to become pregnant, assuming that a clinical examination would be performed and would allow diagnosing sterility.³⁵

The main extensions to the original demographic simulation model^{34,35} were (1) couples would not persist indefinitely in their pregnancy attempt. A 10% yearly stopping rate (including divorce, modeled separately, see above) was assumed, independently of age.³⁶ (2) For each woman, date of birth has been defined, drawn uniformly between 1 January 1955 and 31 December 1999.

The reproductive life of each woman was divided into consecutive episodes, which were saved. Each episode was characterized by five attributes: its end-date; the matrimonial situation (single, married/having a stable relationship, widowed, divorced); the intention with respect to reproduction (not trying to become pregnant: no partner; not trying: enough children; not trying: too soon; trying; pregnant); the outcome in case of pregnancy (miscarriage or live birth); and the treatment status (yes/no). For each woman, a new episode is created when any one of these attributes changes owing to a randomly drawn variable.

Sampling According to the Compared Designs

From the simulated population of 1,000,000 individuals, we drew women according to each of the four study designs that we considered (eFigure 1; <http://links.lww.com/EDE/B404> and Table 1). The actual sampling was supposed to take place or start on one specific day.

The pregnancy-based approach focused on all women who had had a planned pregnancy (leading to a live birth or a spontaneous abortion) in the 15-year period before the inclusion date (assumed to be 1 July 2015); the outcome was TTP of the most recent pregnancy, defined as the duration elapsed between the start of the period of unprotected intercourse and date of conception. We did not consider unplanned pregnancies because their TTP is not defined. Some couples would retrospectively describe these pregnancies as planned in a pregnancy-based study, and possibly assign them a (short) TTP, which corresponds to pregnancy planning bias.³⁷ By ignoring these unplanned pregnancies, our model essentially assumes that there is no planning bias that would occur if the probability of having an unplanned

pregnancy was related to fecundability, which would influence all designs.

The current duration design corresponded to a cross-sectional sampling of all women having unprotected intercourse at the inclusion date (eFigure 1; <http://links.lww.com/EDE/B404>, and Table 1). The outcome was the current duration of unprotected intercourse, corresponding to the duration elapsed between the start of the pregnancy attempt and the inclusion date.

The prevalent cohort design relied on the population sampled in the current duration design, except that in this case it was followed up; the outcome was time to pregnancy, defined as the time elapsed between the start of the period of unprotected intercourse and the date of conception. We also implemented a prevalent cohort approach in which only couples with a current duration of less than 6 months at inclusion (the “recent initiators”) were eligible, a restriction used in several prevalent cohort fecundity studies.^{25,38}

The incident cohort study sampled all women wishing to start a period of unprotected intercourse within an accrual period of 12 months starting from the sampling date. For the prevalent cohort and incident cohort designs, length of follow-up was assumed to be 12 months, or less if a pregnancy occurred before this duration.

We assumed a lack of any measurement error for all four designs.

Estimation of the Cumulative Probability of Pregnancy

The eligibility rate of each design was defined as the ratio of the number of eligible women to the total number of women 18–44 years of age in the population (here, 1,000,000).

The cumulative probability of pregnancy was estimated using the Kaplan-Meier approach for the pregnancy-based and incident cohort designs, and Kaplan-Meier approach with left truncation (i.e., delayed entry) at the date of inclusion for the prevalent cohort design. For the current duration design, we used a parametric approach assuming an underlying generalized gamma distribution with censoring at 36 months.^{39,40}

TTP was in addition right censored at the date of occurrence of an infertility treatment in the pregnancy-based and in the two cohort designs; for the current duration analysis, we provide estimates in which couples with treatment were either excluded or included, as previously done²³; we have also repeated the pregnancy-based analysis excluding couples with treatment, instead of censoring them, to mimic what would happen in a society where fecundity treatments are not widespread. Pregnancy rates were provided at 3, 6, and 12 months and compared taking the incident cohort estimate as a reference.

As a basic test of the validity of our approach, we also simulated a population without fecundability heterogeneity, age dependence of fecundity, sterility, and stopping behavior. We estimated the corresponding survival curves, expecting

that, in this homogeneous population, all sampling designs would give identical results.

Estimation of the Impact of a Risk Factor on Fecundity

We assumed the existence of a risk factor A that impacted on fecundability without affecting the proportion of sterile couples; fecundability was multiplied by 0.7 in subjects exposed to A. The prevalence of exposure to factor A was assumed to be independent of all characteristics of the subjects and corresponded to 20% of the whole population. We drew subjects from a new simulated population according to each study design as indicated above, the probability of inclusion being independent of exposure status.

The impact of the risk factor was estimated using the Cox proportional hazard model, with delayed entry for the prevalent cohort analysis. The Cox model estimates hazard ratios (HR) of pregnancy, a value below one indicating that the probability of pregnancy is decreased in exposed compared with nonexposed subjects. For the pregnancy-based, incident, and prevalent cohort designs, we used a continuous Cox model with the Efron method to handle ties. Note that though the simulation is in units of months, the data are in units of days, because we generated birth dates with a resolution at the level of days. For the current duration approach, we used a modification of the “semiparametric” adaptation of the Cox model proposed by McLain et al.⁴¹ Observations were

censored at 12 months for the incident and prevalent cohort designs (which was the maximum length of follow-up we assumed). The Cox proportional hazards model is known to suffer from attenuation bias.^{27,28} We tried to limit the attenuation bias identified in early simulation runs by censoring observations at 6 months. For the current duration design, this was done by right-truncating the observed distribution at 6 months before applying the McLain algorithm. Models were run without and with adjustment for age.

