



A randomized pragmatic clinical trial of gestational diabetes screening (ScreenR2GDM): Study design, baseline characteristics, and protocol adherence



Kathryn L. Pedula^{a,*}, Teresa A. Hillier^a, Keith K. Ogasawara^b, Kimberly K. Vesco^a, Suzanne Lubarsky^c, Caryn E.S. Oshiro^d, Jan VanMarter^a

^a Center for Health Research, Kaiser Permanente Northwest, 3800 N Interstate Ave, Portland, OR 97227, USA

^b Department of Obstetrics & Gynecology, Hawaii Permanente Medical Group, Kaiser Permanente, 3288 Moanalua Road, Honolulu, HI 96819, USA

^c Department of Perinatology, Northwest Permanente, Kaiser Permanente, 10180 SE Sunnyside Rd, Clackamas, OR 97015, USA

^d Center for Health Research, Kaiser Permanente Hawaii, 501 Alakawa St, Suite 201, Honolulu, HI 96817, USA

ARTICLE INFO

Keywords:

Gestational diabetes screening
Randomized clinical trial
Adherence

ABSTRACT

Background: ScreenR2GDM is a pragmatic randomized clinical trial designed to investigate if one of two gestational diabetes (GDM) screening and treatment protocols results in improved outcomes in the context of standard clinical care.

Methods: Pregnant women are randomized to one of two GDM screening strategies: *1-step*: 2-h, 75 g, oral glucose tolerance test (OGTT) or *2-step*: 1-h, 50 g glucose challenge test (GCT) followed by 3-h, 100 g OGTT if GCT-positive. Providers are prompted within the electronic medical record to order the assigned test but were given the option to order the alternate test. Collected data include maternal and pregnancy characteristics, GDM testing, and outcomes for mother and newborn. We describe the study design and baseline characteristics and evaluate characteristics associated with adhering to the randomized protocol.

Results: Baseline characteristics of the 23,792 randomized pregnancies were comparable between the two groups. Adherence to assigned test differed between the two strategies: 66.1% for 1-step and 91.7% for 2-step ($p < .0001$). 27% of the women randomized to receive the 1-step completed the 2-step test vs 2% randomized to the 2-step who completed the 1-step ($p < .0001$). Patient characteristics related to adherence included obesity, age, prior GDM, Medicaid insurance, race and nulliparity. Clinician characteristics related to adherence included provider type, age and gender.

Conclusions: Both patient and provider characteristics were related to adherence to the randomized GDM screening protocol. Analytical techniques that incorporate these findings into the formal evaluation of the two protocols on GDM-associated outcomes will be necessary to account for potential biases introduced by non-adherence.

1. Introduction

In 2013, the NIH held a consensus conference about screening and diagnosis of gestational diabetes (GDM) because there was great controversy about two clinically recommended screening strategies for GDM- a 2-step screening strategy and a 1-step, 2-h 75 g oral glucose tolerance test (OGTT) with pregnancy criteria. One recommendation of this conference was that ideal evidence would result from a head-to-head comparison of the two strategies [1]. Two randomized placebo-

controlled trials have demonstrated that treatment of women with mild hyperglycemia improves maternal and perinatal outcomes in pregnancies with GDM diagnosed using a 2-step screening strategy [2,3]. The Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) Study, a large multi-center prospective cohort study, showed a linear relationship with elevated glucose levels in pregnancy in women without diabetes and four related perinatal outcomes, based on a single 75 g OGTT [4]. Based on the HAPO Study, the American Diabetes Association (ADA) recommended that the 1-step 75 g screening approach for

* Corresponding author at: Hawaii Permanente Medical Group, Kaiser Permanente, 501 Alakawa St, Suite 201, Honolulu, HI 96817, USA.

