American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2023.

Vol. 192, No. 12
https://doi.org/10.1093/aje/kwad153
Advance Access publication:
July 4, 2023

Original Contribution

Large-Scale Data Harmonization Across Prospective Studies

The Preconception Period Analysis of Risks and Exposures Influencing Health and Development (PrePARED) Consortium

Ke Pan, Lydia A. Bazzano, Kalpana Betha, Brittany M. Charlton, Jorge E. Chavarro, Christina Cordero, Erica P. Gunderson, Catherine L. Haggerty, Jaime E. Hart, Anne Marie Jukic, Sylvia H. Ley, Gita D. Mishra, Sunni L. Mumford, Enrique F. Schisterman, Karen Schliep, Jeffrey G. Shaffer, Daniela Sotres-Alvarez, Joseph B. Stanford, Allen J. Wilcox, Lauren A. Wise, Edwina Yeung, and Emily W. Harville*

* Correspondence to Dr. Emily W. Harville, Department of Epidemiology, Tulane School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA 70112 (e-mail: eharvill@tulane.edu).

Initially submitted September 18, 2022; accepted for publication June 29, 2023.

The Preconception Period Analysis of Risks and Exposures Influencing Health and Development (PrePARED) Consortium creates a novel resource for addressing preconception health by merging data from numerous cohort studies. In this paper, we describe our data harmonization methods and results. Individual-level data from 12 prospective studies were pooled. The crosswalk-cataloging-harmonization procedure was used. The index pregnancy was defined as the first postbaseline pregnancy lasting more than 20 weeks. We assessed heterogeneity across studies by comparing preconception characteristics in different types of studies. The pooled data set included 114,762 women, and 25,531 (22%) reported at least 1 pregnancy of more than 20 weeks' gestation during the study period. Babies from the index pregnancies were delivered between 1976 and 2021 (median, 2008), at a mean maternal age of 29.7 (standard deviation, 4.6) years. Before the index pregnancy, 60% of women were nulligravid, 58% had a college degree or more, and 37% were overweight or obese. Other harmonized variables included race/ethnicity, household income, substance use, chronic conditions, and perinatal outcomes. Participants from pregnancy-planning studies had more education and were healthier. The prevalence of preexisting medical conditions did not vary substantially based on whether studies relied on self-reported data. Use of harmonized data presents opportunities to study uncommon preconception risk factors and pregnancy-related events. This harmonization effort laid the groundwork for future analyses and additional data harmonization.

consortia; data harmonization; preconception period; pregnancy; pregnancy complications

Abbreviations: aRD, adjusted risk difference; BMI, body mass index; CDE, common data element; CI, confidence interval; GDM, gestational diabetes mellitus; GH, gestational hypertension; NA, North American; PE, preeclampsia; PrePARED, Preconception Period Analysis of Risks and Exposures Influencing Health and Development.

Recent initiatives have focused on data harmonization as a key component of pregnancy-related research (1, 2). The Preconception Period Analysis of Risks and Exposures Influencing Health and Development (PrePARED) Consortium was formed in 2018 to address gaps in preconception research (3). Harmonizing multiple studies increases the utility of the data and improves statistical power to study uncommon exposures or outcomes. Well-harmonized data can also be used to evaluate effect modification, appropriately account for multiple confounding factors, increase the generalizability of study results, and boost power for population subgroups.

Despite these benefits, the potential of the harmonized data depends on the quantity and quality of the original

data (4). A major challenge is to determine the compatibility of collected data (5–7). Studies may vary in data sources (questionnaires vs. medical records or vital statistics data) and follow-up methods (in-person visits vs. mailings or online questionnaires), and data collection instruments can include questions which appear similar but are subtly different (8). Whether variables from questionnaires can be harmonized depends on the questions, what prompts are used, how ambiguous answers are coded, how missing data are treated, and how variables are constructed or derived after data have been collected (5).

Moreover, it can be challenging to harmonize studies with populations that differ by legal and governmental jurisdiction, eligibility criteria, sociodemographic characteristics, health-care system, and social context (9). For studies of pregnancy or preconception, another challenge is to account for the possibility of more than 1 pregnancy and the varying times between exposure assessment and pregnancy. Prospective studies of reproductive-age women usually enroll either individuals who are actively planning to conceive or individuals regardless of pregnancy intention (10). Enrolling individuals regardless of pregnancy intention allows capture of the natural history of reproductive events but creates more variability in exposure assessment relative to pregnancy, while in studies recruiting pregnancy planners only, participant eligibility may depend on time to pregnancy and fertility treatments (11–14). Changes in health habits and exposures may also be related to pregnancy intentions (15, 16).

Most of the literature has focused on analyzing harmonized data instead of describing harmonization processes and methods, even though these methods may affect study validity (5). This report describes the steps taken to harmonize data across studies in the PrePARED Consortium. In addition, we compare the prevalence of preconception risk factors and the incidence of adverse pregnancy outcomes by study type (studies that recruit pregnancy planners vs. studies that recruit participants regardless of pregnancy intent) and data collection modality (exclusively self-reports vs. medical records or a combination of medical/birth certificate records and self-reports).

METHODS

Study and participants

The PrePARED Consortium includes 12 studies pooling data from more than 114,762 female participants studied during the years 1973–2021. Detailed inclusion and exclusion criteria may be found elsewhere (3). Studies in the PrePARED Consortium include individuals who are actively planning pregnancy (referred to as "planners"), as well as studies examining reproductive-age women generally that also collect data on pregnancy characteristics and outcomes (for brevity, referred to as "unselected" participants, since they are an unselected group with respect to pregnancy intention). The studies that shared individual-level data through December 2021 were included in this article (see Table 1; more information about each study is provided in the Web Appendix, available at https://doi.org/10.1093/aje/kwad153). Although 3 studies collected data on couples, harmonization of data in the present article was limited to female participants.

For this article, we defined an overall sample and a birth group. First, participants from the general-population studies were included in the overall sample if they had at least 2 follow-up evaluations during ages 18–50 years. While this limited the overall sample to participants who were not lost to follow-up, it ensured that the preconception data were collected prospectively relative to pregnancy and that studies which enrolled children captured sufficient information in adulthood. We also included participants aged 18–50 years at enrollment from the studies of pregnancy planners. Next, we created a "birth group" to include participants who had an "index pregnancy" from all studies (Figure 1). To ensure that preconception information was captured, the "index pregnancy" for each participant was defined as the first pregnancy in which conception occurred after baseline and lasted more than 20 weeks or resulted in livebirth or stillbirth. In other words, participants who never became pregnant after enrollment or who became pregnant but whose pregnancies lasted less than 20 weeks were excluded from this group.

Harmonization approach

We applied the crosswalk-cataloging-harmonization process (17) (Figure 2). During the crosswalk step, we used a spreadsheet to document available variables from each cohort and organized them by data concept (e.g., education, income, tobacco use). During the cataloging step, we identified common data elements (CDEs) from PhenX (18) or the National Institutes of Health CDE Repository (https://cde.nlm.nih.gov/home) within each data concept for sociodemographic, lifestyle, and pregnancy-related baseline variables (Web Table 1). If a CDE could not be identified or an identified CDE could not be applied across studies, a definition was created to incorporate the maximum amount of information from each study (5).

During the harmonization step, when variables across studies were not identical, we used 2 different approaches: calibration—converting data into the same unit of measurement—or standardization, such as using study-specific quartiles. For example, information on numbers of alcoholic drinks (beer, wine, liquor) consumed was collected differently across studies (per week or per day), but data were harmonized to the same time period. For time-varying preconception covariates that were measured repeatedly, the latest measurement taken before the index pregnancy was used. The length of time between the preconception measurement and the start of pregnancy was also calculated.

Variables for harmonization

Table 2 shows data availability for variables that were not available across all studies.

Demographic characteristics. The following variables were harmonized for all participants in the overall sample.

Age at baseline. Age at enrollment was shared as a categorical variable in 1 study and as a continuous variable in all others. We converted the categorical data from the 1 study

Table 1. Studies Included in the PrePARED Consortium, 1973–Present

Cohort Study	Original Sampling Frame	Study Location(s)	Study Period	Only Persons Who Intended to Conceive?
ALSWH	Representative samples of Australian women	Australia (national)	1996–present	No
BHS	Black and White semirural men and women	Louisiana, United States	1973–present	No
CARDIA Study	50% Black and 50% White men and women	Alabama, Illinois, Minnesota, and California, United States	1985–present	No
EAGeR Study	Women with 1–2 prior pregnancy losses attempting pregnancy	Utah, New York, Pennsylvania, and Colorado, United States	2006–2012	Yes
EPS	Women from the Research Triangle area of North Carolina	North Carolina, United States	1982–2010	Yes
GUTS	Children of nurses in North America	United States (national)	1996–present	No
HCHS/SOL	Self-identified Hispanic/Latino men and women in the United States	New York, Illinois, Florida, and California, United States	2008–present	No
HOPE Study	Heterosexual couples attempting pregnancy	Utah, United States	2011–2015	Yes
LIFE India	Married women	Telangana State, India	2009–present	Yes
NHS3	Nurses	United States and Canada (national)	2010–present	No
PRESTO	Couples attempting pregnancy	United States and Canada (national)	2013–present	Yes
TTP Study	Women attempting pregnancy	Utah, United States	2003–2005	Yes

Abbreviations: ALSWH, Australian Longitudinal Study on Women's Health; BHS, Bogalusa Heart Study; CARDIA, Coronary Artery Risk Development in Young Adults; EAGeR, Effects of Aspirin in Gestation and Reproduction; EPS, Early Pregnancy Study; GUTS, Growing Up Today Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; HOPE, Home Observation of Periconceptional Exposures; LIFE India, Longitudinal Indian Family Health Pilot Study; NHS3, Nurses' Health Study 3; PrePARED, Preconception Period Analysis of Risks and Exposures Influencing Health and Development; PRESTO, Pregnancy Study Online; TTP, Time to Pregnancy in Couples of Proven Fecundity.

into a continuous variable by taking the midpoint of each age interval.