RESULTS

Study Population

The eligibility rate was highest for the pregnancy-based design (35% of women 18–44 years of age were eligible, Table 2) and lowest for the current duration and prevalent cohort designs (about 2% of women). The current duration and prevalent cohort designs over-represented sterile couples (2.5% sterility for the current duration design), compared with an incident cohort (1.0% of couples were sterile, Table 2), an overrepresentation that is corrected for in the statistical analyses.

Cumulative Probability of Pregnancy

In the situation without sterility, age dependency, or any other source of heterogeneity in fecundability and without stopping behavior, all four designs showed as expected

TABLE 2. Eligibility Rate, Sample Composition, and Number of Events in the Population Studied According to Each of the Four Considered Study Designs: Simulation of 1,000,000 Women

Characteristics	Study Design				
	Incident Cohort	Current Duration	Prevalent Cohort	Prevalent Cohort <6 Months ^a	Pregnancy-based
Number of eligible women (%)	43,024 (4.3)	20,396 (2.0)	20,396 (2.0)	14,620 (1.5)	345,035 (35.0)
Age at start of attempt (years) ^b	26.8 (24.3, 29.9)	27.5 (24.7, 30.9)	27.5 (24.7, 30.9)	27.4 (24.7, 30.7)	27.8 (25.1, 30.8)
Number of children at inclusion					
No child	47.4%	49.1%	49.1%	48.3%	36.3%
1	38.7%	37.3%	37.3%	37.9%	44.9%
2	11.7%	11.4%	11.4%	11.6%	15.7%
>2	2.1%	2.2%	2.2%	2.2%	3.2%
Proportion of couples sterile at start of pregnancy attempt (n)	1.0% (460)	2.5% (513)	2.5% (513)	1.9% (285)	1.0% (3,556)
Fecundability of nonsterile couples at start of attempt ^b	0.22 (0.14, 0.31)	0.15 (0.08, 0.25)	0.15 (0.08, 0.25)	0.18 (0.11, 0.25)	0.22 (0.14, 0.29)
Exposure months ^b	2.2 (0.9, 4.5)	3.3 (1.3, 6.5)	3.3 (1.4, 6.4)	4.0 (1.5, 7.7)	3.1 (1.2, 7.3)
Proportion of couples with a pregnancy during follow-up (n)	60.0% (25,793)	0 (0)	57.1% (11,642)	69.6% (10,169)	85.0% (293,340)
Proportion of couples treated for infertility (n) ^c	0.8% (337)	0.5% (109)	1.3% (273)	1.3% (197)	5.9% (20,499)

^aPrevalent cohort approach restricted to recent initiators, i.e., to couples who have started the period of unprotected intercourse for less than 6 months at the time of inclusion.

^bMedian and (25th–75th) percentiles.

^cProportion of couples treated for involuntary infertility during the period “at risk of pregnancy”, i.e., between inclusion and end of follow-up (incident and prevalent cohorts); between start and end of the last period of unprotected intercourse leading to a pregnancy (pregnancy-based design) or between the start of the current period of unprotected intercourse and inclusion (current duration design).

Downloaded from http://journals.lww.com/epidem by BhdMfsePHkav1zEounf1tQIN4a+kLIEZgshHd4XMI0hCwwCX1 on 07/18/2024

TABLE 3. Bias of Estimates of the Cumulative Pregnancy Rates at Various Follow-up Times for Each Study Design, Including Couples Treated for Infecundity, Unless Stated Otherwise: Simulation of 1,000,000 Women

Study Design	Proportion Pregnant, % (95% CI)	Bias ^a
Incident cohort		
At 3 months	48.1 (47.7, 48.6)	0 (ref)
At 6 months	68.3 (67.9, 68.8)	0 (ref)
At 12 months	83.1 (82.7, 83.5)	0 (ref)
Current duration ^b		
At 3 months	49.1 (46.1, 53.2)	0.010
At 6 months	68.6 (66.9, 70.4)	0.003
At 12 months	83.6 (82.8, 84.4)	0.005
Current duration, excluding treated couples ^b		
At 3 months	49.4 (46.0, 52.6)	0.013
At 6 months	69.3 (67.3, 71.0)	0.009
At 12 months	84.8 (84.0, 85.6)	0.018
Prevalent cohort		
At 3 months	47.7 (45.8, 49.6)	-0.004
At 6 months	67.9 (66.6, 69.1)	-0.005
At 12 months	83.2 (82.5, 83.9)	0.002
Prevalent cohort <6 months ^c		
At 3 months	47.7 (45.8, 49.6)	-0.004
At 6 months	67.9 (66.6, 69.1)	-0.005
At 12 months	83.4 (82.6, 84.1)	0.003
Pregnancy-based		
At 3 months	49.0 (48.9, 49.2)	0.009
At 6 months	69.8 (69.6, 69.9)	0.014
At 12 months	85.4 (85.3, 85.5)	0.023
Pregnancy-based, excluding treated couples		
At 3 months	52.1 (51.9, 52.3)	0.040
At 6 months	74.1 (73.9, 74.2)	0.058
At 12 months	90.4 (90.3, 90.5)	0.073

^aDifference between the design-specific estimate and the reference value, as estimated by the incident cohort for the same duration of follow-up.