E-mail addresses: kathy.pedula@kp.org (K.L. Pedula), teresa.hillier@kpchr.org (T.A. Hillier), keith.k.ogasawara@kp.org (K.K. Ogasawara), kimberly.k.vesco@kpchr.org (K.K. Vesco), suzanne.l.lubarsky@kp.org (S. Lubarsky), caryn.es.oshiro@kp.org (C.E.S. Oshiro), jan.vanmarter@kpchr.org (J. VanMarter).

<https://doi.org/10.1016/j.cct.2019.105829>

Received 18 March 2019; Received in revised form 5 August 2019; Accepted 13 August 2019

Available online 16 August 2019

1551-7144/ © 2019 Elsevier Inc. All rights reserved.

diagnosing GDM be adopted by clinicians to use in their practice [5].

The American College of Obstetrics & Gynecology (ACOG) continues to recommend using the 2-step screening strategy because it alone has been shown to improve perinatal outcomes in randomized trials and they did not deem there to yet be enough evidence to recommend adopting the 1-step approach [6].

Our pragmatic randomized trial of all pregnant women within a diverse population was designed to compare maternal and perinatal outcomes between the two screening and treatment protocols by randomizing over 17,000 diverse women to GDM screening (2-step vs. 1-step) as part of their clinical care in the Kaiser Permanente Northwest (KPNW) and Hawaii (KPH) regional health plans.

This report describes the development and methodology for the ScreenR2GDM study (clinicaltrials.gov identifier number NCT02266758), presents the baseline characteristics of randomized pregnancies and the factors associated with following the assigned random protocol. The Kaiser Permanente Northwest and Hawaii Institutional Review Boards approved the study and all consent procedures. All study procedures are carried out in accord with the Code of Ethics of the World Medical Association (Declaration of Helsinki) [7].

2. Methods

2.1. Study population

Pregnant women from the Kaiser Permanente Northwest (KPNW) and Hawaii (KPHI) regions were eligible for the study. Kaiser Permanente is a federally qualified not-for-profit, group-model health maintenance organization that provides integrated, comprehensive medical care to > 500,000 individuals in the greater Portland, Oregon area and > 200,000 members in Hawaii.

Women with diabetes prior to pregnancy were not randomized to receive GDM screening.

2.2. Randomization and blinding

Two GDM screening protocols are evaluated in this study. The 1-step strategy utilizes a fasting 2-h 75 g OGTT with thresholds recommended by International Association of the Diabetes and Pregnancy Study Groups (IADPSG) and the ADA, including a GDM diagnosis if one of the three values is elevated [8,9] The alternate 2-step strategy involves a non-fasting 50 g glucose challenge test (GCT); in the women who fail this, a fasting 3-h 100 g oral glucose tolerance test (OGTT) follows [10] (see Table 1).

Using a randomization scheme incorporated within the electronic medical record (EMR) and activated at the initial perinatal obstetric

Table 1
Gestational diabetes mellitus (GDM) screening methods and diagnostic thresholds.

	GDM screening test	
	1-Step	2-Step
75 g OGTT (fasting 2 h)		Step 1: 50 g glucose screen GCT (nonfasting 1 h) Step 2: 100 g OGTT (fasting 3 h) if GCT +
ADA glucose thresholds (mg/dL) [9]		Carpenter & coustan glucose thresholds (mg/dL) [10]
Fasting	92	95
1-h	180	180
2-h	153	155
3-h		140
GDM diagnosis	≥ 1 value abnormal	≥ 2 values abnormal

clinical visit in KPNW and KPHI, women were randomized to one of the two screening protocols. Both regions' IRBS approved a waiver of consent to allow randomization as part of clinical care as both screening strategies were clinically recommended minimal risk tests. The IRB approval included the pragmatic ethical requirement that providers or patients could "opt-out" of the randomized strategy and choose the alternate testing strategy. When the provider placed the order for GDM screening, they were presented with the randomly allocated test for order entry. Therefore, they were not blinded to their patients' random assignment. However, they did retain the option of ordering any test they deemed best for their patient.

