Comparison of gestational dating methods and implications for exposure–outcome associations: an example with $PM_{2.5}$ and preterm birth

Kristen M Rappazzo,[1] Danelle T Lobdell,[1] Lynne C Messer,[2] Charles Poole,[3] Julie L Daniels[3]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/oemed-2016-103833).

[1]National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Office of Research and Development, Chapel Hill, North Carolina, USA
[2]School of Community Health—College of Urban and Public Affairs, Portland State University Portland, Portland, Oregon, USA
[3]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

Correspondence to
Dr Kristen Rappazzo, U.S. EPA/ORD/NHEERL/EPHD, 109 T.W. Alexander Dr, Mail Code: 58C, Research Triangle Park, NC 27711, USA; Rappazzo.kristen@epa.gov

Received 9 May 2016
Revised 27 September 2016
Accepted 10 October 2016
Published Online First 25 October 2016

ABSTRACT

Objectives Estimating gestational age is usually based on date of last menstrual period (LMP) or clinical estimation (CE); both approaches introduce potential bias. Differences in methods of estimation may lead to misclassification and inconsistencies in risk estimates, particularly if exposure assignment is also gestation-dependent. This paper examines a 'what-if' scenario in which alternative methods are used and attempts to elucidate how method choice affects observed results.

Methods We constructed two 20-week gestational age cohorts of pregnancies between 2000 and 2005 (New Jersey, Pennsylvania, Ohio, USA) using live birth certificates: one defined preterm birth (PTB) status using CE and one using LMP. Within these, we estimated risk for 4 categories of preterm birth (PTBs per 10^6 pregnancies) and risk differences (RD (95% CIs)) associated with exposure to particulate matter ($PM_{2.5}$).

Results More births were classified preterm using LMP (16%) compared with CE (8%). RD divergences increased between cohorts as exposure period approached delivery. Among births between 28 and 31 weeks, week 7 $PM_{2.5}$ exposure conveyed RDs of 44 (21 to 67) for CE and 50 (18 to 82) for LMP populations, while week 24 exposure conveyed RDs of 33 (11 to 56) and −20 (−50 to 10), respectively.

Conclusions Different results from analyses restricted to births with both CE and LMP are most likely due to differences in dating methods rather than selection issues. Results are sensitive to choice of gestational age estimation, though degree of sensitivity can vary by exposure timing. When both outcome and exposure depend on estimate of gestational age, awareness of nuances in the method used for estimation is critical.

What this paper adds

- Choice of gestational dating method (clinical estimate or last menstrual period) may influence effect estimation through either outcome or exposure timing.
- Previous research has explored differences in demographic distributions of gestational dating methods, but there is limited work exploring alterations in effect estimates.
- In our study, divergences in risk differences were increased for outcome defined at later gestational ages, and also with later exposure weeks. Our results may indicate that the observed differences in risk differences are due to differences in the dating methods themselves rather than issues of selection.

INTRODUCTION

In research examining pregnancy-based outcomes and exposures, the estimation of gestational age (GA) is a critical step. This estimation is usually based on date of last menstrual period (LMP) or a clinical estimate of gestation (CE). There are benefits and detriments to either choice. LMP is simple, low cost, has a standard definition and is widely available. However, LMP may be inaccurate for a number of reasons: it may be differential based on length of time required for recall or by education level; it shows digit preference, does not capture menstrual irregularities, light bleeding or spotting during early pregnancy which may be mistaken for a period, and produces a bimodal distribution in birth weights.[1–6] CE is a mixture of ultrasound dating and neonatal dating, typically based on Ballard scoring;[3 7 8] it is not reliant on maternal recall, is unaffected by menstrual irregularities, and produces a single mode in birth weight distribution for births classified as preterm.[1 5] However, since CE is not a single method, it may be inaccurate in ways that are difficult to predict. Ballard scoring, an evaluation of neuromuscular and physical maturation, is performed within 96 hours of birth by an attending physician. This method is 'valid and accurate' but relies on typical maturation and tends to overestimate GA by 2–4 days for preterm births.[8] Ultrasounds are given as standard care during the first trimester prenatal care visits and are the most accurate method of gestational dating.[9] Since rates of prenatal care initiated in the first trimester have increased steadily since the 1990s,[10] CE are often based primarily on ultrasounds and therefore generally accurate. Even through 20 weeks of completed gestation, ultrasounds are considered accurate as there is limited variability in fetal growth during this period.[11] However, many women, ~16% in 2011, either do not initiate prenatal care until the third trimester or never, and so do not get an ultrasound for dating purposes;[10] usage of prenatal care and ultrasounds also differs by education, and accuracy of ultrasounds may differ by race/ethnicity.[3 7 10] Finally, ultrasounds

