



Exercise During the First Trimester of Pregnancy and the Risks of Abnormal Screening and Gestational Diabetes Mellitus

Diabetes Care 2021;44:425–432 | <https://doi.org/10.2337/dc20-1475>

Samantha F. Ehrlich,^{1,2} Assiamira Ferrara,¹
Monique M. Hedderson,¹ Juanran Feng,¹
and Romain Neugebauer¹

OBJECTIVE

To estimate the effects of exercise during the first trimester on the risks of abnormal screening and gestational diabetes mellitus (GDM).

RESEARCH DESIGN AND METHODS

Data come from PETALS, a prospectively followed pregnancy cohort ($n = 2,246$, 79% minorities) receiving care at Kaiser Permanente Northern California. A Pregnancy Physical Activity Questionnaire was used to assess exercise. Glucose testing results for screening and diagnostic tests were obtained from electronic health records. Inverse probability of treatment weighting and targeted maximum likelihood with data-adaptive estimation (machine learning) of propensity scores and outcome regressions were used to obtain causal risk differences adjusted for potential confounders, including prepregnancy BMI, exercise before pregnancy, and gestational weight gain. Exercise was dichotomized at 1) the cohort's 75th percentile for moderate- to vigorous-intensity exercise (≥ 13.2 MET-h per week or ≥ 264 min per week of moderate exercise), 2) current recommendations (≥ 7.5 MET-h per week or ≥ 150 min per week of moderate exercise), and 3) any vigorous exercise.

RESULTS

Overall, 24.3% and 6.5% had abnormal screening and GDM, respectively. Exercise meeting or exceeding the 75th percentile decreased the risks of abnormal screening and GDM by 4.8 (95% CI 1.1, 8.5) and 2.1 (0.2, 4.1) fewer cases per 100, respectively, in adjusted analyses.

CONCLUSIONS

Exercise reduces the risks of abnormal screening and GDM, but the amount needed to achieve these risk reductions is likely higher than current recommendations. Future interventions may consider promoting ≥ 38 min per day of moderate-intensity exercise to prevent GDM.

The Physical Activity (PA) Guidelines for Americans (1), the American College of Obstetricians and Gynecologists (ACOG) (2,3), and the American College of Sports Medicine (4) recommend that pregnant women participate in at least 150 min of moderate-intensity activity per week. The 2020 ACOG committee opinion on PA and exercise during pregnancy states that in the absence of obstetric or medical complications or contraindications, PA during pregnancy is safe and desirable, and pregnant women should be encouraged to continue or initiate safe PA (2). The

¹Division of Research, Kaiser Permanente Northern California, Oakland, CA

²Department of Public Health, University of Tennessee, Knoxville, Knoxville, TN

Corresponding author: Samantha F. Ehrlich, sehrlic1@utk.edu or sehrlich@utmck.edu

Received 16 June 2020 and accepted 18 November 2020

This article contains supplementary material online at <https://doi.org/10.2337/figshare.13262768>.

© 2020 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.

committee opinion highlights decreased risk of gestational diabetes mellitus (GDM) as a benefit of exercise during pregnancy.

A 2018 systematic review and meta-analysis found that exercise-only interventions reduced the odds of GDM by 38% compared with no intervention (5), although most of these exercise-only interventions included did not reduce the risk of GDM on their own (5). It is important to supplement data obtained from randomized controlled trials of exercise interventions conducted in motivated volunteers with the findings of population-based cohort studies assessing free-living PA because the later are the basis for PA recommendations. To further improve upon recommendations for exercise in pregnancy, there is the need to better define the minimum threshold for volume of exercise needed to improve health outcomes since volume incorporates the exercise metrics of intensity, duration, and frequency into a single value. Studies examining such thresholds are needed to inform the development of lifestyle interventions to prevent GDM.

This prospective cohort study examined the association of exercise during the first trimester of pregnancy with abnormal glucose values obtained from 50-g, 1-h screening for GDM and for GDM according to Carpenter and Coustan criteria. We additionally sought, a priori, to investigate potential differences by prepregnancy weight status (5–7) and to use contemporary causal inference methods to address this question (8–11).

RESEARCH DESIGN AND METHODS

The study setting was the Kaiser Permanente Northern California (KPNC) health care delivery system, which serves 3.6 million members. The KPNC population is representative of the geographic area served and is racially/ethnically and socioeconomically diverse (12). Data for the current study come from the Pregnancy Environment and Lifestyle Study (PETALS) birth cohort study. Beginning in 2013, PETALS recruited KPNC women initiating prenatal care at <11 gestational weeks; the participation rate was 75% (13).

At 10–13 gestational weeks, participants provided consent and had height and weight measured. The study survey was the Pregnancy Physical Activity

Questionnaire (PPAQ), which participants completed on their own. The PETALS PPAQ (14) was slightly modified from the original (15) and updated to reflect technological advances (e.g., watching TV or a video was updated to watching TV, a movie, or video clip). The PPAQ asked women to report time spent in 36 population-appropriate activities during the previous 2 months. Response options included ranges for the amount of time spent in each activity (e.g., none, <1/2 h per day, 1/2 to almost 1 h per day, 1 to almost 2 h per day, 2 to almost 3 h per day, or ≥ 3 h per day). To provide a conservative estimate, the minimum value of the range selected (duration and frequency of activity) were multiplied by the intensity, expressed in METs, to derive an estimate of the volume of PA (MET-hours per day). Compendium-based MET values were used for activities other than walking and household tasks, which came from field-based measurements of pregnant women (15).