Universal GDM Screening is typically done at 24–28 weeks gestation. However, both KPNW and KPHI regions have implemented a best practice to screen obese and high-risk women in the first trimester (Early Screen). Women who screen negative Early are then rescreened at the usual 24–28 weeks (Usual Screen). The electronically allocated random assignment accompanied a woman throughout her entire pregnancy, so that if she required screening more than once, the same test was provided to the clinician when he/she placed the order. Once the pregnancy ended, this random assignment was deleted from the EMR. Thus, for women who had more than one pregnancy during the study, each pregnancy was randomized independently.

2.3. Data collection and study measures

Patient data recorded in the regions' EMR is used to obtain the outcome measures. The EMR contains information on outpatient and inpatient encounters, diagnoses, prescribed medications, laboratory test orders, lab results, procedures ordered and completed, health problems, vital statistics and progress notes.

Primary outcome measures for the trial include diagnosis of GDM based on laboratory results, need for insulin treatment, maternal gestational hypertension, cesarean section, fetal macrosomia, and a composite perinatal outcome that includes any of the following: neonatal death, stillbirth, shoulder dystocia, bone fracture or nerve palsy. Other outcomes of interest include birthweight assessed as large-for-gestational age (LGA), gestational weight gain, neonatal complications (respiratory distress, jaundice, hypoglycemia), maternal development of postpartum diabetes and depression, and future childhood overweight and obesity.

Demographic and clinical characteristics were collected and compared in the present paper. Maternal pre-pregnancy weight and BMI were obtained from the electronic medical record. Measured pre-pregnancy weight was defined as the last measurement prior to estimated date of conception (up to 3 months); if that was not available, we used the earliest measurement during pregnancy (up to 12 weeks gestation). Maternal age, baby gender, Medicaid insurance and self-reported race and ethnicity were collected during health-plan enrollment. Prior GDM diagnosis and parity were obtained from the EMR; parity information is supplemented with data obtained from our locally maintained pregnancy databases. Provider type, gender and age were obtained from health plan administrative databases.

2.4. DSMB

A Data Safety Monitoring Board (DSMB) was established to oversee the study, including to suggest early termination for serious adverse effects, futility of study because of recruitment issues, or logistical or data quality problems so severe that correction was not feasible.

In addition to review of primary outcomes, the DSMB recommended and reviewed the following safety outcomes at mid-study: admissions to ICU or NICU, neonatal sepsis, preterm birth, and induction of labor. Findings from interim analyses were reviewed by the DSMB at study midpoint.

The PI (Hillier) and all study investigators and staff were blinded to the DSMB data review, with the exception of study biostatisticians. The

PI remained blinded through December 31, 2017, until the requisite number of women had been randomized. In early 2018, upon DSMB approval, the PI was unblinded.

3. Statistical methods

3.1. Sample size and power

The proposed primary analysis will evaluate the association of GDM testing protocol with maternal and perinatal complications in an intention-to-treat comparison of rates between these two screening groups: 1-step and 2-step. In addition to comparing testing protocols, we will evaluate the effect of GDM diagnosis on main outcomes as well as the interaction between screening test type and GDM diagnosis. The primary method of analyses will be generalized linear mixed model regression with dichotomous outcome, binomial error distribution and logit link function, a technique that can adjust for correlated data that may occur due to multiple pregnancies per woman during the study period. If there are few women with multiple pregnancies (i.e., many clusters with only one observation), we will randomly select one pregnancy per individual and use standard logistic regression analyses. In these models, we will include the 2 main effects (randomized screening test and test result [GDM diagnosis]) and their interaction, plus other covariates that may be related to the main effects or the outcomes. These include ethnicity, baseline BMI, maternal age, maternal weight gain, parity, baby gender and KP region in our models.

This study was designed to evaluate pregnancies occurring in each KP region, with initial estimates indicating that over 19,000 pregnancies would be eligible for screening during the study period. We designed the electronic randomization to exclude women with pre-existing diabetes. Women with prior bariatric surgery, < 18 years old, or those pregnancies resulting in multiple births (e.g. twins, triplets) would not be included in the formal analyses. Thus, our original power estimates were based on a sample size of 17,626 women and their offspring (> 35,000 total).