To cite: Rappazzo KM, Lobdell DT, Messer LC, *et al*. *Occup Environ Med* 2017;**74**:138–143.

have no way to detect if a fetus is unusually large or physically advanced for its actual gestational age.[11]

In studies of environmental exposures during pregnancy, gestational dating may affect exposure assignment and method differences may lead to differences in risk estimates. It is important to understand how the method of gestational dating used in analyses may affect estimated associations. In this analysis, we expanded on a previous work, which examined particulate matter ($PM_{2.5}$)–preterm birth (PTB) associations for live singleton births during 2000–2005 across three states (Pennsylvania, Ohio and New Jersey, USA),[12] by examining differences in associations when using alternate methods of GA dating.

METHODS

We constructed two analytic populations based on different availability of LMP or CE data in birth records. In one population, we assigned GA based on CE, and in the other, GA was based on self-reported LMP. In the CE method, the GA reported on birth certificates was used, and estimated LMP (for purposes of exposure assignment) was calculated by subtracting the CE from the date of birth. In the LMP method, the LMP reported was used and GA was calculated by subtracting LMP from date of birth. Both analytic populations were constructed from all live birth records provided by the State Health Departments of Pennsylvania, New Jersey and Ohio, USA for the years 2000–2005, described in detail in Rappazzo *et al*.[12] As in the previous analyses,[12] the populations were restricted to singleton pregnancies with no recorded birth defects and with geocodeable addresses. Included births were required to have achieved gestational week 20 no earlier than 1 January 2000 and gestational week 45 no later than 31 December 2005, which ensured that each pregnancy was entirely observable within the study period, and to avoid fixed-cohort bias. Pregnancies missing information on covariates were also excluded (maternal race/ethnicity, maternal education, marital status and maternal age at delivery).

From all birth records (n=2 495 350), these restrictions led to analytic populations of 1 781 527 pregnancies for the CE method, and of 1 592 478 pregnancies for the LMP method. Note that in these analytic populations, women are overlapping but not identical; for example, women in the CE analytic population may also have had a recorded LMP, but dating for the CE cohort was based solely on the CE. Inclusion of women based on CE was similar across the three states (proportion of women in the CE analytic population compared to initial birth records: New Jersey=78%, Ohio=76%, and Pennsylvania, USA=79%), though LMP-based inclusion varied (New Jersey=76%, Ohio=68%, and Pennsylvania, USA=66%). To further examine any differences in results, we also performed an analysis using only women who had both an LMP and a CE (n=1 588 595). This analysis allowed us to consider whether the observed differences may be due to population selection (ie, women with just one or the other dating method may be different, which could shift effect estimates) or whether observed differences may be due to differences in the dating methods and their influence on exposure and outcome assignment.

For the outcome, PTBs were divided into four categories based on definitions from the WHO:[13] extremely PTB (ExPTB) 20–27 completed gestation weeks; very PTB (VPTB) 28–31 completed gestation weeks; moderate PTB (MPTB) 32–34 completed gestation weeks; and late PTB (LPTB) 35–36 completed gestational weeks. Term births were between 37 and 45 completed gestational weeks.

Exposure to air pollution was assigned based on a monitor-corrected Community Multiscale Air Quality Model estimate of $PM_{2.5}$ exposure during pregnancy, as detailed in Rappazzo *et al*.[12] Model development is detailed in Hogrefe *et al*.[14] Exposures were assigned daily from LMP (estimated based on CE or reported as LMP) to the birth date, then averaged to weekly concentrations for each week of gestation.

Risk differences (RD) and 95% CIs for each PTB category were estimated for a 1 $\mu g/m^3$ increase in $PM_{2.5}$ using linear risk regression models, and were adjusted for: maternal race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other) education (<8th grade, some high school, high school diploma, some college, bachelor's degree, graduate school), marital status and maternal age at delivery (3-knot restricted quadratic spline). We used an at-risk approach in our models: all births that could have experienced birth during the weeks of interest were included in the model. For example, VPTBs were included in ExPTB models as non-events, but ExPTBs were not included in VPTB models. RD measures were used in this study as a continuation of previous work, and because we believe they are informative for public health impact and decision-making, and outcome severity. Effect measure modifiers were not considered in this study, as previous work showed no evidence for modification by infant sex, maternal parity, race/ethnicity, smoking status, state or region of birth, and census tract population density.[12] All analyses were performed using SAS V.9.3 (SAS Institute, . Cary, North Carolina, USA).