The focus of the current study is the sports and exercise activity domain: PA that is intentional for health and wellness or to increase fitness and results in energy expenditure beyond the demands of everyday living. This domain includes 10 PPAQ activities of moderate intensity (range 3.2–6 METs for walking, swimming, etc.) plus 2 PPAQ activities of vigorous intensity (6.5 and 7 METs for walking quickly up hills and jogging, respectively) performed for fun or exercise (14). It includes an item for weight lifting, resistance exercise (i.e., of moderate intensity), in addition to items assessing for activities that combine resistance and cardiovascular exercises (e.g., swimming). The sum of the volume (MET-hours per day) of all activities in this domain provided an overall estimate of volume of moderate-to vigorous-intensity sports and exercise activity (referred to as exercise hereafter). Cronbach α was 0.75 for the PETALS PPAQ overall (36 items) and 0.70 for the sports and exercise domain (12 items).

To inform public health recommendations pertaining to PA during pregnancy, the exposure was defined by several thresholds. Since PA questionnaire data are self-report and subject to bias, they are most appropriate for ranking individuals with regard to volume of activity. Thus a high volume of exercise was first defined as meeting or exceeding the cohort-specific 75th percentile (i.e., volume of sports and exercise activity ≥ 13.2 MET-hours per

week) (16). Exercise was then defined as meeting or exceeding the lower bound of the current PA guidelines (1,2) recommending 150–300 min per week of moderate exercise, 75–150 min per week of vigorous exercise, or the equivalent volume of moderate/vigorous exercises combined (i.e., ≥ 7.5 MET-hours per week) (1,16,17). Finally, although vigorous exercise is believed to be safe for most healthy pregnant women, particularly those who participated in vigorous exercise prior to pregnancy, the 2020 ACOG committee opinion expressed the need for more data (2). As such, performing any vigorous exercise was examined.

The study survey also provided information on age, race/ethnicity, marital status, parity, education, and exercise during the year prior to pregnancy (6,7). It included a Block food frequency questionnaire (18,19) assessing dietary intake between conception and the study visit, specifically daily caloric intake in kcal. Data on gestational age were obtained from the electronic health record. Clinic-measured prepregnancy weight (i.e., within 1 year of the last menstrual period date) was available in the electronic health record for the majority of the cohort; otherwise, a clinic-measured weight before 10 weeks gestation from the electronic health record or self-reported prepregnancy weight from the study survey was used. BMI was calculated as weight in kilograms divided by height in meters squared and classified according to standard thresholds (20). Prepregnancy weight was subtracted from weight measured at the study visit (i.e., by PETALS research staff) to isolate gestational weight gain through the study visit (21) since weight gain occurring after the exercise exposure could lie on the causal pathway between exercise and the development of abnormal screening and GDM.

Results of all blood glucose testing were obtained from the KPNC Gestational Diabetes and Pregnancy Glucose Tolerance Registry (22); all plasma glucose measurements in this setting are performed using the hexokinase method at the KPNC regional laboratory. Women with recognized pre-GDM were identified in the KPNC Diabetes Registry (23) and ineligible to participate in PETALS. The KPNC Diabetes Registry identifies patients from four data sources: primary hospital discharge diagnoses of diabetes, two or more outpatient visit diagnoses of

diabetes, any prescription for a diabetes-related medication, or any record of an abnormal HbA_{1c} test (>50 mmol/mol [6.7%] [23]).

The outcomes—abnormal screening and GDM—were defined according to the screening and diagnosis protocol used in this clinical setting. At KPNC, the two-step approach is used to identify GDM: At 24–28 weeks gestation, women are screened with a 50-g, 1-h glucose challenge test (>95% of pregnancies screened), and those with glucose \geq 140 mg/dL (7.8 mmol/L) (referred to as abnormal screening hereafter) go on to a diagnostic 100-g, 3-h oral glucose tolerance test (OGTT). For the current study, only women with two or more glucose values on the OGTT meeting or exceeding thresholds proposed by Carpenter and Coustan (24) are considered to have GDM (fasting: 95 mg/dL [5.3 mmol/L]; 1 h: 180 mg/dL [10.1 mmol/L]; 2 h: 155 mg/dL [8.7 mmol/L]; 3 h: 140 mg/dL [7.8 mmol/L]).

PETALS participants with PPAQ data who delivered between October 2013 and October 2017 were eligible for the current study ($N = 2,501$). Those with a contraindication to PA during pregnancy (3) who were diagnosed before the study visit were excluded ($n = 6$) as were 72 women with implausible volumes of PA (25,26). Participants missing blood glucose data ($n = 177$) were then excluded (e.g., not screened or abnormal screening but no follow-up OGTT). The final analytic cohort consisted of 2,246 women.

Statistical Analyses

Causal inference methods were used to obtain estimates of the average difference in risk of an abnormal screening test result and GDM. Under the assumptions discussed below, all estimates can be interpreted as the difference in risk had all women exercised above one of the three activity levels considered minus the risk had none exercised above the same level during the first trimester. Adjusted risk differences were estimated by four strategies: inverse probability of treatment weighting (IPTW), then targeted maximum likelihood estimation (TMLE) (8–11) with user-defined parametric models, and finally, TMLE implemented with two sets of machine learning algorithms (i.e., the defaults and an expanded set) to potentially improve

upon the robustness and precision of the estimates.

For the IPTW analyses, the propensity for exercise was estimated by logistic regression and included the following baseline covariates (i.e., adjusted for): age (continuous), prepregnancy BMI (continuous), daily caloric intake (kcal) (continuous), marital status (married [reference], not married), race/ethnicity (Hispanic [largest group, reference], White, Asian/Pacific Islander, African American, and other), parity (0 [reference], 1, 2+), education (high school or less, some college [reference], college graduate, postgraduate), and exercise during the year prior to pregnancy (did not exercise, did exercise [reference]).