Race/ethnicity and country of current residence. Data on self-identified race and ethnicity were collected in North American (NA) studies only. In the Australian Longitudinal Study on Women's Health, participants were asked about the continent on which they were born (Australian-born, other English-speaking background, Europe, Asia, or other). Self-identified race/ethnicity and country of current residence were categorized as follows: NA Hispanic/Latina; NA non-Hispanic White; NA non-Hispanic Black or African-American; NA non-Hispanic Asian; NA non-Hispanic other race or multiracial; Indian (living in India); or Australian.

Education. Self-reported information on education was available in all studies, and the variable was harmonized as follows: less than 12 years of schooling (i.e., less than high school), 12 years of schooling (i.e., high school or equivalent), 13–14 years of schooling (i.e., associate's degree or some college), and more than 14 years of schooling (i.e., college degree or more). In the 2 studies where years of education was collected, it was converted into 4 categories according to the education system of the participant's country.

Income. Data on self-reported household income were available in 10 studies. To adjust for differences in currency across countries and between different years, self-reported annual household income from a single year and country were categorized into quartiles.

Health conditions, body mass index, and lifestyle behavioral factors. The variables listed below were harmonized for all participants in the overall sample. Whether information was self-reported or measured is summarized in Web Table 2.

Preexisting health conditions. Data on self-reported preexisting medical history of type 1 or type 2 diabetes and chronic hypertension, regardless of medication use, were available in 10 studies and were categorized as binary variables separately. In 2 studies, fasting blood glucose and glucose tolerance and blood pressure for diagnosis

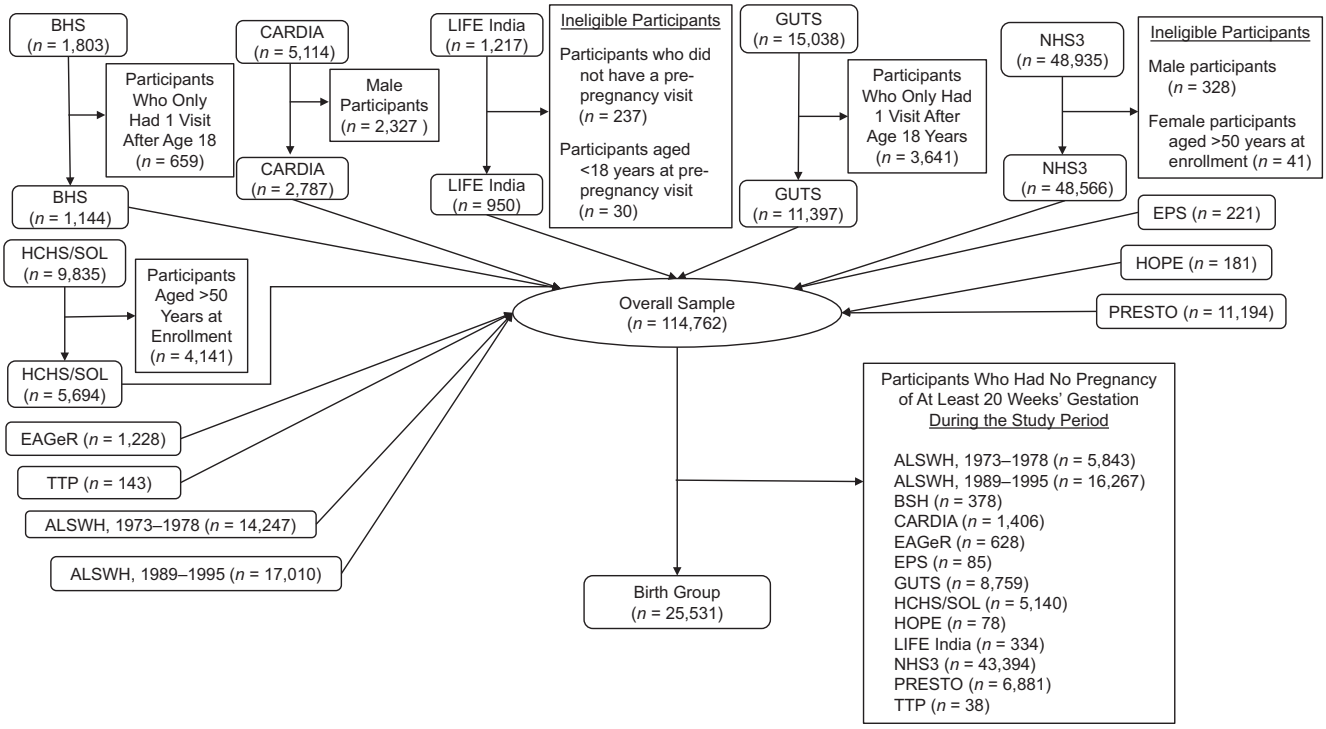

Figure 1. Compilation of the overall sample and the birth group in the Preconception Period Analysis of Risks and Exposures Influencing Health and Development (PrePARED) Consortium, 1973–present. ALSWH, Australian Longitudinal Study on Women's Health; BHS, Bogalusa Heart Study; CARDIA, Coronary Artery Risk Development in Young Adults; EAGeR, Effects of Aspirin in Gestation and Reproduction; EPS, Early Pregnancy Study; GUTS, Growing Up Today Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; HOPE, Home Observation of Periconceptional Exposures; LIFE India, Longitudinal Indian Family Health Pilot Study; NHS3, Nurses' Health Study 3; PRESTO, Pregnancy Study Online; TTP, Time to Pregnancy in Couples of Proven Fecundity.

of hypertension were measured and classified, which allowed differentiation of gestational diabetes mellitus (GDM) from overt diabetes and differentiation of preexisting chronic hypertension from gestational hypertension (Web Table 2).

Body mass index. Body mass index (BMI; weight (kg)/height $(m)^2$) was self-reported/self-measured in 5 studies (76% of eligible participants) and measured by research staff in the other 7 studies. Baseline BMI and the latest BMI before the index pregnancy were grouped into 6 categories: <18.5 (underweight), 18.5–24.9 (normal-weight), 25.0–29.9 (overweight), 30.0–34.9 (obese class I), 35.0–39.9 (obese class II), and ≥40.0 (obese class III).

Tobacco use. Data on self-reported history of tobacco use (cigarettes) before the index pregnancy were available in all studies. Tobacco use was categorized as never smoker, former smoker, or current smoker. Individuals who had never smoked more than 100 cigarettes or smoked regularly in their lives were categorized as never smokers; ever smokers were further categorized into former smokers and current smokers.

Alcohol use. Data on self-reported alcohol intake before the index pregnancy were available in all studies. Participants were asked about their alcohol intake patterns for the period of the past year (3 studies), the past month (5 studies), or the present ("now"; 3 studies). We categorized alcohol intake in terms of average daily volume of alcohol consumed: nondrinker, moderate drinker (≤7 standard drinks/week), or regular heavy drinker (>7 standard drinks/week) (19).

Recent cannabis use. Self-reported information on recent cannabis use was collected in 8 studies and was verified by urine sample in 1 of those studies (the Effects of Aspirin in Gestation and Reproduction (EAGeR) Study). Data included whether participants had used cannabis recently (in the past year or month; yes/no) and frequency of recent cannabis use (Web Table 3). Questions regarding recent cannabis use status varied across studies in terms of time period and frequency. To harmonize variables across studies, we assumed that recent cannabis use "now or over the past 1–3 months" represented the respondent's cannabis use status in the past year. We categorized information on cannabis use frequency in the past year into a 4-level variable: never, less than once a week, weekly but not daily, and daily.

Other drug use. Data on self-reported use of other drugs (apart from cannabis) were collected in 5 studies (Web Table 4). Use before the index pregnancy was categorized as a binary variable (yes/no). Four studies asked about other drug use status over the past 12 months, while 1 study asked about other drug use status in the past month.