We conducted sample size estimates necessary to detect a range of effect sizes for comparison of main effects (in terms of relative risk given an anticipated prevalence of outcomes, a 5% [two-sided] Type 1 rate and 80% power). We assumed equal sample sizes in the 1-step vs 2-step comparisons. With these assumptions, we determined we could detect increased risk as low as 1.2 comparing the 2 testing protocols and an increased risk of 1.4 or larger in comparing GDM positive to GDM negative for all but the composite perinatal outcome. For the composite perinatal outcome, we can detect an increased risk of 1.4 in testing protocol and 1.6 between test results. Using the same assumptions, we can detect interaction (testing protocol x GDM status) effect sizes of 1.5–2.5. These estimates include adjustments for other possible covariates that have a squared correlation with the independent variables as high as 0.5.

3.2. Randomization and ordering of tests

The electronic randomization within the EMR was implemented May 28, 2014 in KPNW and in July 7, 2014 in KPHL. We conducted weekly monitoring of the randomization to ensure that approximately 50% of pregnancies were randomized to each testing protocol. While the testing protocol was assigned at the initial obstetrical visit, the actual test ordering was delayed until the appropriate gestational age: 1st trimester for obese and high-risk patients; approximately 22 weeks gestation for all others. In both regions, the initial OB visit occurred at an average of 10 weeks gestation, so we were unable to adequately evaluate initial protocol adherence until about 14 weeks into the study (approximately 24 weeks gestation). At that time, we determined that while the random assignment was balanced, there was a significant imbalance in ordering of the two GDM screening tests. At this point, the principal investigator and co-investigators met with the providers and

OB/GYN department stake-holders in each region to stress the value of the study and encourage adherence to random assignment. We continued to monitor adherence and provide the Obstetrics Chiefs in each region with updates. While, the adherence to random assignment improved slightly, there continued to be imbalance in test ordering.

At study midpoint, we conducted interim analyses and presented these to the DSMB. Due to a larger number of pregnancies than originally estimated, we were on track to randomize nearly 21,000. Yet, because of the imbalance in ordering of the tests, our overall estimates of GDM prevalence were much lower in the group randomized to 1-step because many patients received the alternate 2-step which diagnoses fewer GDM due to the higher cut points and requirement of 2 or more abnormal values (see Table 1). At the recommendation of the DSMB, we extended the study randomization an additional year (approximately 4 years total), adding approximately 3000 pregnancies to the newly estimated 21,000. As a result, the smaller GDM screening group (i.e. the group who received the 1-step despite balanced randomization) would have approximately the originally estimated number of pregnancies.

3.3. Adherence measures

Adherence to study protocol based on glucose testing was evaluated at two timepoints during the pregnancy: 1st trimester for obese and high-risk women (Early); 24–28 weeks for women without risk factors and those not diagnosed with GDM by Early screening. (Usual). To evaluate adherence to randomized study protocol (1-step vs 2-step testing), we evaluated whether a GDM screening test was completed and, if so, which test was conducted. If the randomly assigned test was administered, we classified adherence as “Assigned Study Test”. If the alternate test was used (e.g. randomized to 1-step but 2-step testing conducted instead), adherence was classified as “Alternate Study Test”. If another clinically acceptable test was conducted, adherence was classified as “Other Acceptable Test”. Other Acceptable Early Tests include hemoglobin A1c (HbA1c) and fasting plasma glucose (FPG), as both of these tests are also recommended for Early GDM screening in the first trimester [9]. For Usual testing, an elevated FPG (> 92 mg/dL) was commonly diagnosed and treated as GDM. Thus, if FPG is elevated (> 92 mg/dL) at Usual test, it was classified as “Other Acceptable Test”. All other glucose testing and those who do not receive a test were categorized as “No Test”. Therefore, the adherence measure consisted of 4 levels: Assigned Study Test, Alternate Study Test, Other Acceptable Test, and No Test. Only a value of “Assigned Test” was considered adherence. However, we evaluated “Alternate Study Test”, “Other Acceptable Test”, and “No Test” as separate non-adherent outcomes because factors related to each may be different.