Unadjusted (i.e., unweighted) and adjusted IPTW (with stabilized, untruncated weights) estimates were obtained using Proc Genmod with an independence structure for the variance-covariance matrix (SAS 9.4 software). Stabilized IPTW weights were not truncated because all maximums were <13.5 , except for one subgroup analysis with a maximum IPTW weight of 36.5 (27) (Supplementary Table 1).

R 4.0.2 software was used to implement TMLE, a doubly robust, locally efficient estimator (8–10). Valid inference from TMLE depends on correctly estimating the model for the propensity score or the outcome regression. To avoid incorrect TMLE inference from misspecified parametric models (e.g., logistic) for propensity score and outcome regression, data-adaptive machine learning methods, such as Super Learning (28), have been proposed to estimate nuisance parameters (10,29–32). Three approaches were used to estimate the propensity score and the outcome regression portions of the likelihood: 1) the same logistic regression modeling approach previously described, 2) adaptively with Super Learning (28) using the default “learners” (8,9) (i.e., prediction algorithms) only and the adjustment variables previously described, and 3) adaptively with Super Learning, with the default learners plus six extra learners and the adjustment variables previously described. The default learners (8,9) included Wrapper for Glm, Choose a Model by AIC in a Stepwise Algorithm, and Wrapper Function for SuperLearner Prediction Algorithm, and the extra learners included SL Wrapper for Biglasso, Elastic

Net Regression Including Lasso and Ridge, Wrapper for Kernlab’s SVM Algorithm, Wrapper for Lm, SL Wrapper for Ranger, and Wrapper for Speedglm.

For both adjusted IPTW and TMLE, missing covariate data were addressed using the missingness indicator method (33,34) (i.e., an indicator of missing values was included in the adjustment set, and missing values were replaced with the median of observed values). Baseline categorical covariates with missing values included marital status ($n = 4$), parity ($n = 4$), and education ($n = 2$). Those missing caloric intake ($n = 52$) or with implausible estimates of daily caloric intake (i.e., <400 cal or $>6,000$ cal, $n = 19$) were assigned the cohort median value of 1,448.43 cal; women missing gestational weight gained through study visit ($n = 4$) were assigned the cohort median 1.6 lb.

Analyses stratified by prepregnancy weight status (underweight/healthy weight vs. overweight/obese) estimated subgroup-specific associations (6). Meeting or exceeding the time point–specific thresholds of Carpenter and Coustan on the 100-g, 3-h OGTT (24) were also examined as outcomes. The threshold for statistical significance was 0.05 for all analyses. The study was approved by the KPNC institutional review board.

RESULTS

Characteristics of the analytic cohort ($n = 2,246$) are presented in Table 1. The mean age was 30.2 (SD 5.3) years, and 56% ($n = 1,272$) of the women were overweight or obese as determined by a KPNC clinic-measured prepregnancy weight (i.e., within 1 year of the last menstrual period) for 78% ($n = 1,753$) of the cohort, a KPNC clinic-measured weight from early pregnancy (i.e., before 10 weeks gestation) for 19.4% ($n = 436$), and self-reported prepregnancy weight on the study survey for the remaining 2.5% ($n = 57$). The Pearson correlation coefficient was 0.99 for clinic-measured prepregnancy weight and clinic-measured early pregnancy weight ($n = 1,687$; $P < 0.001$), 0.98 for clinic-measured early pregnancy weight and self-reported prepregnancy weight ($n = 2,122$; $P < 0.001$), and 0.98 for self-reported prepregnancy weight and clinic-measured prepregnancy weight ($n = 1,752$; $P < 0.001$). Women excluded from the current study were less likely to be married, more often

Table 1—Characteristics of the analytic cohort overall and stratified by meeting the cohort-specific 75th percentile for moderate- to vigorous-intensity exercise: PETALS, KPNC, 2013–2017

	Overall (N = 2,246)	Cohort's 75th percentile for moderate- to vigorous-intensity exercise	
		Meeting (n = 561) ^a	Not meeting (n = 1,685) ^b
Volume of moderate- to vigorous-intensity exercise (MET-hours per week), median (IQR)	4.8 (11.6)	21.7 (13.0)	2.7 (5.6)
Age (years)			
≤25	465 (20.7)	103 (18.4)	362 (21.5)
25–35	1,310 (58.3)	328 (58.5)	982 (58.3)
≥35	471 (21.0)	130 (23.2)	341 (20.2)
Prepregnancy BMI			
Underweight	60 (2.7)	16 (2.9)	44 (2.6)
Normal weight	914 (40.7)	218 (38.9)	696 (41.3)
Overweight	646 (28.8)	169 (30.1)	477 (28.3)
Obese	626 (27.9)	158 (28.2)	468 (27.8)
Exercised prior to pregnancy ^c	1,980 (88.2)	547 (97.5)	1,433 (85.0)
Diet (kcal) (n = 2,175)			
<1,000	397 (18.3)	95 (17.2)	302 (18.6)
1,000–2,000	1,322 (60.8)	325 (59.0)	997 (61.4)
>2,000	456 (21.0)	131 (23.8)	325 (20.0)
Married (n = 2,242)	1,521 (67.8)	393 (70.3)	1,128 (67.0)
Race/ethnicity ^c			
Hispanic	937 (41.7)	223 (39.8)	714 (42.4)
White	481 (21.4)	170 (30.3)	311 (18.5)
Asian American/Pacific Islander	530 (23.6)	98 (17.5)	432 (25.6)
African American	219 (9.8)	46 (8.2)	173 (10.3)
Other	79 (3.5)	24 (4.3)	55 (3.3)
Parity (n = 2,242)			
0	986 (44.0)	257 (45.9)	729 (43.3)
1	828 (36.9)	200 (35.7)	628 (37.3)
2+	428 (19.1)	103 (18.4)	325 (19.3)
Education (n = 2,244) ^c			
High school or less	304 (13.6)	57 (10.2)	247 (14.7)
Some college	872 (38.9)	190 (33.9)	682 (40.5)
College graduate	638 (28.4)	190 (33.9)	448 (26.6)
Postgraduate	430 (19.2)	123 (22.0)	307 (18.2)
Abnormal screen ^c	545 (24.3)	115 (20.5)	430 (25.5)
GDM ^c	147 (6.5)	26 (4.6)	121 (7.2)