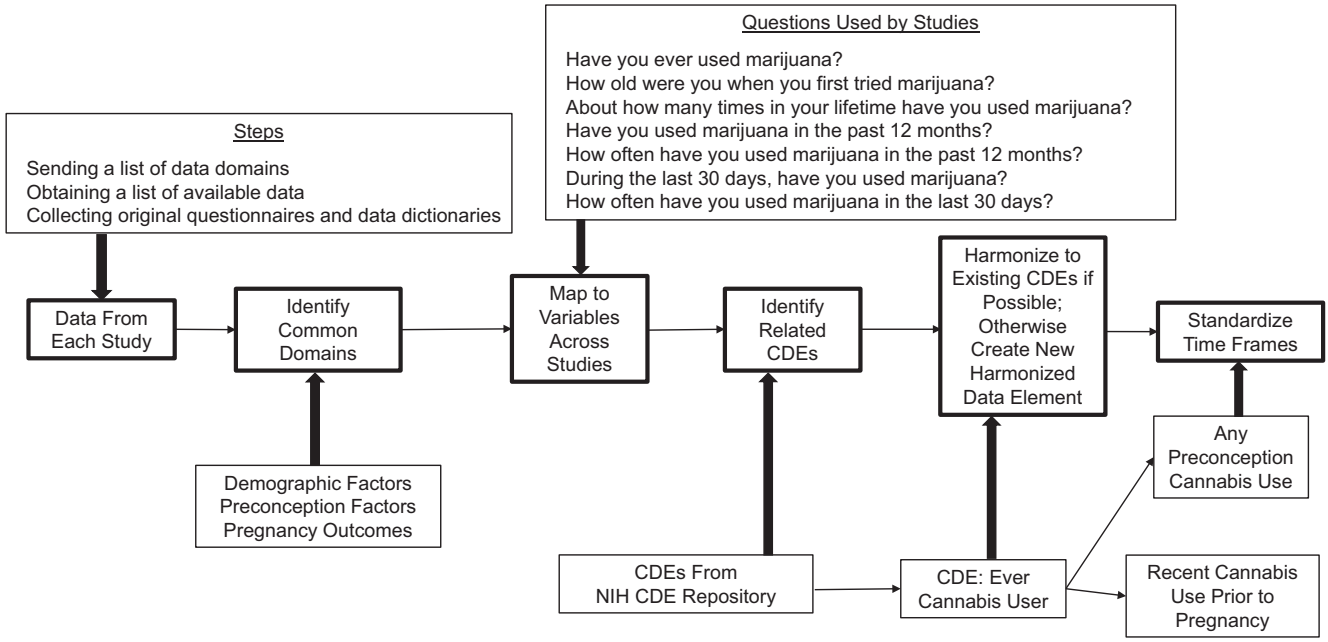

Figure 2. The crosswalk-cataloging-harmonization process used in the Preconception Period Analysis of Risks and Exposures Influencing Health and Development (PrePARED) Consortium, 1973–present. The example depicted pertains to the variable on cannabis use. CDE, common data element; NIH, National Institutes of Health.

Pregnancy-related variables. Pregnancy-related variables were defined for the birth group only.

Hypertensive disorders of pregnancy. Hypertensive disorders of pregnancy (binary variable) comprise gestational hypertension (GH) and preeclampsia (PE) (which may include some cases of eclampsia). As subtypes of hypertensive disorders of pregnancy, GH is defined by the new onset of hypertension at ≥20 weeks' gestation, while PE refers to preexisting or new-onset hypertension with proteinuria and/or significant end-organ dysfunction (20) (for details on the information available in each study, see Web Table 5). GH and PE were assessed exclusively via self-report questionnaires in 7 studies. Cases of GH or PE were identified by medical records directly (2 studies) or confirmed by validation studies in a subset via review of medical records (1 study) (21), birth certificates (1 study), and a combination of medical records and birth certificates (1 study).

Gestational diabetes mellitus. GDM (binary variable) is defined as glucose intolerance with onset or first detection during pregnancy (22). GDM was identified using medical records directly in 2 studies and assessed by self-report questionnaires in 10 studies. Among those studies that assessed GDM using self-report questionnaires, the GDM cases were confirmed by validation studies in a subset via review of medical records (1 study), birth certificates (1 study), and a combination of medical records and birth certificates (1 study) (for specific questions used for self-reporting in each study, see Web Table 6).

Data on the following variables were available in all studies and calculated for the index pregnancy.

Year of delivery. The calendar year of delivery was available in all studies and calculated for the index pregnancy.

Maternal age at delivery. Age at delivery was shared as a categorical variable in 1 study but as a continuous variable in other studies. We converted the categorical data from the one study into a continuous variable by taking the midpoint of each age interval to harmonize it with the age variable in other studies. Information was available in all eligible studies.

Parity or gravidity. Self-reported data on gravidity (number of pregnancies) and parity (number of births at ≥20 weeks' gestation) before the index pregnancy were available in all studies and treated as categorical variables (0, 1, 2, or ≥3).

Plurality. Self-reported or birth/medical record information on whether the pregnancy was a multiple-gestation pregnancy (yes) or a singleton pregnancy (no) at delivery was available in all studies.

Pregnancy intention. Self-reported data on pregnancy intention were available in 2 of the general-population studies. The variable was dichotomized as "yes" if participants reported actively trying to become pregnant for the index pregnancy and "no" otherwise. Individuals enrolled in a pregnancy planning cohort were categorized as "yes."

Statistical plan

After data harmonization, we used the mean value and standard deviation to describe normally distributed continuous variables (assessed by histograms and Q-Q plots);

Table 2. Availability of Data on Baseline Variables in the PrePARED Consortium, 1973–Present

Cohort Study	Availability of Baseline Data				
	Household Income	Pregnancy Intention	Cannabis Use	Other Drug Use	Preexisting Disease
ALSWH[a]					
Birth years 1973–1978	Yes		Yes	Yes	Yes
Birth years 1989–1995			Yes	Yes	Yes
BHS	Yes	Yes	Yes	Yes	Yes
CARDIA Study	Yes		Yes	Yes	Yes
EAGeR Study	Yes	Yes[b]	Yes	Yes	—[c]
EPS		Yes[b]	Yes		—[c]
GUTS					
GUTS I		Yes	Yes	Yes	Yes
GUTS II		Yes	Yes		Yes
HCHS/SOL	Yes				Yes
HOPE Study	Yes	Yes[b]			Yes
LIFE India	Yes	Yes			Yes
NHS3	Yes	Yes	Yes	Yes	Yes
PRESTO	Yes	Yes[b]	Yes		Yes
TTP Study	Yes	Yes[b]			Yes[d]

Abbreviations: ALSWH, Australian Longitudinal Study on Women's Health; BHS, Bogalusa Heart Study; CARDIA, Coronary Artery Risk Development in Young Adults; EAGeR, Effects of Aspirin in Gestation and Reproduction; EPS, Early Pregnancy Study; GUTS, Growing Up Today Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; HOPE, Home Observation of Periconceptional Exposures; LIFE India, Longitudinal Indian Family Health Pilot Study; NHS3, Nurses' Health Study 3; PrePARED, Preconception Period Analysis of Risks and Exposures Influencing Health and Development; PRESTO, Pregnancy Study Online; TTP, Time to Pregnancy in Couples of Proven Fecundity.

[a] ALSWH enrolled participants who were aged 18–23 years (birth years 1973–1978) when surveys began in 1996 and has enrolled an additional 17,000 women aged 18–23 years (1989–1995 cohort) since 2012.

[b] These cohort studies restricted the data to women who were actively planning to conceive.

[c] The study did not enroll individuals with preexisting diabetes or hypertension at baseline.

[d] The study did not enroll individuals with preexisting diabetes.

otherwise, we used the median value and interquartile range. Data for categorical variables are presented as percentages. For the overall sample, age-based probabilities of pregnancy during the study period were calculated using life-table methods to account for loss to follow-up. In order to assess heterogeneity across studies, adjusted for potential confounders (such as maternal age, calendar year at report, and length of follow-up), we used linear regression models to estimate mean differences (and 95% confidence intervals (CIs)) in the mean/percentage of preconception risk factors and adverse pregnancy outcomes according to 1) pregnancy intention (planners vs. general-population groups) and 2) data source (self-reported information vs. combined data sources (including self-reported information, medical records, and/or birth certificates) and measurement by study personnel). Nurses' Health Study 3 and the Growing Up Today Study were not included in regression models due to data-access restrictions (data can be analyzed on a central server but cannot be downloaded). Among all of the studies, only the Hispanic Community Health Study/Study of Latinos and the Australian Longitudinal Study on Women's Health had complex survey designs; we did not account for those survey designs and treated each study's data set as a convenience sample contributing to this pooled analysis. Analyses were conducted in SAS, version 9.4 (SAS Institute Inc., Cary, North Carolina).

RESULTS

Study population

Overall sample. Participants resided in the United States and Canada (n = 82,555), Australia (n = 31,257), and India (n = 950) (Table 1). Studies started between 1973 and 2013, with 7 ongoing at the time of data-pooling in December 2021. The overall sample included 114,762 participants, and the birth group included 25,531 participants (Figure 1). Most participants who did not report an index pregnancy

were younger than age 25 years (47%) or older than age 40 years (12%). The demographic characteristics of participants in the overall sample are presented in Table 3 and Web Table 7. Most participants identified themselves as non-Hispanic White (78%) or Hispanic/Latina (11%) (North America only; 82,341 participants (Table 3)). Most participants (58%) had a college or higher degree.

Birth group. Among participants with an index pregnancy (Table 4, Web Table 8), 1% had diabetes and 3% had hypertension before the index pregnancy. At enrollment, 3 studies intentionally excluded individuals who had type 1 or type 2 diabetes, and 2 studies excluded individuals who had chronic hypertension (Web Table 2). Before the index pregnancy (Table 4, Web Table 8), 22% of eligible participants who had an index pregnancy were overweight and 15% were obese, 16% were current smokers, and 10% had consumed more than 7 alcoholic drinks per week, on average, in the past year. Fifteen percent reported using cannabis in the past year, and about 8% reported use of other drugs in the past year.

Pregnancy-related variables

Among the 6 studies recruiting pregnancy planners only, the age-group–based pregnancy percentages during the study period (6–12 months) ranged from 93% (age 18–24 years) to 43% (age 40–44 years) (Web Table 9). Pregnancy was detected by an increase in urinary human chorionic gonadotropin level followed by an ultrasound (2 studies), a confirmed pregnancy test (2 studies), or self-reported information (8 studies) (Web Table 2). In the birth group, 60% of participants had never conceived previously (gravidity = 0), and 77% either had never conceived or had conceived but experienced only miscarriages (parity = 0). The mean maternal age at delivery in the index pregnancy was 29.70 (standard deviation, 4.61) years, and infants were delivered between 1976 and 2021 (median, 2008). Six percent of pregnant participants developed GDM during the index pregnancy, while 11% developed GH or PE during the index pregnancy. Among studies that differentiated between GH and PE, 8% and 6% of participants with an index pregnancy developed GH and PE, respectively.