3.4. Final diagnostic test

While the adherence outcomes described above were used in the present study to evaluate adherence at the Early and Usual screening, we also need to ascertain which test was ultimately used for GDM diagnosis and treatment. This is especially important for women who had more than one GDM screening test during pregnancy. In an “as-treated” analysis, the “final diagnostic test” would provide the comparison groups for evaluating the clinical trial because it is the test that would prompt treatment and likely have the greatest impact on pregnancy-related outcomes. Thus, we defined the final diagnostic test to be the screening test used for GDM (or elevated glucose) diagnosis and treatment. Identifying factors related to agreement of random assignment with final diagnostic test would provide the greatest insight as to whether protocol adherence needed to be considered in the intent-to-treat analyses. If diagnosed with GDM Early, the final diagnostic test was the test used Early. If diagnosed with GDM Usual, it was the Usual test. If not diagnosed with GDM by 1-step or 2-step, but FPG or HbA1c (Early) were elevated and there was no subsequent GDM screening test (1-step or 2-step), we assumed that the provider treated the elevated

glucose from other testing and the final diagnostic test was “Other Acceptable Test”. If there was no elevated glucose on any test, and one of the randomized GDM screening tests was conducted during the pregnancy, then it was the last of these study tests (1-step or 2-step) that determined the final diagnostic test.

3.5. Evaluation of adherence

We evaluated patient and provider characteristics related to protocol adherence (at both Early and Usual testing periods, as applicable) and to the final diagnostic test. For categorical variables we present frequencies and percentages of the total group of randomized subjects or subgroups of subjects. We summarized continuous variables using mean \pm SD. Statistical comparisons are performed using t- and chi-square tests as appropriate. Multinomial logistic regression was used to evaluate the independent associations of pregnancy and provider characteristics to each type (nominal level) of protocol adherence. We conducted the analyses with pregnancies as independent observations as well as clustered by provider. Intra-cluster correlation coefficients were < 0.008 for all models. Because the trial design was to randomize pregnancies (not clusters) and we randomized ALL pregnancies (no sampling), and the standard errors of the estimates for the analyses that accounted for clustering were similar to those without the cluster component, we report here the statistics from the most parsimonious models that do not consider clustering. As protocol adherence was strongly related to randomized assignment, we analyzed adherence separately for each random group. Analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC).

4. Results

4.1. Baseline comparisons

Demographic and clinical characteristics were similar between the randomly assigned groups (Table 2). Reflecting the differences in the population sizes, approximately 69% of the pregnancies were from the KP Northwest; 31% from KP Hawaii. Average age of women was 29 years; 29% were Obese (BMI > 30 kg/m²); 30% nulliparous; and 16% were on Medicaid (or the State's equivalent). The majority of women (56%) were white and 11% were multi-racial. The gestational age at the 1st screening test was 23 weeks for both groups (12.0 weeks mean gestational age for Early screening and 28.0 weeks for Usual). Of the 246 providers who ordered the tests, 60% were physicians; 26% certified nurse midwife; and 14% nurse practitioners. The remaining providers included Medical residents, Physician Assistants and Locum Tenens physicians (who may not have received the same training on randomization). Panel size ranged from 1 to 501 pregnancies for individual providers. Patient and provider characteristics listed in Table 2 were relatively balanced among the provider types with the following exceptions: Patients of the physicians had a slightly higher level of prior hypertension (9% vs 7% for other groups); physicians were younger, on average, than the other provider groups [mean (SD) 47(10) years for physicians vs 50(10) for others]; and the certified nurse midwife and nurse practitioner groups were over 90% female compared to the physicians for whom 72% were female.