Data are n (%) unless otherwise indicated. IQR, interquartile range. ^aMeeting: ≥13.2 MET-hours per week. ^bNot meeting: <13.2 MET-hours per week. ^c χ^2 P < 0.0001.

multiparous, and attained lower levels of education compared with those included in the analytic cohort (Supplementary Table 2).

Of those included in the current study, 913 (40.7%) women met or exceeded the lower bound of the PA guidelines, and 829 (36.9%) reported any amount of vigorous-intensity exercise; 18.3% reported weight lifting or resistance exercises. Compared with women who were below the cohort-specific 75th percentile for exercise, those who met or exceeded this threshold were more likely to have exercised prior to pregnancy, to be White, and to have attained a college or postgraduate education. Overall, mean gestational age at the PPAQ was 12.8 (SD 2.5) weeks, and mean gestational age at the 50-g, 1-h glucose challenge test was 24.4

(SD 5.7) weeks. The frequency of abnormal screening and GDM were 24.3% and 6.5%, respectively (Table 1).

Table 2 presents the exposures, outcomes, and key confounders (e.g., prepregnancy exercise, gestational weight gain through the study visit) stratified by BMI category. The prevalence of abnormal screening and GDM increased with increasing BMI category, and women with overweight and obesity had lower gestational weight gain than underweight and healthy weight women.

Table 3 presents estimates of the causal risk differences for an abnormal screening test and GDM if all women had exercised at or above the cohort-specific 75th percentile for moderate- to vigorous-intensity exercise, met the PA guidelines,

and participated in any vigorous-intensity exercise versus not. All estimates for exercise at or above the cohort-specific 75th percentile indicated a reduction in the risk of abnormal screening. The estimate from data-adaptive TMLE with extra learners revealed 4.8 (95% CI 1.1, 8.5) fewer abnormal screening tests per 100 women had all women exercised at or above the cohort-specific 75th percentile versus not. Exercise at this level was also associated with a reduced risk of GDM in the unadjusted IPTW, data-adaptive TMLE with defaults, and data-adaptive TMLE with extra learners (2.6 [0.2, 4.9], 2.0 [0.2, 4.2], and 2.1 [0.2, 4.1] fewer cases of GDM per 100 women, respectively). There was the suggestion of a reduced risk of GDM for exercise meeting or

Table 2—Exercise, gestational weight gain, abnormal screening, and GDM by prepregnancy BMI category: PETALS, KPNC, 2013–2017

	Underweight (<i>n</i> = 59)	Healthy weight (<i>n</i> = 914)	Overweight (<i>n</i> = 647)	Obese (<i>n</i> = 626)
Exercised prior to pregnancy	50 (84.8)	798 (87.3)	573 (88.6)	559 (89.3)
Exercise during the first trimester				
Cohort 75th percentile (≥ 13.2 MET-hours per week)	16 (27.1)	218 (23.9)	169 (26.1)	158 (25.2)
PA guidelines (≥ 7.5 MET-hours per week)	20 (33.9)	372 (40.7)	269 (41.6)	252 (40.3)
Any vigorous intensity	17 (28.8)	336 (36.8)	262 (40.5)	214 (34.2)
Abnormal screen ^a	11 (18.6)	186 (20.4)	166 (25.7)	182 (29.1)
GDM ^a	1 (1.7)	41 (4.5)	46 (7.1)	59 (9.4)
Volume of moderate- to vigorous-intensity exercise during the first trimester (MET-hours per week), median (IQR)	4.0 (12.7)	4.8 (11.3)	5.2 (11.9)	4.6 (11.8)
Gestational age at study visit (weeks), median (IQR)	13.0 (4.0)	13.0 (3.0)	13.0 (3.0)	13.0 (4.0)
Gestational weight gained through study visit (lb) (<i>n</i> = 2,242), median (IQR) ^b	2.3 (6.3)	2.4 (7.2)	0.7 (8.9)	1.0 (10.0)

Data are *n* (%) unless otherwise indicated. IQR, interquartile range. ^a $\chi^2 P < 0.001$. ^bKruskal-Wallis $\chi^2 P < 0.0001$.

exceeding the lower bound of the PA guidelines in the unadjusted IPTW analysis (2.0 [−0.1, 4.1] fewer cases of GDM per 100 women; $P = 0.06$). Any vigorous-intensity exercise reduced the risk of GDM by 2.5 (0.4, 4.7) per 100 women with unadjusted IPTW.