Pregnancy intention

Tables 3 and 4 also show a comparison of variables among the 2 types of study populations (planners vs. unselected group). The studies from the planners group enrolled participants between 1982 and 2013 (median, 2008), while the studies from the unselected group started enrolling participants between 1973 and 2010 (median, 1996). A greater percentage of participants in the planners group identified as non-Hispanic White or Asian as compared with the unselected group. Educational attainment of participants tended to be higher (college graduation or more) in the planners group than in the unselected group (adjusted risk difference (aRD) = 36.64%, 95% CI: 35.16, 38.12).

In the birth group, participants in the planners group had their index pregnancy at slighter older ages (aRD = 0.95 years, 95% CI: 0.44, 0.86). Compared with the unselected group, the planners group had lower percentages of participants who were overweight or obese, had used cannabis more frequently in the past year, currently smoked, drank heavily in the past year, or had recently used other drugs before the index pregnancy (Table 4). The planners group also had lower percentages of participants who had preexisting diabetes or hypertension, partly because 3 studies excluded individuals who had hypertension or diabetes at enrollment (Web Table 2). The planners group had lower percentages of participants who were diagnosed with GDM or PE (Table 4) during the index pregnancy.

Source of data

Compared with studies using self-reported information, studies using laboratory results had more cases of diabetes and hypertension during study follow-up (diabetes: aRD = 4.95% (95% CI: 4.05, 5.84); hypertension: aRD = 15.05% (95% CI: 13.81, 16.28)) (Web Table 2; (Table 5). There were also higher percentages of participants with prepregnancy BMIs of 25–29 (overweight; aRD = 7.08%, 95% CI: 4.61, 9.55) and ≥30 (obese class I; aRD = 4.74%, 95% CI: 3.03, 6.45) in studies that measured height and weight by study personnel than in studies using self-reported information. Compared with the studies exclusively using self-reported information, there was a lower percentage of GDM cases and a higher percentage of GH/PE cases in studies that relied on multiple sources (e.g., medical records or a combination of self-reported outcomes and birth certificates/medical records) (Table 5).

DISCUSSION

This harmonization project provides a unique opportunity to assess preconception risk factors for reproductive and pregnancy health outcomes. By combining data across prospective studies involving individuals regardless of pregnancy intention and prospective studies involving individuals actively planning for pregnancy in the United States, Canada, Australia, and India, we generated a data set including more than 110,000 participants aged 18–50 years. More than 25,000 participants had at least 1 postbaseline pregnancy lasting more than 20 weeks with preconception information collected during active follow-up. Guided by a crosswalk-cataloging-harmonization process (17) and using CDEs from PhenX (18) or the National Institutes of Health CDE Repository (https://cde.nlm.nih.gov/home), we harmonized data on numerous variables, including demographic characteristics (age, race/ethnicity, education, household income), preconception reproductive history and exposure variables (parity, gravidity, BMI, cannabis use, tobacco use, alcohol use, other drug use, chronic conditions), and pregnancy-related variables (pregnancy intention, multiple gestation, gestational age at delivery, GDM, GH, and PE). We evaluated heterogeneity across studies by comparing variables with different target populations and studies using different primary data collection modalities, which will help investigators in future studies use and analyze the harmonized data appropriately. Lastly, the harmonized

Table 3. Demographic Characteristics of Participants in Studies Included in the PrePARED Consortium (Overall Sample), 1973–Present[a]

Variable	Participant Group								Difference Between Planners and Unselected Participants			
	Total (*n* = 114,762)		NHS3 and GUTS (*n* = 59,963)		Pregnancy Intention							
					Planners[b] (*n* = 13,917)		Unselected[c] (*n* = 40,882)		Unadjusted RD, %	95% CI	Adjusted RD, %	95% CI
	No.	%	No.	%	No.	%	No.	%				
Race/ethnicity[d]												
NA, Hispanic/Latina	9,353	11	2,826	5	833	6	5,694	59	−52.73	−53.7, −51.7		
NA, White, non-Hispanic	64,401	78	51,415	86	10,930	84	2,056	21	62.97	61.96, 63.90		
NA, Black, non-Hispanic	4,004	5	1,734	3	395	3	1,875	19	−16.43	−17.20, −15.67		
NA, Asian, non-Hispanic	1,764	2	1,515	3	249	2	0	0	1.92	1.65, 2.20		
NA, other, non-Hispanic	2,819	3	2,265	4	554	4	0	0	4.27	3.87, 4.68		
Education												
Less than high school	4,903	5	5	0	710	5	4,188	10	−5.18	−5.74, −4.63	−22.44[e]	−23.26, −21.61
High school	9,891	10	318	1	830	6	8,743	22	−15.56	−16.29, −14.84	−25.02[e]	−26.17, −23.87
Associate's degree or some college	28,342	27	14,178	29	3,160	23	11,004	27	−4.20	−5.05, −3.35	10.81[e]	9.47, 12.15
College degree or more	60,624	58	35,040	71	9,047	66	16,537	41	24.95	24.01, 25.89	36.64[e]	35.16, 38.12
Ever had diabetes	3,373	3	1,361	2	187	1	1,825	5	−3.13	−3.49, −2.77	−8.58[f]	−9.15, −8.0
Ever had hypertension	10,505	9	6,018	10	172	1	4,315	11	−9.34	−9.86, −8.82	−11.64	−12.42, −10.8
Smoked more than 100 cigarettes or regularly in life	37,354	33	16,013	27	3,058	22	18,283	45	−22.94	−23.86, −22.02	−1.90[f]	−3.35, −0.45
Ever drank alcohol heavily[g] in the past year	17,430	21	5,900	18	1,603	12	9,927	26	−14.01	−14.80, −13.22	7.43[f]	6.09, 8.76
Ever used other drugs	17,613	27	6,309	20	<5	0	11,300	36	−35.55	−38.24, −32.86	−45.35[f]	−48.89, −41.82

Abbreviations: ALSWH, Australian Longitudinal Study on Women's Health; BHS, Bogalusa Heart Study; CARDIA, Coronary Artery Risk Development in Young Adults; CI, confidence interval; EAGeR, Effects of Aspirin in Gestation and Reproduction; EPS, Early Pregnancy Study; GUTS, Growing Up Today Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; HOPE, Home Observation of Periconceptional Exposures; LIFE India, Longitudinal Indian Family Health Pilot Study; NA, North American; NHS3, Nurses' Health Study 3; PrePARED, Preconception Period Analysis of Risks and Exposures Influencing Health and Development; PRESTO, Pregnancy Study Online; RD, risk difference; TTP, Time to Pregnancy in Couples of Proven Fecundity.

[a] We compared variable distributions among studies that recruited pregnancy planners (the "planners" group) with those among studies that recruited individuals regardless of pregnancy intention (the "unselected" group). GUTS and NHS3 were not included in the comparison because of restrictions on data access.

[b] Planners group (study only recruited women who were planning pregnancy): EAGeR Study, EPS, HOPE Study, LIFE India, PRESTO, and TTP Study.

[c] Unselected group (study recruited women regardless of pregnancy intention): ALSWH, BHS, CARDIA Study, and HCHS/SOL.

[d] The NA American Indian/Alaska Native, NA Native Hawaiian or Pacific Islander, NA other, and NA mixed-race groups were combined as "NA other." ALSWH and LIFE India were not included in the race/ethnicity breakdown as "NA other" studies.

[e] Adjusted for year at baseline and country of residence.

[f] Adjusted for year at baseline, country of residence, and length of follow-up.

[g] Consuming more than 7 drinks/week, on average.

Table 4. Prepregnancy and Pregnancy Characteristics Associated With Participants' Index Pregnancies in the PrePARED Consortium (Birth Group), 1973–Present[a]