^b95% CI for the current duration design were estimated by bootstrapping.

^cPrevalent cohort approach restricted to recent initiators, i.e., to couples who have started the period of unprotected intercourse for less than 6 months at the time of inclusion.

very similar results (eFigure 3; <http://links.lww.com/EDE/B404>). When we considered a more realistic population with age dependence of fecundity, with additional heterogeneity in fecundability, with sterile couples and including stopping behavior, the cumulative pregnancy rates after 6 and 12 months of unprotected intercourse were 68.3% (95% confidence interval [CI], 67.9, 68.8) and 83.1% (95% CI, 82.7, 83.5, Table 3 and Figure 1) for the incident cohort design. The absolute bias in the estimated rate of 12-month involuntary infertility (the incident cohort being taken as a reference) was close to or below 2% for most designs, but for the pregnancy-based analysis in which couples with an infertility treatment were excluded; in this case, the rate of 12-month involuntary

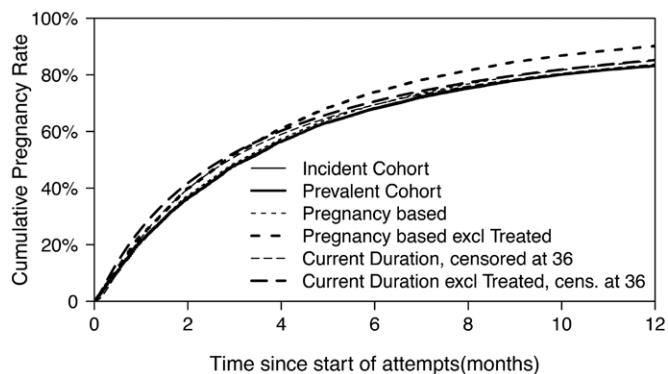


FIGURE 1. Cumulative pregnancy chances by four sampling approaches, on simulated data, in the simulation run including heterogeneity in fecundability, age dependency, and sterility.

infertility was underestimated by 7% (estimated rate, 10%, compared with 17% for the incident cohort). Including treated couples (with censoring) limited the bias of the pregnancy-based design (Table 3).

Impact of a Risk Factor on Fecundity

The HR associated with factor A tended to be attenuated (biased toward one) in all designs, the bias being lowest for the current duration design (age- and parity-adjusted HR, 0.78, 95% CI, 0.73, 0.84, compared with a theoretical value among fecund couples of 0.70) and highest for the prevalent cohort design (HR, 0.85, 95% CI, 0.81, 0.89, Table 4). The plot of the hazard rates allowed investigating the origin of this bias and showed that, both for the incident and prevalent cohort designs, the hazard ratio associated with exposure to A attenuated over time to reach a value close to 1.0 after 10–15 months of unprotected intercourse (Figure 2A). The number of subjects still trying to become pregnant was, as expected, maximum during the first month of unprotected intercourse in the incident cohort design, whereas it was lowest during the first month of unprotected intercourse and increased to reach a maximum for a follow-up of about 6 months in the prevalent cohort design, after which the number decreased because of the restriction to recent initiators (Figure 2B).

Censoring at 6 months led to decreases in the number of observations in the current duration design and in the prevalent cohort design without restriction to recent initiators (in which many couples were included after more than 6 months of unprotected intercourse) but not for the other designs (Table 4). HRs of pregnancy were all further away from the null value and closer to the theoretical value of 0.70 after censoring at 6 months: for example, the HR of pregnancy associated with factor A was 0.78 in the incident cohort design after censoring at 6 months, compared with 0.79 when censoring at 12 months (Table 4). The HR of pregnancy estimated by the pregnancy-based design decreased from 0.82 to 0.79, while that of the prevalent cohort decreased from 0.85 to 0.77, reaching a value very close to the incident cohort estimate

Downloaded from <http://journals.lww.com/epidem> by BhDMf5ePHkav1zEoum1tQIN44+kLUEZgshHd4XMI0hCwvCX1

TABLE 4. Estimates of the Effect of a Risk Factor A on the Probability of Pregnancy

Study Design	Sample Size	Number of Events ^a	Prevalence of Risk Factor A (%)	Hazard Ratio (95% CI) ^b	
				Not Adjusted	Adjusted
Censoring at 12 months					
Incident cohort	43,523	25,345	20.2	0.79 (0.76, 0.81)	0.79 (0.76, 0.81)
Current duration ^c	21,431	-	23.1	0.78 (0.73, 0.84)	0.78 (0.73, 0.84)
Current duration, excluding treatments ^c	20,825	-	23.0	0.78 (0.73, 0.84)	0.78 (0.73, 0.84)
Prevalent cohort	20,825	11,593	23.0	0.86 (0.82, 0.90)	0.85 (0.81, 0.89)
Prevalent cohort, entry <6 months ^d	14,725	10,037	22.1	0.85 (0.81, 0.89)	0.85 (0.81, 0.89)
Pregnancy-based	348,268	292,564	20.5	0.82 (0.81, 0.82)	0.82 (0.81, 0.83)
Censoring at 6 months					
Incident cohort	43,523	23,394	20.2	0.77 (0.75, 0.80)	0.78 (0.75, 0.80)
Current duration ^c	15,048	-	22.0	0.77 (0.68, 0.89)	0.72 (0.57, 0.92)
Current duration, excluding treatments ^c	14,725	-	22.1	0.75 (0.65, 0.86)	0.73 (0.58, 0.91)
Prevalent cohort	14,725	6,248	22.1	0.77 (0.73, 0.82)	0.77 (0.73, 0.82)
Prevalent cohort, entry <6 months ^d	14,725	6,248	22.1	0.77 (0.73, 0.82)	0.77 (0.73, 0.82)
Pregnancy-based	348,268	236,556	20.5	0.79 (0.78, 0.80)	0.79 (0.78, 0.80)