Table 4 displays estimates of the causal risk differences for abnormal screening and GDM stratified by prepregnancy weight status (i.e., underweight or healthy weight vs. overweight or obesity). Among underweight or healthy weight women ($n = 973$), 197 (20.3%) had abnormal screening and 42 (4.3%) had GDM, and all risk difference estimates for abnormal screening attained statistical significance in this subgroup. Data-adaptive TMLE with extra learners indicated that 7.1 (95% CI 1.5, 12.7) fewer underweight or healthy weight women per 100 would have an abnormal screening test had all exercised at or above the 75th percentile versus not. In women with overweight or obesity ($n = 1,272$), 348 (27.3%) had abnormal screening and 105 (8.3%) had GDM, and all adjusted estimates for GDM attained statistical significance. If all the women with overweight or obesity had exercised at or above the 75th percentile, there would be 2.9 (0.3, 5.4) fewer cases of GDM according to data-adaptive TMLE with extra learners.

Estimates of the causal risk differences for meeting or exceeding the time point-specific diagnostic thresholds of Carpenter and Coustan for the 100-g, 3-h OGTT if all women had exercised at or above the 75th percentile (13.2 MET-hours per week), versus not, are presented in Supplementary Table 3. Estimates for the 2-h threshold

attained significance in several adjusted analyses; data-adaptive TMLE with extra learners indicated that 2.5 (95% CI 0.3, 4.7) fewer women per 100 would meet the 2-h threshold had all exercised at or above the 75th percentile versus not. There was a decreased risk of meeting the fasting threshold for exercise at or above the 75th percentile using adjusted IPTW (1.6 [−0.3, −3.0] fewer per 100 women). No associations were observed for the 1-h and 3-h thresholds.

CONCLUSIONS

The results of this study suggest that meeting current recommendations for exercise during the first trimester of pregnancy does not confer reductions in the risks of abnormal screening and GDM. However, exercise at or above the cohort-specific 75th percentile, a higher minimum threshold, was found to reduce the risks of abnormal screening and GDM by 4.8 and 2.1 per 100 women, respectively. Although a meta-analysis of the results of randomized controlled trials of exercise-only interventions demonstrated a reduced odds of GDM, most of the individual studies of exercise-only interventions did not reduce GDM when evaluated on their own (5). Indeed, randomized controlled trials of pregnancy lifestyle interventions of diet and exercise have largely been unsuccessful in preventing GDM (35). The results of this study suggest that future lifestyle interventions among pregnant women who are free of complications and contraindications to PA (3) and monitored by an obstetric care provider (2) should prescribe a volume of exercise

that is higher than the currently recommended minimum in order to reduce GDM risk.

ACOG's (2,3) recommendation for at least 150 min per week of moderate-intensity PA (i.e., ≥ 7.5 MET-hours per week [1,16,17]) is based on the PA guideline recommendations for the general American population (1), which prescribe either moderate-intensity exercise, vigorous-intensity exercise, or a comparable combination of moderate- and vigorous-intensity exercises. Moderate-intensity exercise encompasses the 3–6-MET range, and the lowest value (i.e., 3 METs) is typically used to construct the recommendation; moderate-intensity exercise is also recommended for pregnant women (2). The cohort-specific 75th percentile examined in the current study was 13.2 MET-hours per week and would be achieved through ≥ 264 min per week of moderate-intensity exercise (i.e., at 3 METs) or alternatively stated, ≥ 38 min per day of moderate-intensity exercise. Since walking is the mostly commonly reported exercise among pregnant women in the U.S. (36,37), future lifestyle interventions should consider prescribing this amount of moderate-intensity walking to reduce the risk of GDM.

Davenport et al. (5) recommended at least 10 MET-hours per week of moderate-intensity exercise to achieve at least a 25% reduction in the odds of GDM, gestational hypertension, and preeclampsia. Their meta-analysis included 6,934 women from 26 randomized controlled trials of exercise-only interventions and found that among motivated trial volunteers, exercise-only interventions reduced the

Table 3—Estimates of the causal risk differences for an abnormal screening test and GDM if all had exercised at or above the cohort-specific 75th percentile for moderate- to vigorous-intensity exercise, met the PA guidelines, and participated in any vigorous-intensity exercise vs. not: PETALS, KPNC, 2013–2017

	Abnormal screen		GDM by Carpenter and Coustan	
	Risk difference (95% CI)	P value	Risk difference (95% CI)	P value
Cases, n (%)	545 (24.3)		147 (6.5)	
Cohort's 75th percentile (≥ 13.2 MET-hours per week)				
Exposed cases, n (%)	115 (5.1)		26 (1.2)	
Crude				
Unadjusted IPTW	-0.0502 (-0.0911, -0.00926)	0.02	-0.0255 (-0.0491, -0.00184)	0.03
Adjusted				
Stabilized IPTW ^a	-0.0564 (-0.101, -0.0114)	0.01	-0.0177 (-0.0453, 0.00990)	0.21
TMLE				
User specified ^b	-0.0562 (-0.102, -0.0105)	0.02	-0.0161 (-0.0447, 0.0126)	0.27
Data adaptive with defaults ^c	-0.0495 (-0.0889, -0.0101)	0.01	-0.0202 (-0.0420, 0.00165)	0.07
Data adaptive with defaults and extra learners ^d	-0.0476 (-0.0845, -0.0108)	0.01	-0.0212 (-0.0408, -0.00172)	0.03
PA guidelines (≥ 7.5 MET-hours per week)				
Exposed cases, n (%)	209 (9.3)		49 (2.2)	
Crude				
Unadjusted IPTW	-0.0232 (-0.0593, 0.0130)	0.21	-0.0199 (-0.0407, 0.000979)	0.06
Adjusted				
Stabilized IPTW ^a	-0.0261 (-0.0640, 0.0118)	0.18	-0.0180 (-0.0396, 0.00360)	0.10
TMLE				
User specified ^b	-0.0249 (-0.0622, 0.0124)	0.19	-0.0159 (-0.0374, 0.00562)	0.15
Data adaptive with defaults ^c	-0.0175 (-0.0530, 0.0180)	0.33	-0.0158 (-0.0359, 0.00435)	0.12
Data adaptive with defaults and extra learners ^d	-0.0146 (-0.0486, 0.0194)	0.40	-0.0165 (-0.0353, 0.00225)	0.08
Any vigorous intensity				
Exposed cases, n (%)	196 (8.7)		41 (1.8)	
Crude				
Unadjusted IPTW	-0.00987 (-0.0466, 0.0269)	0.60	-0.0254 (-0.0465, -0.00416)	0.02
Adjusted				
Stabilized IPTW ^a	-0.00270 (-0.0412, 0.0357)	0.89	-0.0196 (-0.0402, 0.00110)	0.06
TMLE				
User specified ^b	-0.000826 (-0.0382, 0.0366)	0.97	-0.0170 (-0.0373, 0.00332)	0.10
Data adaptive with defaults ^c	0.00633 (-0.0300, 0.0426)	0.73	-0.0147 (-0.0352, 0.00591)	0.16
Data adaptive with defaults and extra learners ^d	0.00824 (-0.0255, 0.0420)	0.63	-0.0147 (-0.0341, 0.00465)	0.14