Variable	Time to Index Pregnancy, months[b]	Participant Group								Difference Between Planners and Unselected Participants			
		Total (n = 25,531)		NHS3 and GUTS (n = 7,810)		Pregnancy Intention							
						Planners[c] (n = 5,873)		Unselected[d] (n = 11,848)		Unadjusted RD, %	95% CI	Adjusted RD, %	95% CI
		No.	%	No.	%	No.	%	No.	%				
Race/ethnicity[e]													
NA, Hispanic/Latina		1,102	7	266	3	282	5	554	21	−15.14	−16.53, −13.76		
NA, White, non-Hispanic		12,875	82	7,099	91	4,586	87	1,190	44	43.21	41.37, 45.05		
NA, Black, non-Hispanic		1,107	7	64	1	86	2	957	35	−33.79	−35.17, −32.42		
NA, Asian, non-Hispanic		188	1	90	1	98	2	0	0	1.86	1.35, 2.38		
NA, other, non-Hispanic		486	3	283	4	203	4	0	0	3.86	3.14, 4.59		
Gravidity													
0		15,136	60	5,456	71	2,388	42	7,292	62	−20.07	−21.61, −18.53	4.99[f]	1.91, 8.07
1		5,285	21	1,279	17	1,692	30	2,314	20	9.94	8.62, 11.25	3.67[f]	1.00, 6.35
2		2,500	10	569	7	922	16	1,009	9	7.55	6.57, 8.53	1.88[f]	−0.11, 3.86
≥3		2,304	9	384	5	727	13	1,193	10	2.59	1.60, 3.57	−10.54[f]	−12.53, −8.54
Parity													
0		19,471	77	6,351	83	3,386	59	9,734	82	−23.25	−24.58, −21.92	−4.49[f]	−7.12, −1.87
1		3,977	16	844	11	1,671	29	1,462	12	16.76	15.58, 17.94	9.56[f]	7.18, 11.93
2		1,388	6	374	5	545	9	469	4	5.53	4.80, 6.26	0.13[f]	−1.36, 1.61
≥3		411	2	102	1	138	2	171	1	0.96	0.55, 1.37	−5.19[f]	−6.03, −4.35
BMI group[g]	10 (2–25)												
Underweight		1,164	5	218	3	355	6	591	5	0.83	0.11, 1.55	3.15[f]	1.66, 4.63
Normal-weight		14,184	58	4,562	64	3,013	52	6,609	59	−7.11	−8.67, −5.55	23.01[f]	19.69, 26.63
Overweight		5,309	22	1,473	21	1,320	23	2,516	22	0.25	−1.07, 1.57	−10.99[f]	−13.81, −8.18
Obese class I		2,101	9	531	7	613	10	957	8	1.99	1.08, 2.90	−7.89[f]	−9.83, −5.94
Obese class II		908	4	218	3	309	5	381	3	1.90	1.28, 2.52	−4.41[f]	−5.74, −3.09
Obese class III		592	2	136	2	238	4	218	2	2.14	1.63, 2.64	−2.86[f]	−3.95, −1.78
Education	11 (2–27)												
Less than high school		1,738	7	1	0	42	7	1,315	11	−3.83	−4.77, −2.89	−4.00[f]	−5.77, −2.24
High school or equivalent		3,631	15	81	1	263	5	3,287	28	−23.28	−24.49, −22.06	−17.28[f]	−19.73, −14.83
Associates or some college		5,081	21	1,205	17	1,111	19	2,765	23	−4.18	−5.48, −2.88	7.10[f]	4.41, 9.79
College degree or more		14,323	58	5,896	82	3,982	69	4,445	38	31.29	29.78, 32.79	14.18[f]	11.19, 17.18
Annual household income[h]	16 (1–23)												
Quartile 1		2,165	18	2	4	1,087	19	1,976	17	2.57	1.18, 3.95		
Quartile 2		2,921	24	13	27	1,793	32	1,115	18	14.59	13.07, 16.11		
Quartile 3		2,460	20	16	33	1,042	19	1,402	22	−3.36	−4.81, −1.91		
Quartile 4		4,456	37	18	37	1,663	30	2,775	44	−13.80	−15.52, −12.08		

Table continues

Table 4. (Continued)

Variable	Time to Index Pregnancy, months[b]	Participant Group								Difference Between Planners and Unselected Participants			
		Total (n = 25,531)		NHS3 and GUTS (n = 7,810)		Pregnancy Intention							
						Planners[c] (n = 5,873)		Unselected[d] (n = 11,848)		Unadjusted RD, %	95% CI	Adjusted RD, %	95% CI
		No.	%	No.	%	No.	%	No.	%				
Had preexisting diabetes	7 (1–24)	214	1	56	1	34	1	124	1	−0.47	−0.76, −0.17	−1.34[j]	−1.94, −0.75
Had preexisting hypertension	10 (2–26)	846	3	260	3	77	1	509	4	−2.99	−3.55, −2.43	−2.73[j]	−3.88, −1.57
Used cannabis in the past year		2,498	15	628	18	492	10	1,378	18	−7.78	−9.02, −6.54	−11.49[j]	−13.42, −9.57
Frequency of cannabis use in the past year	8 (1–22)												
None		8,632	86	2,934	82	4,559	90	1,139	83	7.45	5.58, 9.33	13.99[j]	9.87, 18.11
Less than once a week		888	9	508	14	286	6	94	7	−1.17	−2.58, 0.23	−2.04[j]	−5.14, 1.05
Weekly, but not daily		300	3	67	2	104	2	129	9	−7.33	−8.43, −6.23	−10.18[j]	−12.60, −7.75
Daily		161	2	52	1	97	2	12	1	1.05	0.28, 1.82	−1.77[j]	−3.46, −0.07
Tobacco use status[k]	9 (1–25)												
Never smoker		17,084	69	5,657	80	4,813	82	6,614	56	25.65	24.21, 27.10	6.37[j]	3.40, 9.35
Former smoker		3,606	15	904	13	651	11	2,051	17	−6.39	−7.51, −5.26	−0.75[j]	−3.07, 1.58
Current smoker		3,995	16	526	7	403	7	3,066	26	−19.27	−20.48, −18.05	−5.63[j]	−8.12, −3.13
Alcohol use status in the past year[l]	11 (1–27)												
No drinking		7,114	32	953	20	4,601	79	1,560	14	64.80	63.64, 65.96	23.27[j]	20.73, 25.81
Light drinking		12,585	57	3,456	73	1,088	19	8,041	71	−52.12	−53.49, −50.75	−10.66[j]	−13.69, −7.63
Heavy drinking		2,294	10	348	7	171	3	1,775	16	−12.68	−13.66, −11.71	−12.61[j]	−14.86, −10.35
Used other drugs in the past year	17 (7–29)	969	8	350	8	<20	0	617	9	−8.29	−10.54, −6.04	−10.67[j]	−14.56, −6.78
Maternal age at delivery		29.70 (4.61)[m]		31.02 (4.25)[m]		29.12 (4.16)[m]		29.11 (4.88)[m]		0.0124	−0.13.33, 0.1581	0.9500[n]	0.4392, 0.8607
Year of delivery		2008 (1976–2021)[o]		2014 (2000–2020)[o]		2017 (1983–2021)[o]		2004 (1976–2019)[o]		11.5687	11.3567, 11.7808		
Planned pregnancy		11,022	89	5,149	78	5,873	100						
Multiple pregnancy		476	2	202	3	50	1	224	2	−1.04	−1.43, −0.66		
Had a pregnancy-related condition in index pregnancy													
GDM		1,505	6	491	6	304	6	710	7	−0.44	−1.27, 0.40	−8.33[p]	−10.02, −6.65
GH or PE		2,688	11	864	11	530	11	1,294	12	−1.16	−2.24, −0.09	−1.14[p]	−3.33, 1.05
GH		1,051	8	557	8	343	8	151	8	0.09	−1.40, 1.58	2.13[p]	−0.18, 4.45
PE		798	6	451	6	198	5	149	8	−2.88	−4.10, −1.66	−2.28[p]	−4.20, −0.36

Table continues

Table 4. (Continued)

Variable	Time to Index Pregnancy, months[b]	Participant Group								Difference Between Planners and Unselected Participants			
		Total (n = 25,531)		NHS3 and GUTS (n = 7,810)		Pregnancy Intention							
						Planners[c] (n = 5,873)		Unselected[d] (n = 11,848)		Unadjusted RD, %	95% CI	Adjusted RD, %	95% CI
		No.	%	No.	%	No.	%	No.	%				
Ever had a pregnancy-related condition during study													
GDM		2,012	9	575	7	309	6	1,128	11	−3.70	−4.65, −2.75	−10.50[q]	−12.83, −8.17
GH or PE		3,114	13	957	12	538	11	1,619	15	−3.96	−5.11, −2.81	−0.46[q]	−3.23, 2.30
GH		1,337	9	630	8	351	8	356	11	−1.93	−3.23, −0.63	4.70[q]	1.87, 7.52
PE		1,143	7	522	7	198	5	423	13	−7.21	−8.41, −6.02	−1.12[q]	−3.71, 1.47

Abbreviations: ALSWH, Australian Longitudinal Study on Women's Health; BHS, Bogalusa Heart Study; CARDIA, Coronary Artery Risk Development in Young Adults; CI, confidence interval; EAGeR, Effects of Aspirin in Gestation and Reproduction; EPS, Early Pregnancy Study; GDM, gestational diabetes mellitus; GH, gestational hypertension; GUTS, Growing Up Today Study; HCHS/SOL, Hispanic Community Health Study/Study of Latinos; HDP, hypertensive disorders of pregnancy; HOPE, Home Observation of Periconceptional Exposures; LIFE India, Longitudinal Indian Family Health Pilot Study; NA, North American; NHS3, Nurses' Health Study 3; PE, preeclampsia; PrePARED, Preconception Period Analysis of Risks and Exposures Influencing Health and Development; PRESTO, Pregnancy Study Online; RD, risk difference; TTP, Time to Pregnancy in Couples of Proven Fecundity.

[a] We compared variable distributions among studies that recruited pregnancy planners (the "planners" group) with those among studies that recruited individuals regardless of pregnancy intention (the "unselected" group). GUTS and NHS3 were not included in the comparison because of restrictions on data access.

[b] Time between data collection and estimated start of the index pregnancy, for time-varying variables only; values are expressed as median (interquartile range).

[c] Planners group (study only recruited women who were planning pregnancy): EAGeR Study, EPS, HOPE Study, LIFE India, PRESTO, and TTP Study.

[d] Unselected group (study recruited women regardless of pregnancy intention): ALSWH, BHS, CARDIA Study, and HCHS/SOL.

[e] The NA American Indian/Alaska Native, NA Native Hawaiian or Pacific Islander, NA other, and NA mixed-race groups were combined as "NA other." ALSWH and LIFE India were not included in the race/ethnicity breakdown as "NA other" studies.

[f] Adjusted for year at birth and country of residence.

[g] BMI was calculated as weight (kg)/height (m)^2. BMI groups: underweight, BMI $<$18.5; normal-weight, BMI 18.5–24.9; overweight, BMI 25.0–29.9; obese class I, BMI 30.0–34.9; obese class II, BMI 35.0–39.9; obese class III, BMI $\geq$40.

[h] Income data collected in a single year and study were categorized into quantiles and harmonized with quantiles from other years and studies.

[i] Adjusted for year at report and country of residence.

[j] Adjusted for year at report and legality of cannabis use at report.