This research was approved by the University of North Carolina at Chapel Hill's Office of Human Research Ethics, the Pennsylvania Department of Health Bureau of Health Statistics and Research, New Jersey Department of Health and Senior Services Institutional Review Board, and the Ohio Department of Health Human Subjects Institutional Review Board. Informed consent was not required for this study as it was a secondary data analysis of existing data and no participant contact was attempted.

RESULTS

More births were classified as preterm in the LMP population (16%) than in the CE population (8%), and distributions of covariates between the two populations were similar based on visual inspection and 95% CI, which almost completely overlapped (table 1). Distribution of prenatal care initiation did vary somewhat, with the CE population missing more information (5.74% for CE, 3.40% for LMP, varying by outcome classification). Of the women not missing prenatal care information, ~90% initiated care before the fifth month of pregnancy; note that owing to birth certificate changes in 2003, PA prenatal care initiation for 2003–2005 births is estimated based on GA. The largest difference in GA category between the two populations appeared in post-term (≥43 weeks of completed gestation), with 992 births (0.06%) classified as post-term with the CE method and 57 114 births (3.6%) classified as post-term with the LMP method. For exposure, $PM_{2.5}$ distributions were slightly lower in the LMP population (~1 $\mu g/m^3$) than in the CE population (table 2).

For the population restricted to women with both an LMP and a CE, loss of participants from the initial analytic population was limited based on the LMP measure (n=3883 births, <0.5%), but fairly substantial based on the CE measure (n=192,932, >10%) (ie, almost all women who had an LMP also had a CE, while the reverse was not the case). Covariate distributions were similar across all analytic populations (see online supplementary materials eTable 2), and also among women with a CE but not LMP (data not shown). Preterm birth distributions in the restricted populations were similar to the distributions among the non-restricted populations.

Table 1 Distribution of demographic characteristics by gestational outcome classification and gestational dating method: Number (percentages)

	ExPTB		VPTB		MPTB		LPTB		Term	
	CE	LMP	CE	LMP	CE	LMP	CE	LMP	CE	LMP
	8664	10 789	12 004	18 525	31 446	56 493	90 037	167 865	1 639 376	1 338 806
Factor	(<1)	(<1)	(<1)	(1)	(2)	(4)	(5)	(11)	(92)	(84)
Maternal education										
Some college	3476 (40)	4220 (39)	5188 (43)	7931 (43)	14 535 (46)	25 972 (46)	44 635 (50)	87 942 (52)	901 061 (55)	754 669 (56)
High school diploma	3221 (37)	4062 (38)	4227 (35)	6458 (35)	10 789 (34)	19 187 (34)	29 566 (33)	53 639 (32)	491 888 (30)	394 839 (29)
<High school	1967 (23)	2507 (23)	2589 (22)	4136 (22)	6122 (19)	11 334 (20)	15 836 (18)	26 284 (16)	246 427 (15)	189 298 (14)
Maternal race/ethnicity										
Non-Hispanic white	4120 (48)	5311 (49)	6549 (55)	10 208 (55)	18 848 (60)	33 921 (60)	58 868 (65)	112 190 (67)	1 152 731 (70)	954 358 (71)
Non-Hispanic black	3279 (38)	3723 (35)	3671 (31)	4913 (27)	7781 (25)	12 076 (21)	17 034 (19)	27 627 (16)	225 430 (14)	167 757 (13)
Hispanic	256 (3)	372 (3)	447 (4)	817 (4)	1382 (4)	2653 (5)	4131 (5)	8406 (5)	83 507 (5)	70 379 (5)
Other	1009 (12)	1383 (13)	1337 (11)	2587 (14)	3435 (11)	7843 (14)	10 004 (11)	19 642 (12)	177 708 (11)	146 312 (11)
Maternal age										
<20	1354 (16)	1702 (16)	1549 (13)	2491 (13)	3549 (11)	6482 (11)	9133 (10)	14 545 (9)	139 192 (8)	105 099 (8)
20–29	4145 (48)	5196 (48)	5566 (46)	8705 (47)	14 748 (47)	26 872 (48)	43 259 (48)	78 884 (47)	790 998 (48)	642 307 (48)
30–39	2888 (33)	3539 (33)	4441 (37)	6666 (36)	11 971 (38)	21 178 (37)	34 657 (38)	68 740 (41)	664 625 (41)	555 450 (41)
40+	277 (3)	352 (3)	448 (4)	663 (4)	1178 (4)	1961 (3)	2988 (3)	5696 (3)	44 561 (3)	35 950 (3)
Marital status										
Married	3891 (45)	4822 (45)	6055 (50)	9539 (51)	17 481 (56)	32 611 (58)	55 582 (62)	110 762 (66)	1 118 053 (68)	939 300 (70)
Unmarried	4773 (55)	5967 (55)	5949 (50)	8986 (49)	13 965 (44)	23 882 (42)	34 455 (38)	57 103 (34)	521 323 (32)	399 506 (30)

CE, clinical estimate of gestation; ExPTB, 20–27 completed gestation weeks; LMP, date of last menstrual period; LPTB, 35–36 completed gestational weeks; MPTB, 32–34 completed gestation weeks; VPTB, 28–31 completed gestation weeks.