A hazard ratio of pregnancy below one indicates reduced fecundability. Simulation of 1,000,000 women, assuming that factor A reduced fecundability by a multiplicative factor 0.7.
^aEvents correspond to pregnancies. The “-” for the current duration approach indicates no pregnancies as in this design women are recruited during the period of unprotected intercourse without follow-up.

^bHazard ratio of pregnancy, as estimated from Cox proportional hazards model. Adjusted hazard ratios have been adjusted for woman's age at the start of the period at risk and parity.

^cThe hazard ratio in the current duration approach is derived from McLain's method.⁴¹ Observations with current duration above 12 (or 6) months are excluded instead of censored.

^dPrevalent cohort in which only couples who have been trying for less than 6 months at inclusion are recruited.

(HR, 0.78). The current duration estimate was the least biased (HR, 0.72), but had by far the widest confidence interval (0.57, 0.92, compared, e.g., with 0.75, 0.80 for the incident cohort).

DISCUSSION

This study is to our knowledge the first systematic comparison of most available models for the analysis of human fecundability in the general population. Within a realistically simulated population, we have applied four main different study designs to describe the fecundity level of the population and the impact of a risk factor on fecundity. In terms of ability to describe the fecundity level of the population, the pregnancy-based design tended, as expected, to over-estimate the fecundity level of the population, which was particularly obvious if couples with infecundity treatment were excluded or, equivalently, in populations where efficient infecundity treatments are not widely available. Regarding the ability to highlight the impact of a risk factor on fecundity, all four designs analyzed by proportional hazards regression suffered from attenuation bias. The attenuation was reduced by censoring analyses at 6 months follow-up.

Population

Our approach was based on a carefully simulated population that had many of the features of a real population.^{7,34,42}

The model assumed that it is possible to identify and recruit a random subsample of women before they start a period of unprotected intercourse (in the incident cohort design), which is something that not all couples plan long in advance.⁴³ We also assumed that one was able to identify and recruit all couples starting a period of unprotected intercourse

within a 1-year period, which may also be practically very challenging. For these reasons, the eligibility rate of the incident cohort approach is probably strongly overestimated. As an illustration, in Denmark, screening among 52,255 women members of a trade union, living as a couple, and 20–35 years of age allowed to recruit and follow-up 430 women, a rate of 0.8% encompassing ineligibility and refusals. This rate is much lower than the rate of 4% of women assumed to be eligible in the incident cohort design in our simulation.

More generally, we assumed a lack of selection bias, with the exception of the fact that the pregnancy-based design excluded infertile couples. This is optimistic; indeed, couples with long TTP might be more prone to participate in fecundity studies with collection of biological samples, compared with couples with shorter TTP.^{44,45}

We did not consider a retrospective design in which one would attempt to collect the TTP of unsuccessful attempts at pregnancy, in addition to that of periods of unprotected intercourse leading to a pregnancy. Identification of such unsuccessful attempts at pregnancy has been advocated¹⁶ and attempted.^{46,47} This corresponds to the so-called historically prospective cohort, which we expect to be less biased than the pregnancy-biased design implemented here. To efficiently consider the historically prospective cohort design in our simulation study, one would have needed realistic information on the structure of recall error of unsuccessful attempts at pregnancy and their duration. In practice, it seems relevant to try collecting information on these unsuccessful attempts in retrospective studies, if only to use them in sensitivity analyses.^{16,46,47}