[k] Tobacco use status: never smoker, never having smoked more than 100 cigarettes or smoked regularly in one's life; former smoker, having smoked more than 100 cigarettes or smoked regularly in one's life, but not currently smoking; current smoker, having smoked more than 100 cigarettes or smoked regularly in one's life and currently smoking.

[l] Light drinking was defined as 1–7 drinks/week, on average; heavy drinking was defined as $>$7 drinks/week, on average.

[m] Values are expressed as mean (standard deviation).

[n] Adjusted for country of residence.

[o] Values are expressed as median (range).

[p] Adjusted for year at birth, country of residence, and maternal age at delivery.

[q] Adjusted for year at enrollment, length of follow-up, country of residence, and maternal age at delivery.

Table 5. Comparison of Information Exclusively Self-Reported With Information Not Exclusively Self-Reported[a] in the PrePARED Consortium, 1973–Present

Variable	Type of Data Source				Difference Between Exclusively Self-Reported and Not Exclusively Self-Reported Data[b]			
	Not Exclusively Self-Reported		Exclusively Self-Reported		Unadjusted RD, %	95% CI	Adjusted RD, %	95% CI
	No.	%	No.	%				
Had diabetes during study	390	14	1,622	3	10.77	10.05, 11.49	4.95[c]	4.05, 5.84
Had hypertension during study	1,178	32	3,309	7	24.84	23.94, 25.74	15.05[c]	13.81, 16.28
Prepregnancy BMI group[d]								
Underweight	380	9	566	4	4.92	4.12, 5.71	−1.63[e]	−2.94, −0.32
Normal	2,113	52	7,509	58	−6.17	−7.91, −4.43	−12.97[e]	15.90, −10.05
Overweight	915	22	2,921	22	−0.13	−1.60, 1.33	7.08[e]	4.61, 9.55
Obese class I	413	10	1,157	9	1.18	0.17, 2.19	4.74[e]	3.03, 6.45
Obese class II	164	4	526	4	−0.04	−0.73, 0.65	1.95[e]	0.78, 3.12
Obese class III	117	3	339	3	0.25	−0.32, 0.81	0.83[e]	−0.12, 1.79
Had preexisting diabetes before index pregnancy	14	1	144	1	0.08	−0.45, 0.61	−0.09[f]	−0.84, 0.66
Had preexisting hypertension before index pregnancy	125	6	461	3	3.18	2.33, 4.03	0.30[f]	−1.15, 1.74
Had a pregnancy-related condition in index pregnancy								
GDM	448	7	566	6	0.29	−4.89, 1.07	−4.39[f,g]	−6.16, −2.62
GH or PE	832	12	992	11	1.29	0.28, 2.30	6.54[f,g]	4.24, 8.84
GH	408	9	86	6	2.41	0.77, 4.05	6.60[f]	4.42, 8.77
PE	270	6	77	5	0.23	−1.13, 1.59	2.08[f]	0.27, 3.89
Ever had a pregnancy-related condition during study								
GDM	548	8	891	10	−2.49	−3.37, −1.61	−5.67[c,g]	−7.76, −3.58
GH or PE	970	14	1,189	13	1.16	0.08, 2.24	7.26[c,g]	4.72, 9.79
GH	626	10	92	7	3.15	1.46, 4.84	6.84[c]	4.53, 9.15
PE	543	8	86	6	2.39	0.83, 3.94	2.07[c]	0.04, 4.19

Abbreviations: BMI, body mass index; CI, confidence interval; GDM, gestational diabetes mellitus; GH, gestational hypertension; PE, preeclampsia; PrePARED, Preconception Period Analysis of Risks and Exposures Influencing Health and Development; RD, risk difference.

[a] Information not exclusively self-reported included information obtained directly from medical records or from a combination of medical/birth records and self-reports.

[b] The Growing Up Today Study and Nurses' Health Study 3 were not included in the comparison because of restrictions on data access.

[c] Adjusted for year at enrollment, country of residence, length of follow-up, and age at baseline.

[d] BMI was calculated as weight (kg)/height (m)^2. BMI groups: underweight, BMI <18.5; normal-weight, BMI 18.5–24.9; overweight, BMI 25.0–29.9; obese class I, BMI 30.0–34.9; obese class II, BMI 35.0–39.9; obese class III, BMI ≥40.

[e] Adjusted for year at report and country of residence.

[f] Adjusted for year at report, country of residence, and maternal age at delivery.

[g] If the 2 studies that validated case status by using birth certificates or a combination of medical and birth certificate records were excluded, results were as follows: 1) during the index pregnancy—for GDM, −3.83% (95% CI: −5.92, −1.74), and for GH/PE, 6.20% (95% CI: 3.44, 8.97); 2) ever had the condition during follow-up—for GDM, −6.68% (95% CI: −9.24, −4.13), and for GH/PE, 7.18% (95% CI: 4.10, 10.25).

data set provides an opportunity to evaluate uncommon exposures (e.g., cannabis use) or outcomes (e.g., PE) and effect modification during a crucial time (the preconception period) for setting the stage for lifelong wellness for both parents and offspring.

Recently, there have been calls for paying more attention to preconceptional and interconceptional health (23–28), especially as a strategy to reduce health disparities (29, 30). Due to studies' lack of information about the time period prior to pregnancy, limited sample size, or restricted

populations that included pregnancy planners only, little is known about the preconception period and its impact on both adverse pregnancy outcomes and future health outcomes (3). Different study designs have advantages and disadvantages: Studies of pregnancy planners exclude unintended pregnancies and might over- or underrecruit fertile couples, and studies focused on chronic health conditions tend to target an older population (3), while studies that recruit without regard to pregnancy planning may not have sufficient numbers of pregnancies or births. Thus, these types of studies provide important contextual comparison for each other. For instance, in the unselected group, there was often a specific focus on recruiting persons of non-White race/ethnicity (31–34), while the planners studies had relatively fewer recruits from minority racial/ethnic groups.

Preconception risk factors also differed. Planners were slightly older at delivery than the unselected group and were more likely to have a higher education, consistent with other studies (35). Moreover, the planners group had a higher percentage of participants with lower parity than the unselected group. Individuals planning pregnancy (those who had not reached their desired family size) may have been more likely to delay childbearing because of their education and occupation. The proportions of pregnancy in participants aged ≥25 years (range, 43%–70%) seemed slightly lower than what is typically reported (80%–90% being able to conceive within 12 months (35, 36)), which might indicate that individuals recruited for pregnancy planning studies include some people who had unprotected intercourse or used less effective methods of contraception in the past but did not conceive (subfertile individuals) (35).

In concurrence with our observations, pregnancy planning has been found to also be associated with healthier lifestyles (e.g., less smoking) (37, 38). Although we did not evaluate whether pregnancy planners actually changed their behaviors before pregnancy, this might suggest that the preconception period is a window of opportunity in which women are willing to adopt healthier behaviors (39, 40). Our project did not find an important difference in the prevalence of hypertension or diabetes before the index pregnancy between studies exclusively using self-reported information and studies relying on in-person examination measurements and blood sample tests or other sources, after adjusting for year of report, country where the studies were conducted, and maternal age. Participants for whom study personnel measured height and weight were more likely to be classified as overweight or obese than participants in studies using self-reported information, in line with previous findings that people tend to slightly overestimate their height and underestimate their weight, resulting in lower BMI (41).

Previous studies have shown that the reliability of self-reporting of adverse pregnancy outcomes is variable. Compared with studies exclusively using self-reported information, we found fewer GDM cases and more GH/PE cases during the index pregnancy in studies using objective record sources or a combination of self-reported and objective record sources. When previous studies have validated self-reports, GDM has generally been found to be more accurately reported than GH or PE; overall, self-reported GH/PE has a low sensitivity but a high specificity (21), and GDM is reported with both high sensitivity and high specificity (21, 42–46). Therefore, in future studies using the harmonized data set, researchers need to carefully assess the difference in study-specific or study-type–specific estimates before combining the results. In addition, diagnostic criteria and screening programs for GDM and GH/PE differ across populations or calendar years (47–49), and approaches to screening for GDM can differ even within a country within the same time period. Therefore, adjusting for calendar year and/or geographic region is recommended to evaluate associations involving GDM, GH, or PE. Moreover, use of self-reported pregnancy outcomes might have failed to capture women who had early pregnancy losses. Compared with individuals in studies recruiting women of reproductive age regardless of pregnancy intention (examining pregnancy within the life course but not focused on pregnancy), individuals recruited for pregnancy planning studies may be more aware of the onset of pregnancy because of better health practices and/or concerns due to earlier pregnancy problems. Therefore, the bias related to time of enrollment (relative to the period of unprotected intercourse) might point in unpredictable directions and might not be correctable with ordinary life-table analysis.

Another challenge was harmonization of variables for which data were collected using different questions and response categories. For instance, data on self-identified race/ethnicity were collected differently in Australia than in the United States. In the Australian Longitudinal Study on Women's Health, participants were asked not about their race/ethnicity but about where they were born: Australian-born, other English-speaking background, Europe, Asia, or other. Combining the Asian-American group with the Asian group in the Longitudinal Indian Family Health Pilot Study might not be appropriate, because race/ethnicity is a social construct and the lived experience of Asian women in these countries is likely to differ markedly (50). Therefore, the comparison of race/ethnicity across groups in our project was done among North American studies only. Additionally, although our project included studies in a range of geographic regions within the United States, Canada, Australia, and India, the majority of participants identified as White and resided in developed countries. In addition, there was a lack of standardization of questions and response categories in cannabis-related variables across studies in our project. Some studies asked about ever use of cannabis/age of first use and use in the recent past, while others asked about recent cannabis use status only. Moreover, the recent use question varied from current use to the past 12 months. Therefore, sensitivity analysis may be necessary to assess different definitions of "recent use." Lastly, although the harmonized data set included participants enrolled between 1973 and 2013 (overall study period), only 3 studies enrolled participants between 1973 and 1996 (4,152 participants; 3.62% of overall participants), and only 1 study enrolled participants between 1973 and 1982 (1,144 participants; 1.00% of overall participants). Therefore, the harmonized data set might be less representative of the targeted population in the earlier overall study period than in the later overall study period. In future studies, researchers analyzing the current

harmonized data set to understand the effect of preconception factors on pregnancy should consider calendar year of delivery to address the secular trends in environmental exposures, lifestyles, and quality of health care, and should be cautious in interpreting the generalizability of their results to earlier cohorts.