Table 2 Distribution of $PM_{2.5}$ ($\mu g/m^3$) concentrations, averaged over all weeks of exposure

	ExPTB		VPTB		MPTB		LPTB		Term births	
Statistic	CE	LMP	CE	LMP	CE	LMP	CE	LMP	CE	LMP
Min	3.73	2.62	3.55	2.4	3.15	1.94	2.84	1.72	2.45	1.39
25th	11	9.91	10.93	9.82	10.87	9.74	10.76	9.69	10.74	9.7
50th	13.8	12.73	13.72	12.61	13.65	12.51	13.54	12.45	13.51	12.46
75th	17.33	16.29	17.24	16.12	17.17	16.02	17.04	15.95	16.98	15.94
Max	50.82	48.84	50.87	52.36	53.33	54.13	55.19	55.9	58.25	58.43
Mean	14.62	13.56	14.54	13.43	14.47	13.34	14.36	13.28	14.31	13.28
SD	5.07	5.1	5.04	5.05	5.03	5.03	5.01	5.01	4.98	4.99
IQR	6.33	6.38	6.31	6.3	6.3	6.28	6.28	6.27	6.24	6.24

CE, clinical estimate of gestation; ExPTB, 20–27 completed gestation weeks; LMP, date of last menstrual period; LPTB, 35–36 completed gestational weeks; MPTB, 32–34 completed gestation weeks; VPTB, 28–31 completed gestation weeks.

Distributions of PTB category differed substantially between the two methods (table 3), with the Spearman correlation coefficient between PTB categories at 0.52.

Differences in RDs (PTBs per 10^6 pregnancies) between the two populations arise both by exposure window and by age of PTB grouping (figure 1, online supplementary materials eTable 1). Exposures in the later weeks of gestation appear to have the greatest divergence; though there is some variation, divergence primarily occurs in exposures after week 20. For example, the RD (95% CI) for VPTB with exposure at week 7 was 45.6 (22.5 to 68.7) for the CE cohort and 44 (12.4 to 75.7) for the LMP cohort, while with exposure at week 24 RDs were 33.4 (11.3 to 55.6) for the CE cohort and −20 (−50 to 10.3) for the LMP cohort. RDs also diverge with outcome classification of PTB, with differences appearing to increase with later GA outcomes. ExPTB, VPTB and MPTB show mostly similar associations, except in later weeks (eg, after week 20 in VPTB, after week 23 in MPTB) of exposure, whereas associations for the LPTB outcome are rarely in agreement between the two populations. For example, RDs for exposures in gestational week 24 were: for VPTB, LMP −20 (−50 to 10.3) and CE 35.9 (13.7 to 58); for MPTB, LMP −69 (−125 to −13) and CE 9.8 (−23 to 43.1); and for LPTB, LMP −30 (−129 to 70.2); CE 81.7 (16 to 147.4). When they diverge, RDs for the two populations are often in opposite directions, with RDs for the LMP analytic population being negative and RDs for the CE analytic population being positive. Results for the populations, restricted to women with both an LMP and CE, revealed patterns similar to those produced when all women were included for each GA dating method (figure 2, online supplementary materials eTable 1). RDs using the LMP cohort did not change between full and restricted analyses, while RDs using the CE cohort did shift somewhat, though not in any particular direction and not enough to change general patterns of association or interpretation of results.

DISCUSSION

In this analysis, we expanded on previous work examining differences in $PM_{2.5}$–PTB associations by using alternate methods of GA dating to create analytic cohorts. We found differences in associations by GA dating method based on both the exposure

Table 3 Distribution of preterm birth category assignment by clinical estimate of gestation method (CE, horizontal) and date of last menstrual period method (LMP, vertical) for population restricted to women with both methods

	CE					
LMP	**ExPTB**	**VPTB**	**MPTB**	**LPTB**	**Term**	**Total**
ExPTB	5946	1611	366	371	2155	10 449
VPTB	490	6580	4342	1280	5583	18 275
MPTB	115	806	16 946	19 113	19 123	56 103
LPTB	57	233	2269	43 543	121 290	167 392
Term	180	563	2395	13 764	1 319 474	1 336 376
Total	6788	9793	26 318	78 071	1 467 625	1 588 595

CE, clinical estimate of gestation; ExPTB, 20–27 completed gestation weeks; LMP, date of last menstrual period; LPTB, 35–36 completed gestational weeks; MPTB, 32–34 completed gestation weeks; VPTB, 28–31 completed gestation weeks.

window and outcome classification. For exposure, RDs with early exposure windows were more likely to be null when the LMP method was used, while RDs with later exposure windows were more likely to be negative when the LMP method was used. For outcome classification, RDs were most divergent across the two gestational dating methods for PTBs classified as LPTB, occurring between 35 and 36 weeks of GA.