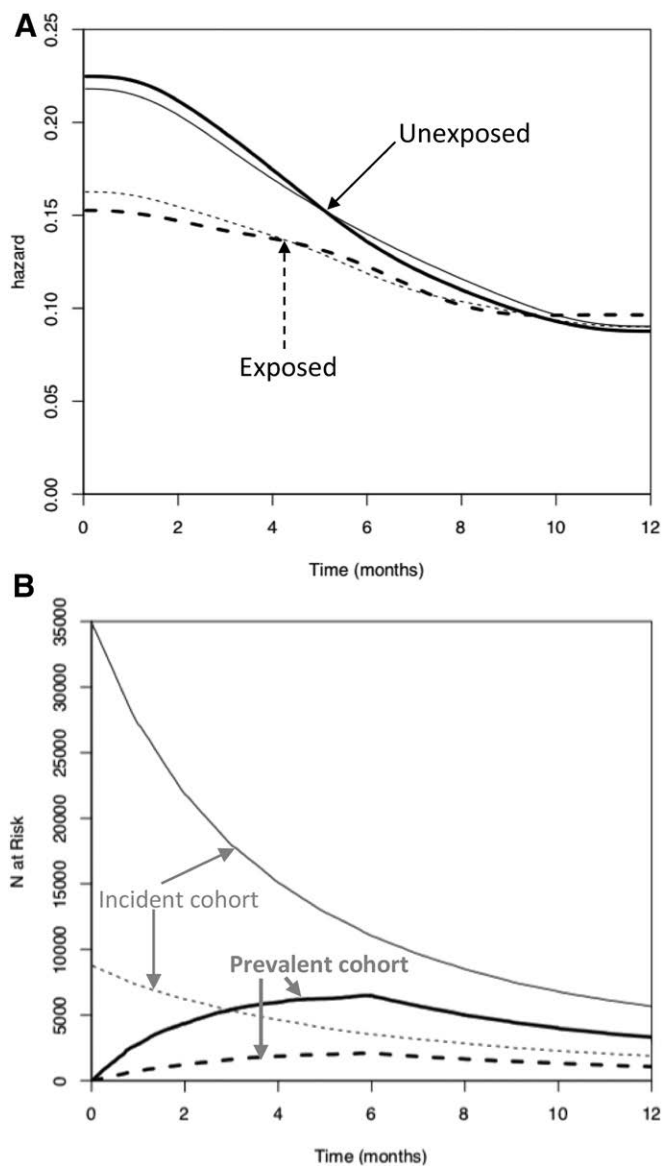


FIGURE 2. Comparison of the incident and prevalent cohort designs. (A) Hazard of pregnancy and (B) Number of women at risk of pregnancy among subjects exposed (dots) and not exposed (continuous curves) to factor A for the incident (thin lines) and prevalent (thick lines) cohorts designs. Exposure prevalence was assumed to be 20% in the source population.

Measurement Error and Confounding

We assumed that time to pregnancy was assessed without error, which is clearly optimistic for the pregnancy-based approach. Indeed, although the distribution of retrospectively assessed TTP has been shown to be relevant at the population level,⁴⁸ recall error exists at the individual level.⁴⁹ Such errors are likely to lead to decreased power in studies aiming at characterizing the effect of exposures on fecundity.⁵⁰ A simulation study characterized measurement error on TTP from the differences in TTP reported using a simple questionnaire and

a more refined one taken as a gold standard (thus probably underestimating measurement error compared with a perfect measure of TTP). Measurement error entailed a decrease in power by about 0.2 for a given population size and a bias in the estimated fecundability ratio by one fifth in the estimated fecundability reduction associated with exposure.⁵⁰ If applied to our data, this bias would increase the HR associated with factor A in the pregnancy-based design from 0.79 to 0.84, starting from a theoretical value of 0.70. In other words, with these assumptions, the pregnancy-based design would only see half of the real effect of the risk factor (a 16% instead of a 30% decrease).

Furthermore, we assumed a lack of misclassification for the exposure factor considered. Measurement error will generally be higher for the designs in which there is some retrospective component in the assessment of the exposure, such as the pregnancy-based and the current duration designs. This will, for example, be the case if exposure is assessed from a biospecimen, which can be collected at the start of the follow-up period in incident and prevalent cohorts designs but is collected retrospectively for the pregnancy-based and current duration designs. Biomarkers of many currently produced chemicals have short-term variations,⁵¹ which would preclude reliance on the pregnancy-based and current duration designs. Depending on the structure of the error, bias in the dose-response function can be expected as a result of exposure misclassification (e.g., in the case of classical-type error⁵²) and loss in statistical power (e.g., in the case of classical-type and Berkson-type errors).

Bias due to reverse causality can also be expected in these retrospective settings if one does not make efforts to assess exposure at the start of the period of unprotected intercourse; for example, couples may modify behaviors associated with exposure to factors suspected to alter fecundity after a long period of unprotected intercourse.³⁷ Bias may also exist in the case when there are time trends in exposure,⁵³ as may happen for behaviors recently identified as possibly harmful or beneficial or for contaminants with changing regulatory levels, such as atmospheric pollutants. This bias is of particular concern for designs in which the date of the start of the period of follow-up varies with time to pregnancy, which is the case for the pregnancy-based and current duration designs, but it can be avoided in incident and prevalent cohort designs. Our simulation assumed a lack of such temporal trends in exposure.

We adjusted for age and parity. The estimated effect of the risk factor was unaffected by controlling for effect of age, which is explained by the fact that we assumed independence of age and risk factor in the simulation. In real life, there will be more confounders. Again, designs in which confounders need to be assessed retrospectively may be more prone to residual confounding than designs in which confounders can be assessed prospectively such as the incident and prevalent cohort designs. This is all the more a concern since, in the setting of survival analyses, bias in the dose-response function

is expected not only if factors simultaneously associated with exposure and outcome are not adjusted for, but also if factors only associated with the outcome but not with exposure at the start of follow-up are not considered, or are measured with error.⁵⁴

Handling of Infertility Treatments and Other Censoring Events

The main censoring events considered were fecundity treatments and couples stopping the period of unprotected intercourse before the occurrence of pregnancy. One can adhere strictly to the original “fecundity” purpose of TTP studies—to obtain estimates of human (biological) fecundity in the absence of medical intervention—or take a pragmatic “fertility” view of just estimating ability to conceive in today’s society where fertility treatment is a reality. In this article, we presented results relating to both of these views. In the “fecundity” view, fertility treatment would be considered an endpoint competing with giving up and with becoming pregnant and they would need to be taken into account (by censoring as we did or using competing risk models).