^aLogistic regression for the propensity score model included the baseline covariates maternal age, prepregnancy BMI, marital status, race/ethnicity, educational attainment, parity, daily caloric intake, exercise prior to pregnancy, gestational weight gained through study visit, and gestational age at study visit. *P* values and 95% CIs are from Proc Genmod, with independence as the identity matrix for robust/sandwich estimator of the variance to provide conservative inference. ^bLogistic regression with the same baseline covariates to specify the propensity score and outcome models.

^cSuperLearner with defaults learners only (i.e., Wrapper for Glm, Choose a Model by AIC in a Stepwise Algorithm, and Wrapper Function for SuperLearner prediction algorithm) and the same baseline covariates to specify the propensity score and outcome models. ^dSuperLearner with defaults plus additional learners (i.e., SL Wrapper for Biglasso, Elastic Net Regression Including Lasso and Ridge, Wrapper for Kernlab's SVM Algorithm, Wrapper for Lm, SL Wrapper for Ranger, and Wrapper for Speedglm) and the same baseline covariates to specify the propensity score and outcome models.

odds of developing GDM by 38% compared with no intervention (5). The current study found that 2.1 cases of GDM per 100 women would be prevented with exercise compared with no exercise. The prevalence of GDM was 6.5 per 100 women; thus, exercise at or above 13.2 MET-hours per week reduced the risk of GDM by 32%, similar to that observed in the meta-analysis. It is worth noting that our study examined free-living or natural exercise, a related but distinct exposure from the exercise interventions examined in randomized controlled trials, which often include elements that providers cannot universally recommend and that are not affordable for all women (e.g., the use of exercise equipment such

as a treadmill or stationary bicycle, personalized training sessions).

Valid causal inference with the methods used by the current study requires several assumptions. The first assumption is the positivity or experimental treatment assignment assumption, which posits that all manifestations of the exposure must be possible conditional on the baseline covariates. The few women with contraindications to exercise during early pregnancy were thus excluded from the current study ($n = 6$), and for women without contraindications, it is reasonable to assume that any level of exercise would be possible. A second, untestable assumption is that of no unmeasured confounding. The current study had high-

quality data on several key confounding factors, but the strong possibility for residual confounding remains, as is the case for any observational study of a health behavior. Valid IPTW inference assumes consistent estimation of the propensity score (i.e., the logistic model used in this study being correctly specified). Valid TMLE inference relies on consistent estimation of either the propensity score or the outcome regression; the current study used machine learning (i.e., SuperLearner [31]) in the TMLE analyses to avoid violating this assumption.

Strengths of the current study include the prospective design, the size and racial/ethnic diversity of the cohort, and the availability of measured prepregnancy

Table 4—Estimates of the causal risk differences for an abnormal screening test and GDM if all had exercised at or above the cohort-specific 75th percentile for moderate- to vigorous-intensity exercise vs. not, stratified by prepregnancy weight status: PETALS, KPNC, 2013–2017

Cohort's 75th percentile (≥ 13.2 MET-hours per week)	Abnormal screen		GDM by Carpenter and Coustan	
	Risk difference (95% CI)	<i>P</i> value	Risk difference (95% CI)	<i>P</i> value
Underweight and healthy weight (<i>n</i> = 973)				
Cases, <i>n</i> (%)	197 (20.3)		42 (4.3)	
Crude				
Unadjusted IPTW	−0.0753 (−0.134, −0.0163)	0.01	−0.0175 (−0.0474, 0.0125)	0.25
Adjusted				
Stabilized IPTW ^a	−0.0992 (−0.160, −0.0386)	0.001	−0.0114 (−0.0451, 0.0224)	0.51
TMLE				
User specified ^b	−0.0833 (−0.140, −0.0263)	0.004	−0.00830 (−0.0455, 0.0289)	0.66
Data adaptive with defaults ^c	−0.0710 (−0.127, −0.0150)	0.01	−0.0105 (−0.0415, 0.0205)	0.51
Data adaptive with defaults and extra learners ^d	−0.0714 (−0.127, −0.0154)	0.01	−0.0103 (−0.0401, 0.0196)	0.50
Overweight or obese (<i>n</i> = 1,272)				
Cases, <i>n</i> (%)	348 (27.4)		105 (8.3)	
Crude				
Unadjusted IPTW	−0.0345 (−0.0906, 0.0216)	0.23	−0.0328 (−0.0674, 0.00180)	0.06
Adjusted				
Stabilized IPTW ^a	−0.0391 (−0.103, 0.0251)	0.23	−0.0360 (−0.0670, −0.00490)	0.02
TMLE				
User specified ^b	−0.0466 (−0.115, 0.0217)	0.18	−0.0384 (−0.0712, −0.00564)	0.02
Data adaptive with defaults ^c	−0.0252 (−0.0808, 0.0304)	0.38	−0.0331 (−0.0623, −0.00397)	0.03
Data adaptive with defaults and extra learners ^d	−0.0266 (−0.0753, 0.0220)	0.28	−0.0285 (−0.0542, −0.00285)	0.03