The results arising from analyses based on the two dating methods may differ for two primary reasons. First, if the method used for dating GA resulted in different samples of women depending on the dating method, then differences may be due to covariate differences in the two samples, because the results between the main analyses that included all births with data available for a given method were similar to those for analyses restricted to only births with data available for both gestational dating methods. We did not see such differences, as results in the inclusive versus restricted analyses were the same.

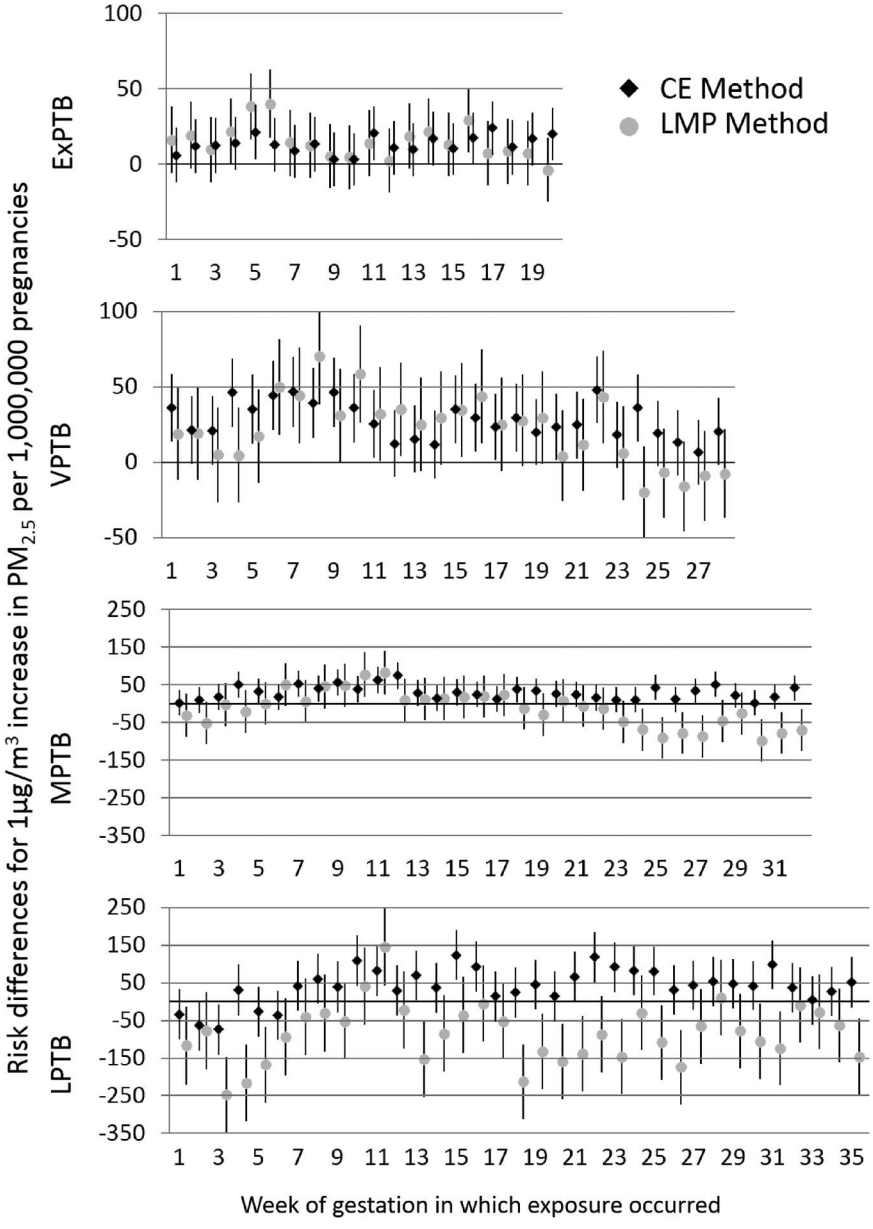

Figure 1 Risk differences per 1 000 000 pregnancies for 1 $\mu g/m^3$ increases in $PM_{2.5}$. Models adjusted for maternal education, race/ethnicity, marital status, age at delivery. Effects for clinical estimate (CE) population are shown in black diamonds, and for date of last menstrual period (LMP) population in grey circles. ExPTB: 20–27 completed gestation weeks; LPTB: 35–36 completed gestational weeks; MPTB: 32–34 completed gestation weeks; $PM_{2.5}$, particulate matter; VPTB: 28–31 completed gestation weeks.