Ability to Describe the Frequency of Involuntary Infertility

The comparison of the various designs should be done keeping in mind that there are differences in the outcomes assessed by each design (the targets of inference, Table 1). The incident and prevalent cohort designs focus on the occurrence of pregnancy; in the pregnancy-based design, the outcome is the occurrence of pregnancy in couples who eventually become pregnant. The current duration approach estimates the minimum of TTP and time to give up. This is shorter than TTP; a similar phenomenon is present in the pregnancy-based design.^{36,55}

For the pregnancy-based design, the estimated probability of involuntary infertility was, as expected, biased toward lower values; indeed, the conditional distribution of TTP, given that it is shorter than the time of giving up, is stochastically smaller than the marginal distribution of TTP. For the current duration design, our expectation was to see shorter estimates of the distribution of “length of trying” than those of TTP from the prospective designs, because of the couples giving up the pregnancy attempt. This is not apparent from Figure 1 and Table 3, where the differences are small so that this issue does not seem to entail consequences with the hypotheses on which our simulation relied.

Ability to Highlight an Effect of Risk Factors on Fecundability

When we simulated the impact of a risk factor influencing fecundability, the most striking feature of our results was the attenuation bias, which impacted all four designs. This was obvious from the fact that the estimated hazard ratios of pregnancy associated with the risk factor were generally in the 0.8–0.9 range while the theoretical value was 0.7, an

underestimation by one-third to two-thirds. Attenuation bias is a serious issue in survival analysis, and a rich literature exists for what we here describe as incident cohorts^{27,28} (see reference 56 for references). After a very long duration, the ratio of the hazard rates of pregnancy between exposed and unexposed subjects converges to one (i.e., a lack of observed effect of exposure), which is a manifestation of attenuation bias. This situation is obvious from Figure 2A, showing that, with the hypotheses of our study, this convergence of the hazard ratio to one is reached after about 10 months of unprotected intercourse. This suggests that durations longer than 9–10 months do not provide any information in terms of the possible effect of the exposure considered. Consequently, authors should see censoring not as a procedure entailing a loss of information but rather as an efficient way to limit bias. The fact that the attenuation was larger in the prevalent cohort than in the incident cohort can be explained by the fact that the number of subjects at risk was largest at the start of the period of unprotected intercourse for the incident cohort, when attenuation is weakest, while for the prevalent cohort, the number of subjects at risk was highest around month six, when attenuation bias was large (Figure 2B); because the overall estimate is a weighted average of the estimates based on the numbers at risk during follow-up, the prevalent cohort will give an estimate that is more biased than that of the incident cohort. Some studies relying on the prevalent cohort design have restricted eligibility to couples who have started the period of unprotected intercourse for 3 months or less, thus further limiting the oversampling of couples with a long duration at inclusion and the resulting attenuation bias.²⁴

Focusing on the start of the period at risk through censoring is a known²⁸ simple cure for attenuation bias. The convergence of the hazard ratio to one around 10 months led us to censor observations around month 6, but shorter censoring duration can be considered to further limit bias. For most designs but the current duration one (for which there were convergence issues with the estimator⁵⁷), further censoring at durations shorter than 6 months tended to slightly further reduce bias, at the cost of increases in the width of confidence intervals (not detailed). Following this logic, a study of atmospheric pollution effects on fecundability has focused on the first month of unprotected intercourse, disregarding later months.¹¹ Most pregnancy-based studies censor observations after 12 months, which on the basis of our simulation may not be enough.

We have used validated statistical approaches to analyze data. Other approaches are possible, including combining the data collected by different designs. For example, it may be possible to make use of the information of the current duration (elapsed between the start of the period of unprotected intercourse and the time of inclusion) when analyzing prospective TTP data from a prevalent cohort. An estimator allowing to combine prospectively collected and retrospectively collected time-to-event information, collected in the same subject, has

been proposed and applied in a study of pneumonia occurrence in HIV subjects.⁵⁸ To our knowledge, a similar approach has not been implemented in the field of fecundity studies yet.

In conclusion, the compared study designs show little bias in assessing the level of infertility in a population (with the exception of a pregnancy-based design excluding couples with infertility treatment), but require special care (e.g., through censoring analyses at 6 months or possibly earlier) to limit attenuation bias when analyzing the effect of an environmental risk factor using standard regression modeling. Disregarding other sources of bias, many previously published studies not censoring observations at short durations may have underestimated the effect, if any, of risk factors. Overall, if issues related to recall error and exposure misclassification due to a retrospective assessment of exposures are ignored, there appear to be no strong differences between designs in terms of bias. The current duration approach was, with our assumptions, the least biased in terms of estimation of the effect of a factor influencing fecundability; however, it was also the one with the largest variance. For factors that are likely to vary strongly over periods of several months (as would, e.g., be the case for many currently produced chemicals or biologic parameters such as hormonal levels) and in the absence of modeling approaches that would allow efficiently predict the level of these factors back in time, incident or prevalent cohort designs should be preferred; these are to be used with careful sensitivity or fine-tuning analyses to quantify and if possible remove attenuation bias.