^aLogistic regression for the propensity score model included the baseline covariates maternal age, prepregnancy BMI, marital status, race/ethnicity, educational attainment, parity, daily caloric intake, exercise prior to pregnancy, gestational weight gained through study visit, and gestational age at study visit. *P* values and 95% CIs are from Proc Genmod, with independence as the identity matrix for robust/sandwich estimator of the variance to provide conservative inference. ^bLogistic regression with the same baseline covariates to specify the propensity score and outcome models. ^cSuperLearner with defaults learners only (i.e., Wrapper for Glm, Choose a Model by AIC in a Stepwise Algorithm, and Wrapper Function for SuperLearner Prediction Algorithm) and the same baseline covariates to specify the propensity score and outcome models. ^dSuperLearner with defaults plus additional learners (i.e., SL Wrapper for Biglasso, Elastic Net Regression Including Lasso and Ridge, Wrapper for Kernlab's SVM Algorithm, Wrapper for Lm, SL Wrapper for Ranger, and Wrapper for Speedglm) and the same baseline covariates to specify the propensity score and outcome models.

weight for most women and objective glucoses measurements to define the outcomes. There are several limitations worth noting. Women with implausible exposure data or missing outcome data were excluded from analyses, which could introduce bias. Since excluded women were more often unmarried, multiparous, and attained lower levels of education, results may not be generalizable to these subgroups. Fortunately, in the current study, only 10% of eligible women were excluded. PETALS also had a 75% participation rate overall (13), suggesting that these data are representative of the underlying source population. The PA exposure data were self-reported and subject to bias. As such, the minimum value for the range of duration and frequency of activity reported was used to mitigate the impact of overreporting and to provide a conservative estimate of the volume of PA. The current study used abnormal screening defined as ≥ 140 mg/dL and GDM by the Carpenter and

Coustan criteria only; thus, findings may not be generalizable to practices that use different screening thresholds or testing procedures and diagnostic criteria for GDM, including the International Association of Diabetes and Pregnancy Study Groups criteria (38).

With the proliferation of commercial PA monitors in recent years, there is great potential for future research on the association of objectively measured PA during pregnancy with abnormal screening and GDM. However, device-based measures of PA cannot distinguish between movement that is intentional for health and wellness and that which is occupational in nature, and positive health effects may be limited to the former. It also remains to be determined which specific objective metric of PA (i.e., steps vs. minutes of moderate- to vigorous-intensity PA) should be used to formulate device-based recommendations for PA in the future. Although current recommendations use minutes of moderate-

to vigorous-intensity aerobic activity (on the basis of self-report), device-based measures of steps are more accurately estimated than device-based minutes of moderate- to vigorous-intensity PA, and there is less interdevice variability in the estimation of steps (39,40). However, recommendations that are based on minutes of PA do not require the use of a PA monitoring device. Future studies, both randomized controlled trials and observational cohort studies of pregnant women, should include both device-based measures and validated questionnaires.

In conclusion, the results suggest that exercise during the first trimester of pregnancy is an effective tool for reducing the risks of abnormal screening and GDM. However, more exercise than is currently recommended may be needed to achieve reductions in the risks of these outcomes. Pregnancy lifestyle interventions should encourage, among women who are free of complications and contraindications to PA and monitored by

an obstetric care provider, at least 38 min per day of moderate-intensity exercise to prevent GDM.

Funding. The study was funded by National Institute of Environmental Health Sciences grant R01-ES-019196 to A.F. and supported by National Institutes of Health (NIH) Environmental Influences on Child Health Outcomes (ECHO) Program contract award UG3-OD-023289. S.F.E. was supported by National Institute of Diabetes and Digestive and Kidney Diseases grant K01-DK-105106. A.F. received support from National Institute of Diabetes and Digestive and Kidney Diseases grant P30-DK-092924.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. S.F.E., A.F., and M.M.H. conceived of the idea and designed the study. S.F.E. and R.N. planned the analyses. S.F.E. and J.F. performed the analyses. R.N. verified the analytic methods. S.F.E. drafted the manuscript. A.F., M.M.H., J.F., and R.N. provided critical feedback and helped to shape the final manuscript. S.F.E. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the 79th Scientific Sessions of the American Diabetes Association, San Francisco, CA, 7–11 June 2019.