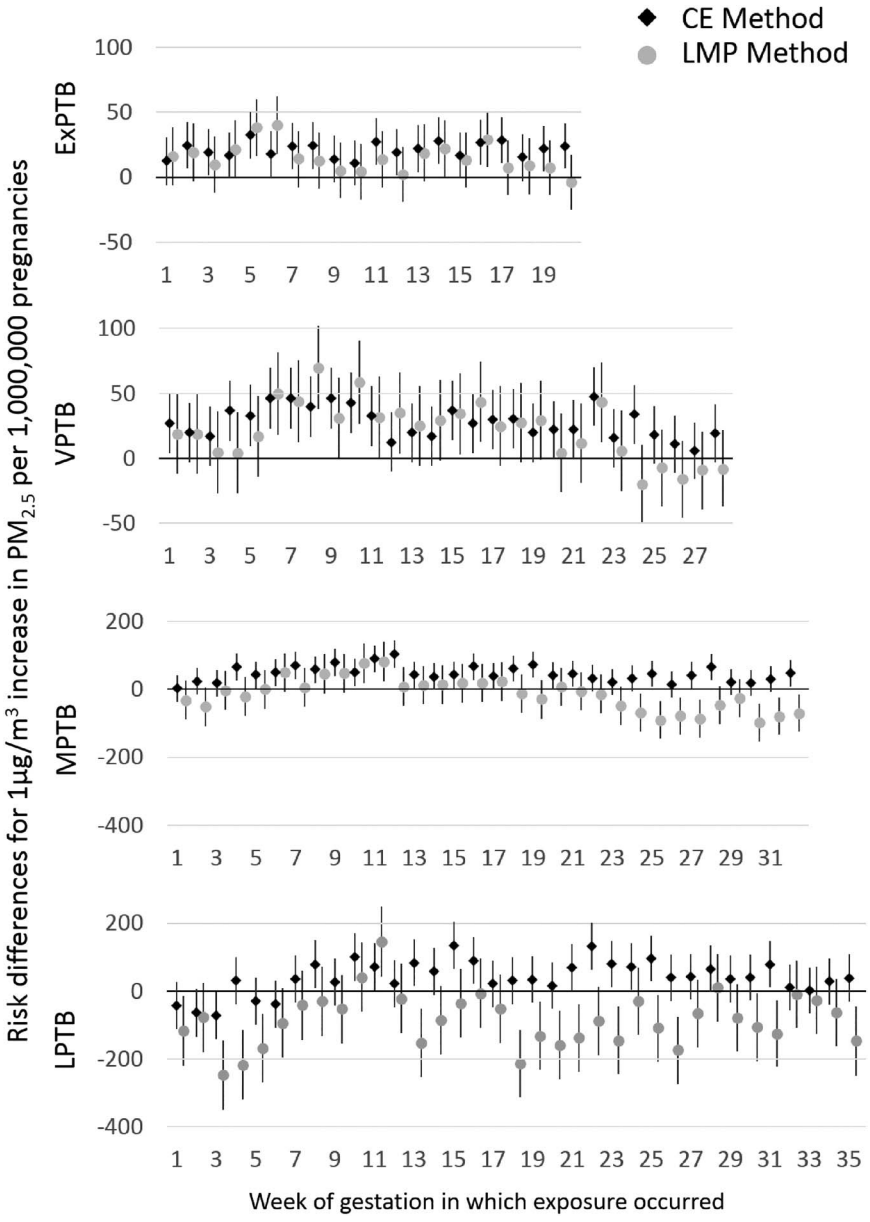

Figure 2 Risk differences per 1 000 000 pregnancies for 1 μg/m^3 increases in $PM_{2.5}$ for populations restricted to those births with both a clinical estimate (CE) and date of last menstrual period (LMP). Models adjusted for maternal education, race/ethnicity, marital status, age at delivery. Effects for clinical estimate (CE) population are shown in black diamonds, and for date of last menstrual period (LMP) population in grey circles. ExPTB: 20–27 completed gestation weeks; LMP, last menstrual period; LPTB: 35–36 completed gestational weeks; MPTB: 32–34 completed gestation weeks; $PM_{2.5}$, particulate matter; VPTB: 28–31 completed gestation weeks.

Another explanation, which seems more likely, is that there are underlying variances in the dating methods that lead to altered exposure or outcome classifications, which contribute to the divergence. The latter explanation is supported by the similarities between results for the main and restricted analyses.