ACKNOWLEDGMENTS

The authors thank Jean Bouyer (Inserm) for useful discussions at the initiation of this project.

REFERENCES

- Joffe M. What has happened to human fertility? *Hum Reprod.* 2010;25:295–307.
- Wood J. *Dynamics of Human Reproduction. Biology, Biometry, Demography.* New-York, NY: Aldine de Guyter; 1994.
- Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. *BMJ.* 1992;305:609–613.
- Auger J, Kunstmann JM, Czyglik F, Jouannet P. Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med.* 1995;332:281–285.
- Rolland M, Le Moal J, Wagner V, Royère D, De Mouzon J. Decline in semen concentration and morphology in a sample of 26,609 men close to general population between 1989 and 2005 in France. *Hum Reprod.* 2013;28:462–470.
- Swan SH, Elkin EP, Fenster L. Have sperm densities declined? A reanalysis of global trend data. *Environ Health Perspect.* 1997;105:1228–1232.
- Leridon H, Slama R. The impact of a decline in fecundity and of pregnancy postponement on final number of children and demand for assisted reproduction technology. *Hum Reprod.* 2008;23:1312–1319.
- Buck Louis GM, Sundaram R, Schisterman EF, et al. Heavy metals and couple fecundity, the LIFE Study. *Chemosphere.* 2012;87:1201–1207.
- Buck Louis GM, Sundaram R, Schisterman EF, et al. Persistent environmental pollutants and couple fecundity: the LIFE study. *Environ Health Perspect.* 2013;121:231–236.
- Chevrier C, Warembourg C, Gaudreau E, et al. Organochlorine pesticides, polychlorinated biphenyls, seafood consumption, and time-to-pregnancy. *Epidemiology.* 2013;24:251–260.
- Slama R, Bottagisi S, Solansky I, Lepeule J, Giorgis-Allemand L, Sram R. Short-term impact of atmospheric pollution on fecundability. *Epidemiology.* 2013;24:871–879.
- Jensen TK, Scheike T, Keiding N, Schaumburg I, Grandjean P. Fecundability in relation to body mass and menstrual cycle patterns. *Epidemiology.* 1999;10:422–428.
- Jensen TK, Henriksen TB, Hjollund NH, et al. Adult and prenatal exposures to tobacco smoke as risk indicators of fertility among 430 Danish couples. *Am J Epidemiol.* 1998;148:992–997.
- Ehrlich S, Williams PL, Missmer SA, et al. Urinary bisphenol A concentrations and implantation failure among women undergoing *in vitro* fertilization. *Environ Health Perspect.* 2012;120:978–983.
- Buck GM, Lynch CD, Stanford JB, et al. Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environ Health Perspect.* 2004;112:79–86.
- Joffe M, Key J, Best N, Keiding N, Scheike T, Jensen TK. Studying time to pregnancy by use of a retrospective design. *Am J Epidemiol.* 2005;162:115–124.
- Keiding N, Kvist K, Hartvig H, Tvede M, Juul S. Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics.* 2002;3:565–578.
- Slama R, Ducot B, Carstensen L, et al. Feasibility of the current-duration approach to studying human fecundity. *Epidemiology.* 2006;17:440–449.
- Weinberg CR, Wilcox AJ. Reproductive epidemiology. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology.* Washington: Lippincott-Raven; 2008:585–608.
- Guzick DS, Overstreet JW, Factor-Litvak P, et al.; National Cooperative Reproductive Medicine Network. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med.* 2001;345:1388–1393.
- Jensen TK, Slama R, Ducot B, et al. Regional differences in waiting time to pregnancy among fertile couples from four European cities. *Hum Reprod.* 2001;16:2697–2704.
- Bonde JP, Ernst E, Jensen TK, et al. Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *Lancet.* 1998;352:1172–1177.
- Slama R, Hansen OK, Ducot B, et al. Estimation of the frequency of involuntary infertility on a nation-wide basis. *Hum Reprod.* 2012;27:1489–1498.
- Wesselink AK, Rothman KJ, Hatch EE, Mikkelsen EM, Sorensen HT, Wise LA. Age and fecundability in a North American preconception cohort study. *Am J Obstet Gynecol.* 2017;217:e1–667.e8.
- Wise LA, Rothman KJ, Mikkelsen EM, Sørensen HT, Riis A, Hatch EE. An internet-based prospective study of body size and time-to-pregnancy. *Hum Reprod.* 2010;25:253–264.
- Jukic AMZ, Wilcox AJ, McConaughey DR, Weinberg CR, Steiner AZ. 25-hydroxyvitamin D and long menstrual cycles in a prospective cohort study. *Epidemiology.* 2018;29:388–396.
- Hernán MA. The hazards of hazard ratios. *Epidemiology.* 2010;21:13–15.
- Henderson R, Oman P. Effect of frailty on marginal regression estimates in survival analysis. *J Roy Stat Soc B.* 1999;61(part 2):367–379.
- Sallmén M, Lindbohm ML, Nurminen M. Paternal exposure to lead and infertility. *Epidemiology.* 2000;11:148–152.
- Slama R, Kold-Jensen T, Scheike T, Ducot B, Spira A, Keiding N. How would a decline in sperm concentration over time influence the probability of pregnancy? *Epidemiology.* 2004;15:458–465.
- Keiding N. Independent delayed entry (with discussion). In: Klein JP, Goel PK, eds. *Survival analysis: State of the Art.* Dordrecht, The Netherlands: Kluwer; 1992:309–326.
- Weinberg CR, Gladen BC. The beta-geometric distribution applied to comparative fecundability studies. *Biometrics.* 1986;42:547–560.
- Ali MM, Marshall T, Babiker AG. Analysis of incomplete durations with application to contraceptive use. *J Roy Stat Soc A.* 2001;164(part 3): 549–563.
- Leridon H. Can assisted reproduction technology compensate for the natural decline in fertility with age? A model assessment. *Hum Reprod.* 2004;19:1548–1553.
- Habbema JD, Eijkemans MJ, Nargund G, Beets G, Leridon H, Te Velde ER. The effect of *in vitro* fertilization on birth rates in western countries. *Hum Reprod.* 2009;24:1414–1419.
- Basso O, Juul S, Olsen J. Time to pregnancy as a correlate of fecundity: differential persistence in trying to become pregnant as a source of bias. *Int J Epidemiol.* 2000;29:856–861.