The strength of our consortium includes the large sample size and the carefully harmonized and curated data across multiple prospective studies. Historically, analyzing data from multiple studies was usually done by conducting standard meta-analysis among published results, which is also limited to specific research questions. Individual-level data across studies can be modeled simultaneously using standardized statistical analysis, adjusting for potential confounders consistently across studies, and analyzed using stratified analyses to evaluate cross-study or cross-nation differences. Moreover, overall and study-specific estimates can be reported to reduce publication bias. In our project, we harmonized individual-level data from studies that included participants regardless of their pregnancy intentions. Our project also suggests the need for standardized measures in future studies, such as recent cannabis use. Although care was taken to harmonize common variables across studies, residual differences likely remained due to how some of the questions were asked and what probes were used. Differences might be addressed through multiple questions and sensitivity/subgroup analysis.

The preconception, pregnancy, and lactation periods form a continuum influencing perinatal and future health outcomes (51). The comprehensive preconception risk profile has been neglected in most research studies, and while the participating studies have attempted to fill this gap individually (52–59), preconception data from this harmonization fill an important need for expansion of data sources to enhance risk stratification of individuals in relation to adverse pregnancy outcomes, identify infertility risk, and personalize targets for development of early interventions before and during pregnancy. By harmonizing studies to include both pregnancy planners and non–pregnancy planners, we can fully capture the preconception period in a more generalizable population, while also maximizing the sample size for pregnancies. The harmonized data in the PrePARED Consortium provide opportunities to evaluate preconception risk factors, and the assessment of heterogeneity across studies can guide the analytical plans of future studies.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Tulane School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, United States (Ke Pan, Lydia A. Bazzano, Sylvia H. Ley, Emily W. Harville); Society for Health Allied Research and Education (SHARE) INDIA, MediCiti Institute of Medical Sciences, Ghanpur, Telangana State, India (Kalpana Betha); Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, United States (Brittany M. Charlton); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Brittany M. Charlton); Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States (Brittany M. Charlton, Jorge E. Chavarro, Jaime E. Hart); Department of Nutrition and Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Jorge E. Chavarro); Department of Psychology, Miller School of Medicine, University of Miami, Miami, Florida, United States (Christina Cordero); Division of Research, Kaiser Permanente Northern California, Oakland, California, United States (Erica P. Gunderson); Department of Health Systems Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, California, United States (Erica P. Gunderson); Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States (Catherine L. Haggerty); Department of Environmental Health, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Jaime E. Hart); Epidemiology Branch, National Institute of Environmental Health Sciences, Durham, North Carolina, United States (Anne Marie Jukic, Allen J. Wilcox); Epidemiology and Biostatistics Division, School of Public Health, University of Queensland, Herston, Queensland, Australia (Gita D. Mishra); Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States (Sunni L. Mumford, Enrique F. Schisterman); Division of Public Health, Department of Family and Preventive Medicine, School of Medicine, University of Utah, Salt Lake City, Utah, United States (Karen Schliep, Joseph B. Stanford); Department of Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, United States (Jeffrey G. Shaffer); Department of Biostatistics and Collaborative Studies Coordinating Center, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (Daniela Sotres-Alvarez); Department of Epidemiology, School of Public Health, Boston University, Boston, Massachusetts, United States (Lauren A. Wise); and Epidemiology Branch, Division of Population Health Research, Division of Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Bethesda, Maryland, United States (Edwina Yeung).

Planning for the PrePARED Consortium was partially supported by National Institute of General Medical Sciences grant U54 GM104940 from the Louisiana Clinical and Translational Science Center, which is funded by the National Institutes of Health (NIH). The Australian Government Department of Health provided funding for the Australian Longitudinal Study on Women's Health. G.D.M. is a National Health and Medical Research Council Leadership Fellow (award APP2009577). The Bogalusa Heart Study has been supported by NIH grants R01HD069587, R01HL121230, R01AG016592, R01AG041200, P50HL015103, and R01HD032194. The

Coronary Artery Risk Development in Young Adults (CARDIA) Study was supported by contracts HHSN268201800003I, HHSN268201800004I, HHSN268201800005I, HHSN268201800006I, and HHSN268201800007I with the National Heart, Lung, and Blood Institute (NHLBI), NIH. The CARDIA pregnancy-derived variables project was supported by grants K01 DK059944, R01 DK090047, and R01 DK106201 (E.P.G., Principal Investigator) from the National Institute of Diabetes and Digestive and Kidney Diseases, NIH. The Effects of Aspirin in Gestation and Reproduction (EAGeR) Study was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), NIH (contracts HHSN267200603423, HHSN267200603424, and HHSN267200603426). E.P.S. was supported by the Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS), NIH (grant Z01ES103333). The Growing Up Today Study and Nurses' Health Study 3 were supported by grant U01HL145386 from the NHLBI. The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) was supported by NHLBI contracts with the University of North Carolina (Chapel Hill, North Carolina; contract N01-HC65233), the University of Miami (Miami, Florida; contract N01-HC65234), Albert Einstein College of Medicine (New York, New York; contract N01-HC65235), Northwestern University (Evanston, Illinois; contract N01-HC65236), and San Diego State University (San Diego, California; contract N01-HC65237). The following institutes/centers/offices contributed to the HCHS/SOL through a transfer of funds to the NHLBI: the National Center on Minority Health and Health Disparities, the National Institute on Deafness and Other Communication Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the Office of Dietary Supplements. The Home Observation of Periconceptional Exposures (HOPE) Study was supported by NIEHS grant 1R01ES020488-01. The Time to Pregnancy in Couples of Proven Fecundity Study was supported by grant 1K23 HD0147901-01A1 from the NICHD and by the Health Studies Fund, Department of Family and Preventive Medicine, University of Utah (Salt Lake City, Utah). The Pregnancy Study Online was supported by extramural grants from the NICHD and NIEHS (grants R01HD086742, R01ES028923, R01ES029951, R01HD105863, R21HD094322, and R21HD072326) and the National Science Foundation (grant NSF1914792).

The data underlying this article will be shared upon reasonable request to the corresponding author.

We thank Tanran Wang (Boston University), Dr. Hongyan Huang (Channing Division of Network Medicine, Brigham and Women's Hospital), and Jie Sun (Channing Division of Network Medicine, Brigham and Women's Hospital) for their help during data access and analysis. Some of the research on which this paper is based was conducted as part of the Australian Longitudinal Study on Women's Health by the University of Queensland (Brisbane, Queensland, Australia) and the University of Newcastle (Newcastle, New South Wales, Australia). We are grateful to the women who provided the Australian survey data.

This work was presented at the 34th Annual Meeting of the Society for Pediatric and Perinatal Epidemiologic Research (virtual), June 10, 2021.

Informed consent was obtained from participants in each study during recruitment. Data-sharing was conducted in accordance with the approvals and data-use agreements of the participating studies. Creation of the overall consortium was approved by the Institutional Review Board of Tulane University.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflict of interest: none declared.

REFERENCES

1. Cohen JM, Cesta CE, Kjerpeseth L, et al. A common data model for harmonization in the Nordic Pregnancy Drug Safety Studies (NorPreSS). *Norsk Epidemiologi.* 2021; 29(1-2):117–123.
2. Eunice Kennedy Shriver National Institute of Child Health and Human Development. *Promoting Data Harmonization to Accelerate COVID-19 Pregnancy Research.* Bethesda, MD: Eunice Kennedy Shriver National Institute of Child Health and Human Development; 2021. https://www.niehs.nih.gov/research/programs/disaster/database/promoting_data_harmonization_to_accelerate_covid19_pregnancy_research_508.pdf. Accessed April 10, 2023.
3. Harville EW, Mishra GD, Yeung E, et al. The Preconception Period Analysis of Risks and Exposures influencing health and Development (PrePARED) consortium. *Paediatr Perinat Epidemiol.* 2019;33(6):490–502.
4. Fortier I, Burton PR, Robson PJ, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *Int J Epidemiol.* 2010;39(5): 1383–1393.
5. Rolland B, Reid S, Stelling D, et al. Toward rigorous data harmonization in cancer epidemiology research: one approach. *Am J Epidemiol.* 2015;182(12):1033–1038.
6. Qi W, Ratanatharathorn A, Gevonden M, et al. Application of data pooling to longitudinal studies of early post-traumatic stress disorder (PTSD): the International Consortium to Predict PTSD (ICPP) project. *Eur J Psychotraumatol.* 2018; 9(1):1476442.
7. Malanchini M, Smith-Woolley E, Ayorech Z, et al. Aggressive behaviour in childhood and adolescence: the role of smoking during pregnancy, evidence from four twin cohorts in the EU-ACTION consortium. *Psychol Med.* 2019; 49(4):646–654.
8. Brinton LA, Cook MB, McCormack V, et al. Anthropometric and hormonal risk factors for male breast cancer: Male Breast Cancer Pooling Project results. *J Natl Cancer Inst.* 2014; 106(3):djt465.
9. Fortier I, Dragieva N, Saliba M, et al. Harmonization of the health and risk factor questionnaire data of the Canadian Partnership for Tomorrow Project: a descriptive analysis. *CMAJ Open.* 2019;7(2):E272–E282.