In this work, the overall classification of preterm status had moderate agreement between the CE and LMP methods, but classification into preterm categories was very different for each method. For example, only 57% of those classified as ExPTB by CE were classified as ExPTB using LMP, with 21% of the others being classified as term births. Combining the LMP and CE methods has previously been used to fill data gaps that may be present when using either measure alone; however, using both the LMP and CE methods may lead to muddled results resulting in more bias than by using a single method.[7] A method introduced by Basso and Wilcox,[15] which combines LMP and CE using standards of birth weight to assess the plausibility of GA estimates, may offer an improved method to simple substitution.

Our analysis shows some similarities to previously published work in this area, such as more births being classified as post-term with the LMP method.[16] However, much of the research examining differences between the LMP and CE methods has focused on agreement between these methods, effects on metrics such as birth weight, or differences in demographic factors, but not the effects on specific exposure–outcome associations.[1 7 17 18]

We focus on identifying differences in the associations between $PM_{2.5}$ exposure and PTB when using the two methods of gestational dating, which demonstrates potential issues with, or effects of, method choice. While in many studies the gestational dating method will affect only the outcome under study, in our analysis, exposures are also affected. This will hold true for any exposure rooted in time, as timing of exposure is affected by the definition of pregnancy start and length. With air pollution exposures, particularly when looking over short-term periods, an inaccurate date will shift the concentration estimate; this is exemplified in our analysis in the $PM_{2.5}$ concentrations being shifted by 1 μg/m^3 between the two populations. While estimates may reflect general exposure levels due to temporal correlation, this shifting does introduce error. This complicates the choice of the GA metric because the impact of misclassification may be amplified, although misclassification is almost certain to be present in either CE or LMP dating alone.

Perturbations of development at particular times may lead to different outcomes, or different levels of outcome severity (eg, exposures early in pregnancy may lead to major birth defects, while later in utero exposures might lead to minor or no defects). In utero exposures to environmental contaminants have been linked to outcomes beyond birth, including childhood cancers and autism.[19 20] Understanding the critical timing of exposure to these contaminants relative to vulnerable developmental windows is an important part of understanding outcome aetiology.

Use of each GA dating method has benefits and limitations, with neither method being explicitly better or worse than the other. CE is not reliant on maternal recall, is unaffected by menstrual irregularities, includes ultrasound dating when available and produces a single mode in birth weight distribution for preterm births; it is also not a single method and may result in unpredictable inaccuracies.[1–3 5 6 11] LMP is simple, low cost, has a standard definition and is widely available; however, it is also likely to be inaccurate and may be differentially so.[21] While neither method is perfect, each has its advantages that should be considered when developing research questions. Both CE and LMP may lead to bias in effect estimates. In our analyses, biases associated with CE may potentially indicate that results are biased in unpredictable directions; we do not know what proportion of CE measures were (more accurate) ultrasounds before 20 weeks, versus ultrasounds after 20 weeks, or neonatal dating. That mixture could introduce outcome misclassification, and is likely to be differential by socioeconomic factors, and thus we cannot accurately predict the direction (or magnitude) of bias. This could also lead to inappropriate distribution of exposure assignment based on pregnancy start date, which would most likely induce exposure misclassification.

Biases associated with LMP may lead to results that are less generalisable to a broad population, as women with no recorded LMP or unreliable LMP (negative, longer than possible pregnancy) are removed (~12% of our initial population had missing or unreliable LMP). We observed divergence in preterm classification and in our estimated associations, and others have observed disagreement between methods with factors such as maternal age, odds of PTB for maternal race, or usage of prenatal care between CE and LMP methods.[7 16 22] Factors such as timing of prenatal care, ultrasound usage, and homogeneity or standards of neonatal practice are useful in the evaluation of the gestational dating method, though not always available to researchers. It would be informative for future research to break down CE into its main components, particularly if ultrasound-based or not, to examine if further differences exist. This information could lead to more accurate gestational dating in research and better effect estimation; unfortunately, those components are not currently recorded in registry data.

In this work, we observed higher effect estimates with CE dating, which may mean that there is the potential for underestimation of effects with LMP gestational dating. Substitution of the two methods for one another is likely to produce more bias in effect estimation than by use of a single method or a more complex combination algorithm. Awareness of methods used in current research, and differences in these methods even within the same person, may help evaluate potential uncertainties within that body of work and inform later policy decisions. In general, relevance for a specific research question should be taken into consideration when choosing the method of GA dating, and that method should be clearly indicated when presenting the study methods to help consumers evaluate inferences of one study within a greater body of literature.