4.2. Adherence to randomization

The relationship between random group assignment and completed test is shown in Table 3. The pregnancies randomized to receive the 1-step test had a substantially higher proportion of receiving the alternate study test compared to those randomized to 2-step protocol. This discordance was present at both Early and Usual screening. Other acceptable tests for diagnosing pre-existing diabetes in Early screening were comparable between the two random groups (10.4% for 1-step and 9.2% for 2-step, $p = .08$). Over 90% of patients randomized to 2-step

Table 2
Baseline characteristics of pregnancies randomized to two GDM screening strategies.

Pregnancy characteristics	Randomized assignment		
	1-step (n = 11,922)	2-step (n = 11,870)	Overall (n = 23,792)
Maternal age: mean(SD)	29.4 (5.5)	29.3 (5.5)	29.4 (5.5)
Gestational age at 1st test (weeks): mean(SD) (n = 22,710)	23.4 (8.0)	23.1 (8.2)	23.3 (8.1)
Parity: mean (SD)	1.5 (1.5)	1.5 (1.5)	1.5 (1.5)
Nulliparous	3642 (30.6%)	3616 (30.5%)	7258 (30.5%)
Maternal BMI (kg/m ²): mean (SD) (n = 23,785)	27.4 (6.7)	27.6 (7.0)	27.5 (6.8)
Maternal obesity (n = 23,785)	3360 (28.2%)	3433 (28.9%)	6793 (28.6%)
Prior hypertension	948 (8.0%)	976 (8.2%)	1924 (8.1%)
Prior GDM	636 (5.3%)	637 (5.4%)	1273 (5.4%)
Medicaid	1914 (16.1%)	1810 (15.3%)	3724 (15.7%)
Race			
Native Hawaiian/Pacific Islander	623 (5.2%)	619 (5.2%)	1242 (5.2%)
Asian	1789 (15.0%)	1782 (15.0%)	3571 (15.0%)
White	6608 (55.4%)	6586 (55.5%)	13,194 (55.5%)
American Indian	49 (0.4%)	50 (0.4%)	99 (0.4%)
Black	329 (2.8%)	328 (2.8%)	657 (2.8%)
Other Single Race	40 (0.3%)	42 (0.4%)	82 (0.3%)
Multiracial	1317 (11.1%)	1310 (11.0%)	2627 (11.0%)
Unknown/missing	1167 (9.8%)	1153 (9.7%)	2320 (9.8%)
Hispanic			
No	4650 (39.0%)	4586 (38.6%)	9236 (38.8%)
Yes	1343 (11.3%)	1280 (10.8%)	2623 (11.0%)
Unknown	5929 (49.7%)	6004 (50.6%)	11,933 (50.2%)
Provider type (n = 22,423) ^a			
Physician	6621 (59.2%)	6744 (60.1%)	13,365 (59.6%)
Certified nurse midwife	2876 (25.7%)	2844 (25.3%)	5720 (25.5%)
Nurse practitioner	1618 (14.5%)	1568 (14.0%)	3186 (14.2%)
Other provider	77 (0.7%)	75 (0.7%)	152 (0.7%)
Provider gender (n = 22,423)			
Female	9206 (82.3%)	9223 (82.1%)	18,429 (82.2%)
Male	1986 (17.7%)	2008 (17.9%)	3994 (17.8%)
Provider age: mean(SD)(n = 19,507)	48.6 (10.9)	48.8 (10.9)	48.7 (10.9)
Site			
KP-Northwest	8203 (68.8%)	8140 (68.6%)	16,343 (68.7%)
KP-Hawaii	3719 (31.2%)	3730 (31.4%)	7449 (31.3%)

^a Provider type is for the provider who ordered the GDM screening test. If no test was ordered, then it is the provider of the initial OB visit. *Other provider* includes Physician Assistants, Locum Tenens Physicians, and Medical Residents who may have not received the randomization training.

testing received the 2-step as the final diagnostic test. This compares to 66.1% of those randomized to 1-step who received the 1-step as the final diagnostic test ($p < .0001$). Other accepted tests used for final diagnosis were rarely conducted for either of the two groups. The group randomized to 1-step testing had slightly higher rates of “No Test” (6.5% vs 5.7%, $p = .005$).