References

- Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for Americans. *JAMA* 2018;320:2020–2028
- The American College of Obstetricians and Gynecologists. Physical activity and exercise during pregnancy and the postpartum period: ACOG committee opinion, number 804. *Obstet Gynecol* 2020;135:e178–e188
- The American College of Obstetricians and Gynecologists. ACOG committee opinion No. 650: physical activity and exercise during pregnancy and the postpartum period. *Obstet Gynecol* 2015;126:e135–e142
- American College of Sports Medicine (ACSM). *ACSM's Guidelines for Exercise Testing and Prescription*. 10th ed. Philadelphia, PA, Wolters Kluwer, 2018
- Davenport MH, Ruchat S-M, Poitras VJ, et al. Prenatal exercise for the prevention of gestational diabetes mellitus and hypertensive disorders of pregnancy: a systematic review and meta-analysis. *Br J Sports Med* 2018;52:1367–1375
- Oken E, Ning Y, Rifas-Shiman SL, Radesky JS, Rich-Edwards JW, Gillman MW. Associations of physical activity and inactivity before and during pregnancy with glucose tolerance. *Obstet Gynecol* 2006;108:1200–1207
- Badon SE, Wartko PD, Qiu C, Sorensen TK, Williams MA, Enquobahrie DA. Leisure time physical activity and gestational diabetes mellitus in the Omega study. *Med Sci Sports Exerc* 2016;48:1044–1052
- Gruber S, van der Laan M, Kennedy C. Targeted Maximum Likelihood Estimation. The Regents of the University of California. Accessed 15 October 2020. Available from <https://CRAN.R-project.org/package=tmlle>
- Gruber S, van der Laan M. *tmlle: an R package for targeted maximum likelihood estimation*. *J Stat Softw* 2012;51:1–35
- van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*, Springer Series in Statistics. New York, NY, Springer Science+Business Media, 2011
- Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol* 2017;185:65–73
- Gordon N, Lin T. The Kaiser Permanente Northern California adult member health survey. *Perm J* 2016;20:15–225
- Zhu Y, Hedderson MM, Feng J, Mevi AA, Ferrara A. The Pregnancy Environment and Lifestyle Study (PETALS): a population-based longitudinal multi-racial birth cohort. *BMC Pregnancy Childbirth* 2017;17:122
- Ehrlich SF, Sternfeld B, Krefman AE, et al. Erratum to: moderate and vigorous intensity exercise during pregnancy and gestational weight gain in women with gestational diabetes. *Matern Child Health J* 2016;20:1989–1992
- Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. *Med Sci Sports Exerc* 2004;36:1750–1760
- Harrod CS, Chasan-Taber L, Reynolds RM, et al. Physical activity in pregnancy and neonatal body composition: the Healthy Start study. *Obstet Gynecol* 2014;124:257–264
- Shiroma EJ, Sesso HD, Moorthy MV, Buring JE, Lee IM. Do moderate-intensity and vigorous-intensity physical activities reduce mortality rates to the same extent? *J Am Heart Assoc* 2014;3:e000802
- Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol* 1986;124:453–469
- Harley K, Eskenazi B, Block G. The association of time in the US and diet during pregnancy in low-income women of Mexican descent. *Paediatr Perinat Epidemiol* 2005;19:125–134
- World Health Organization. *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation*. Geneva, Switzerland, World Health Organization, 2000 (WHO technical report series 894)
- Hedderson MM, Gunderson EP, Ferrara A. Gestational weight gain and risk of gestational diabetes mellitus. *Obstet Gynecol* 2010;115:597–604
- Ferrara A, Kahn HS, Quesenberry CP, Riley C, Hedderson MM. An increase in the incidence of gestational diabetes mellitus: Northern California, 1991–2000. *Obstet Gynecol* 2004;103:526–533
- Selby JV, Ray GT, Zhang D, Colby CJ. Excess costs of medical care for patients with diabetes in a managed care population. *Diabetes Care* 1997;20:1396–1402
- Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. *Am J Obstet Gynecol* 1982;144:768–773
- Ehrlich SF, Sternfeld B, Krefman AE, et al. Moderate and vigorous intensity exercise during pregnancy and gestational weight gain in women with gestational diabetes. *Matern Child Health J* 2016;20:1247–1257
- Ehrlich SF, Hedderson MM, Brown SD, et al. Moderate intensity sports and exercise is associated with glycaemic control in women with gestational diabetes. *Diabetes Metab* 2017;43:416–423
- Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;168:656–664
- van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2008;6:25
- Neugebauer R, Schmittdiel JA, van der Laan MJ. A case study of the impact of data-adaptive versus model-based estimation of the propensity scores on causal inferences from three inverse probability weighting estimators. *Int J Biostat* 2016;12:131–155
- Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Stat Med* 2010;29:337–346
- Neugebauer R, Fireman B, Roy JA, Raebel MA, Nichols GA, O'Connor PJ. Super learning to hedge against incorrect inference from arbitrary parametric assumptions in marginal structural modeling. *J Clin Epidemiol* 2013;66(Suppl.):S99–S109
- van der Laan MJ. Targeted estimation of nuisance parameters to obtain valid statistical inference. *Int J Biostat* 2014;10:29–57
- Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010;25:1–21
- Blake HA, Leyrat C, Mansfield KE, Tomlinson LA, Carpenter J, Williamson EJ. Estimating treatment effects with partially observed covariates using outcome regression with missing indicators. *Biom J* 2020;62:428–443
- Mudd LM, Evenson KR. Review of impacts of physical activity on maternal metabolic health during pregnancy. *Curr Diab Rep* 2015;15:572
- Evenson KR, Wen F. National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999–2006. *Prev Med* 2010;50:123–128
- Connolly CP, Conger SA, Montoyo AHK, et al. Walking for health during pregnancy: a literature review and considerations for future research. *J Sport Health Sci* 2019;8:401–411
- Metzger BE, Gabbe SG, Persson B, et al.; International Association of Diabetes and Pregnancy Study Groups Consensus Panel. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* 2010;33:676–682
- Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act* 2015;12:42
- Reid RER, Insogna JA, Carver TE, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *J Sci Med Sport* 2017;20:578–582