Acknowledgements The authors would like to acknowledge Shannon Grabich and Thomas Luben for their thoughtful reviews and comments.

Contributors KMR conceived and designed the study and statistical analysis plan with substantial input from DTL, LCM, CP and JLD. KMR acquired, cleaned, linked and analysed data, and drafted and revised the manuscript. DTL acquired data, assisted with data interpretation and revised the manuscript. LCM, CP and JLD assisted with data interpretation and revised the manuscript. JLD supervised the study design and statistical analysis plan. KMR is the guarantor.

Funding This project was supported in part by an appointment to the Internship/Research Participation Programme at the Office of Research and Development, US Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and EPA.

Competing interests None declared.

Ethics approval University of North Carolina Chapel Hill, Office of Human Research Ethics.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Ananth CV. Menstrual versus clinical estimate of gestational age dating in the United States: temporal trends and variability in indices of perinatal outcomes. *Paediatr Perinat Epidemiol* 2007;21(Suppl 2):22–30.
2 Butler AS, Behrman RE. *Preterm birth: causes, consequences, and prevention*. National Academies Press, 2007.
3 Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol* 2007;21(Suppl 2):86–96.
4 Savitz DA, Hertz-Picciotto I, Poole C, *et al*. Epidemiologic measures of the course and outcome of pregnancy. *Epidemiol Rev* 2002;24:91–101.
5 Urquia ML, Moineddin R, Frank JW. A mixture model to correct misclassification of gestational age. *Ann Epidemiol* 2012;22:151–9.
6 Wegienka G, Baird DD. A comparison of recalled date of last menstrual period with prospectively recorded dates. *J Womens Health (Larchmt)* 2005;14:248–52.
7 Wingate MS, Alexander GR, Buekens P, *et al*. Comparison of gestational age classifications: date of last menstrual period vs. clinical estimate. *Ann Epidemiol* 2007;17:425–30.
8 Ballard JL, Khoury JC, Wedig K, *et al*. New Ballard Score, expanded to include extremely premature infants. *J Pediatr* 1991;119:417–23.
9 ACOG ACoOaG. Method for Estimating Due Date. Secondary Method for Estimating Due Date. 2015. http://www.acog.org/Resources-And-Publications/Committee-Opinions/Committee-on-Obstetric-Practice/Method-for-Estimating-Due-Date
10 HRSA (US Department of Health and Human Services HRaSA, Maternal and Child Health Bureau). *Child health USA 2007: prenatal care*. Secondary Child Health USA 2007: Prenatal Care, 2008. http://www.mchb.hrsa.gov/chusa07/hsfu/pages/309pc.html
11 Savitz DA, Terry JW, Dole N, *et al*. Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *Am J Obstet Gynecol* 2002;187:1660–6.
12 Rappazzo KM, Daniels JL, Messer LC, *et al*. Exposure to fine particulate matter during pregnancy and risk of preterm birth among women in New Jersey, Ohio, and Pennsylvania, 2000–2005. *Environ Health Perspect* 2014;122:992.
13 Organization WH. Preterm birth. Fact Sheet no. 363. Secondary Preterm birth. Fact Sheet no. 363 2015. http://www.who.int/mediacentre/factsheets/fs363/en/
14 Hogrefe C, Lynn B, Goldberg R, *et al*. A combined model–observation approach to estimate historic gridded fields of PM 2.5 mass and species concentrations. *Atmospheric Environment* 2009;43:2561–70.
15 Basso O, Wilcox A. Mortality risk among preterm babies: immaturity versus underlying pathology. *Epidemiology* 2010;21:521–7.
16 Hoffman CS, Messer LC, Mendola P, *et al*. Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. *Paediatr Perinat Epidemiol* 2008;22:587–96.
17 Callaghan WM, Dietz PM. Differences in birth weight for gestational age distributions according to the measures used to assign gestational age. *Am J Epidemiol* 2010;171:826–36.
18 Hall ES, Folger AT, Kelly EA, *et al*. Evaluation of gestational age estimate method on the calculation of preterm birth rates. *Matern Child Health J* 2014;18:755–62.
19 Kalkbrenner AE, Windham GC, Serre ML, *et al*. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology* 2015;26:30–42.
20 Vahter M. Health effects of early life exposure to arsenic. *Basic Clin Pharmacol Toxicol* 2008;102:204–11.
21 Pearl M, Wier ML, Kharrazi M. Assessing the quality of last menstrual period date on California birth records. *Paediatr Perinat Epidemiol* 2007;21(Suppl 2):50–61.
22 Lazariu V, Davis CF, McNutt LA. Comparison of two measures of gestational age among low income births. The potential impact on health studies, New York, 2005. *Matern Child Health J* 2013;17:42–8.