37. Weinberg CR, Baird DD, Wilcox AJ. Sources of bias in studies of time to pregnancy. *Stat Med*. 1994;13:671–681.
38. Hatch EE, Wesselink AK, Hahn KA, et al. Intake of sugar-sweetened beverages and fecundability in a North American Preconception Cohort. *Epidemiology*. 2018;29:369–378.
39. Yamaguchi K. Accelerated failure-time mover-stayer regression models for the analysis of last-episode data. *Sociol Methodol*. 2003;33:81–110.
40. Keiding N, Hojbjerg Hansen OK, Norbo Sorensen D, Slama R. The current duration approach to estimating time to pregnancy. *Scand J Stat*. 2012;39:185–204.
41. McLain AC, Sundaram R, Thoma M, Buck Louis GM. Semiparametric modeling of grouped current duration data with preferential reporting. *Stat Med*. 2014;33:3961–3972.
42. Leridon H. How effective is assisted reproduction technology? A model assessment. *Rev Epidemiol Sante Publique*. 2005;53(Spec No 2): 2S119–2S127.
43. Slama R, Ducot B, Keiding N, Bouyer J. Correspondence: studying human fertility and environmental exposures. *Environ Health Perspect*. 2004;112:A604.
44. Eustache F, Auger J, Cabrol D, Jouannet P. Are volunteers delivering semen samples in fertility studies a biased population? *Hum Reprod*. 2004;19:2831–2837.
45. Stewart TM, Liu DY, Garrett C, Brown EH, Baker HW. Recruitment bias in studies of semen and other factors affecting pregnancy rates in fertile men. *Hum Reprod*. 2009;24:2401–2408.
46. Slama R, Boutou O, Ducot B, Spira A. Reproductive life events in the population living in the vicinity of a nuclear waste reprocessing plant. *J Epidemiol Community Health*. 2008;62:513–521.
47. Karmaus W, Juul S; on behalf of the European Infertility and Subfecundity Study Group. Infertility and subfecundity in population-based samples from Denmark, Germany, Poland and Spain. *Eur J Public Health*. 1999;9:229–235.
48. Joffe M, Villard L, Li Z, Plowman R, Vessey M. Long-term recall of time-to-pregnancy. *Fertil Steril*. 1993;60:99–104.
49. Cooney MA, Buck Louis GM, Sundaram R, McGuinness BM, Lynch CD. Validity of self-reported time to pregnancy. *Epidemiology*. 2009;20:56–59.
50. Baird DD, Weinberg CR, Rowland AS. Reporting errors in time-to-pregnancy data collected with a short questionnaire. Impact on power and estimation of fecundability ratios. *Am J Epidemiol*. 1991;133:1282–1290.
51. Vernet C, Philippat C, Calafat AM, et al. Within-day, between-day and between-week variability of urinary concentrations of phenol biomarkers in pregnant women. *Env Health Perspect*. 2018;126:037005.
52. Perrier F, Giorgis-Allemand L, Slama R, Philippat C. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. *Epidemiology*. 2016;27:378–388.
53. Weinberg CR, Baird DD, Rowland AS. Pitfalls inherent in retrospective time-to-event studies: the example of time to pregnancy. *Stat Med*. 1993;12:867–879.
54. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regression and omitted covariates. *Biometrika*. 1984;71:431–444.
55. Keiding N, Hansen OKH, Sorensen DN, Slama R. The current duration approach to estimating time to pregnancy. *Scand J Stat*. 2012;39: 185–204.
56. Keiding N, Andersen PK, Klein JP. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Stat Med*. 1997;16:215–224.
57. McLain AC, Sundaram R, Thoma M, Buck Louis GM. Cautionary note on “Semiparametric modeling of grouped current duration data with preferential reporting”. Available at: <https://arxiv.org/abs/1801.00775v1>. Accessed 6 May 2018.
58. Cole SR, Chu H, Allison PD, Gange SJ. Combined analysis of retrospective and prospective occurrences in cohort studies: HIV-1 serostatus and incident pneumonia. *Int J Epidemiol*. 2006;35:1442–1446.