10. Eijkemans MJC, Leridon H, Keiding N, et al. A systematic comparison of designs to study human fecundity. *Epidemiology.* 2019;30(1):120–129.
11. Santelli J, Rochat R, Hatfield-Timajchy K, et al. The measurement and meaning of unintended pregnancy. *Perspect Sex Reprod Health.* 2003;35(2):94–101.
12. Rowland BB, Rocca CH, Ralph LJ. Certainty and intention in pregnancy decision-making: an exploratory study. *Contraception.* 2021;103(2):80–85.
13. Buck Louis GM, Hediger ML, Bell EM, et al. Methodology for establishing a population-based birth cohort focusing on couple fertility and children's development, the Upstate KIDS Study. *Paediatr Perinat Epidemiol.* 2014;28(3):191–202.
14. Stanford JB, Sanders JN, Simonsen SE, et al. Methods for a retrospective population-based and clinic-based subfertility cohort study: the Fertility Experiences Study. *Paediatr Perinat Epidemiol.* 2016;30(4):397–407.
15. Maconochie N, Doyle P, Prior S, et al. Risk factors for first trimester miscarriage—results from a UK-population-based case-control study. *BJOG.* 2007;114(2):170–186.
16. Yu P, Zhou L, Meng F, et al. Developmental trajectories of health-promoting behaviours among pregnant women: a longitudinal study. *J Clin Nurs.* 32(1-2):234–242.
17. Pino EC, Zuo Y, Maciel de Olivera C, et al. Cohort profile: the MULTI sTUdy diabetes rEsearch (MULTITUDE) consortium. *BMJ Open.* 2018;8(5):e020640.
18. Hamilton CM, Strader LC, Pratt JG, et al. The PhenX Toolkit: get the most from your measures. *Am J Epidemiol.* 2011;174(3):253–260.
19. National Institute on Alcohol Abuse and Alcoholism. Alcohol's effects on health. Drinking levels defined. https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking. Accessed April 28, 2020.
20. Tranquilli AL, Dekker G, Magee L, et al. The classification, diagnosis and management of the hypertensive disorders of pregnancy: a revised statement from the ISSHP [editorial]. *Pregnancy Hypertens.* 2014;4(2):97–104.
21. Gunderson EP, Chiang V, Lewis CE, et al. Long-term blood pressure changes measured from before to after pregnancy relative to nonparous women. *Obstet Gynecol.* 2008;112(6):1294–1302.
22. American College of Obstetricians and Gynecologists Committee on Practice Bulletins—Obstetrics. ACOG Practice Bulletin no. 190: gestational diabetes mellitus. *Obstet Gynecol.* 2018;131(2):e49–e64.
23. Ashton DM, Lawrence HC, Adams NL, et al. Surgeon General's Conference on the Prevention of Preterm Birth. *Obstet Gynecol.* 2009;113(4):925–930.
24. Barker M, Dombrowski SU, Colbourn T, et al. Intervention strategies to improve nutrition and health behaviours before conception. *Lancet.* 2018;391(10132):1853–1864.
25. Johnson K, Posner SF, Biermann J, et al. Recommendations to improve preconception health and health care—United States: a report of the CDC/ATSDR Preconception Care Work Group and the Select Panel on Preconception Care. *MMWR Recomm Rep*. 2006;55(RR-6):1–23.
26. Kroelinger CD, Okoroh EM, Boulet SL, et al. Making the case: the importance of using 10 key preconception indicators in understanding the health of women of reproductive age. *J Womens Health.* 2018;27(6):739–743.
27. Stephenson J, Heslehurst N, Hall J, et al. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. *Lancet.* 2018;391(10132):1830–1841.
28. Campaigning for preconception health [editorial]. *Lancet.* 2018;391(10132):1749.
29. Lu MC, Kotelchuck M, Hogan V, et al. Closing the Black-White gap in birth outcomes: a life-course approach. 2010;20(1 suppl 2):S2-62–S2-76.
30. Robbins C, Boulet SL, Morgan I, et al. Disparities in preconception health indicators—Behavioral Risk Factor Surveillance System, 2013–2015, and Pregnancy Risk Assessment Monitoring System, 2013–2014. *MMWR Surveill Summ.* 2018;67(1):1–16.
31. Berenson GS. Bogalusa Heart Study: a long-term community study of a rural biracial (black/white) population. *Am J Med Sci.* 2001;322(5):267–274.
32. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol.* 1988;41(11):1105–1116.
33. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8):629–641.
34. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8):642–649.
35. Allen JW. *Fertility and Pregnancy: An Epidemiologic Perspective*. New York, NY: Oxford University Press; 2010.
36. Kuohung W, Hornstein MD. Overview of infertility. (Topic 7396, version 43.0). https://www.uptodate.com/contents/overview-of-infertility. Literature review current through July 2023. Updated April 24, 2023. Accessed July 9, 2020.
37. Hellerstedt WL, Pirie PL, Lando HA, et al. Differences in preconceptional and prenatal behaviors in women with intended and unintended pregnancies. *Am J Public Health.* 1998;88(4):663–666.
38. Naimi TS, Lipscomb LE, Brewer RD, et al. Binge drinking in the preconception period and the risk of unintended pregnancy: implications for women and their children. *Pediatrics.* 2003;111(suppl 1):1136–1141.
39. Yu Y, Hardy I, Sun W, et al. Diet quality during preconception or pregnancy and gestational weight gain: protocol for a systematic review and meta-analysis. *BMJ Open.* 2020;10(2):e033130.
40. Herman DR, Taylor Baer M, Adams E, et al. Life course perspective: evidence for the role of nutrition. *Matern Child Health J.* 2014;18(2):450–461.
41. Olfert MD, Barr ML, Charlier CM, et al. Self-reported vs. measured height, weight, and BMI in young adults. *Int J Environ Res Public Health.* 2018;15(10):2216.
42. Hinkle SN, Rawal S, Zhu Y, et al. Validation of self-reported diagnosis of gestational diabetes at six weeks postpartum. *Epidemiology.* 2017;28(5):747–752.
43. Gunderson EP, Lewis CE, Tsai AL, et al. A 20-year prospective study of childbearing and incidence of diabetes in young women, controlling for glycemia before conception. *Diabetes.* 2007;56(12):2990–2996.
44. Gresham E, Forder P, Chojenta CL, et al. Agreement between self-reported perinatal outcomes and administrative data in New South Wales, Australia. *BMC Pregnancy Childbirth.* 2015;15(1):161.
45. Devlin HM, Desai J, Walaszek A. Reviewing performance of birth certificate and hospital discharge data to identify births complicated by maternal diabetes. *Matern Child Health J.* 2009;13(5):660–666.
46. Northam S, Knapp TR. The reliability and validity of birth certificates. *J Obstet Gynecol Neonatal Nurs.* 2006;35(1):3–12.
47. Houshmand A, Jensen DM, Mathiesen ER, et al. Evolution of diagnostic criteria for gestational diabetes mellitus. *Acta Obstet Gynecol Scand.* 2013;92(7):739–745.

48. Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. *Can J Diabetes.* 2018;42(suppl 1):A6–A16.
49. American College of Obstetricians and Gynecologists Task Force on Hypertension in Pregnancy. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstet Gynecol.* 2013;122(5):1122–1131.
50. Kaufman JS, Cooper RS. Commentary: considerations for use of racial/ethnic classification in etiologic research. *Am J Epidemiol.* 2001;154(4):291–298.
51. National Academies of Sciences, Engineering, and Medicine. *Nutrition During Pregnancy and Lactation: Exploring New Evidence. Proceedings of a Workshop*. Washington, DC: National Academies Press; 2020.
52. Gunderson EP, Jacobs DR Jr, Chiang V, et al. Childbearing is associated with higher incidence of the metabolic syndrome among women of reproductive age controlling for measurements before pregnancy: the CARDIA Study. *Am J Obstet Gynecol.* 2009;201(2):177.e1–177.e9.
53. Chavarro JE, Rich-Edwards JW, Gaskins AJ, et al. Contributions of the Nurses' Health Studies to reproductive health research. *Am J Public Health.* 2016;106(9):1669–1676.
54. Wise LA, Rothman KJ, Mikkelsen EM, et al. Design and conduct of an Internet-based preconception cohort study in North America: Pregnancy Study Online. *Paediatr Perinat Epidemiol.* 2015;29(4):360–371.
55. Stanford JB, Smith KR, Varner MW. Impact of instruction in the Creighton Model FertilityCare System on time to pregnancy in couples of proven fecundity: results of a randomised trial. *Paediatr Perinat Epidemiol.* 2014;28(5):391–399.
56. Kusneniwar GN, Whelan RM, Betha K, et al. Cohort profile: the Longitudinal Indian Family hEalth (LIFE) Pilot Study, Telangana State, India. *Int J Epidemiol.* 2016;46(3):788–789j.
57. Porucznik CA, Cox KJ, Schliep KC, et al. The Home Observation of Periconceptional Exposures (HOPE) study, a prospective cohort: aims, design, recruitment and compliance. *Environ Health.* 2016;15(1):67.
58. Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. *N Engl J Med.* 1988;319(4):189–194.
59. Schisterman EF, Silver RM, Perkins NJ, et al. A randomised trial to evaluate the Effects of low-dose Aspirin in Gestation and Reproduction (EAGeR): design and baseline characteristics. *Paediatr Perinat Epidemiol.* 2013;27(6):598–609.