As the type of non-adherence differed between the two randomized groups, we evaluated characteristics associated with non-adherence separately for each group. Odds ratios and 95% confidence intervals for the relationship of these characteristics to each type of final diagnostic test non-adherence are shown in Fig. 1. Statistics for Early and Usual adherence are provided in the supplemental tables. Among pregnancies randomized to 1-step, those who received the alternate (2-step) study test were less likely to be Asian, more likely to be on Medicaid, and have providers who were younger, and either nurse practitioners or certified nurse midwives (Fig. 1a).

Among pregnancies randomized to 2-step screening, women who received the alternate (1-step) study test were more likely to be older, have had prior GDM pregnancy, have a certified nurse midwife provider and less likely to be Native Hawaiian/Other Pacific Islander

Table 3
Protocol adherence: GDM screening (early and usual) tests received and GDM diagnostic test by random assignment.

Recommended for early (1st trimester) testing ^a				
Random assignment	Early test received			
	Assigned study test	Alternate study test	Other acceptable test	No test
1-step (n = 3617)	1061 (29.3%)	704 (19.5%)	375 (10.4%)	1477 (40.8%)
2-step (n = 3658)	1969 (53.8%)	19 (0.5%)	335 (9.2%)	1335 (36.5%)
Eligible for usual (24–28 weeks gestation) testing ^b				
Random assignment	Usual test received			
	Assigned study test	Alternate study test	Other acceptable test	No test
1-step (n = 11,335)	7575 (66.8%)	2878 (25.4%)	0 (0.0%)	882 (7.8%)
2-step (n = 11,618)	10,711 (92.2%)	89 (0.8%)	3 (0.03%)	815 (7.0%)
Final diagnostic test ^c				
Random assignment	Test determining GDM (or elevated glucose) diagnosis and treatment			
	Assigned study test	Alternate study test	Other acceptable test	No test
1-step (n = 11,922)	7880 (66.1%)	3247 (27.4%)	17 (0.1%)	778 (6.5%)
2-step (n = 11,870)	10,881 (91.7%)	281 (2.4%)	36 (0.3%)	672 (5.7%)

^a Pregnancies recommended for Early screening include obese and high-risk pregnancies.

^b Pregnancies eligible for Usual testing included those without GDM diagnosis Early.

^c Final diagnostic test is the screening test used for GDM (or elevated glucose) diagnosis and treatment. If diagnosed with GDM Early, the final diagnostic test was the test used Early. If diagnosed with GDM Usual, it was the Usual test. If not diagnosed with GDM by 1-step or 2-step, but FPG or HBA1c (Early) were elevated and there was no subsequent study test (1-step or 2-step), we assumed the provider treated the elevated glucose from other testing and the final diagnostic test was “Other Acceptable Test”. If there was no elevated glucose on any test, and one of the two randomized GDM screening tests was conducted during the pregnancy, then it was the last of these study tests (1-step or 2-step) that determined the final diagnostic test. For all other cases, the final diagnostic test category was “No test”.

(Fig. 1b).

Among those who had both Early and Usual screening, there was no significant relationship of lab results at Early screening to adherence at the Usual screening (data not shown).

5. Discussion

This randomized pragmatic clinical trial is designed to help resolve the current public policy debate on the potential benefits and risks of two accepted GDM testing strategies (1-step vs. 2-step) in clinical practice for which a sufficiently powered head-to-head comparison has not yet been done [10] and for which more evidence is needed to fill this current research gap [11]. It is the first head-to-head randomized comparison of the two protocols to be conducted in all pregnant women in the context of standard clinical care.

Our study allows us to evaluate the effectiveness of each screening strategy in normal practice, including all pregnant women and the

entire spectrum of hyperglycemia that typically is restricted ethically for clinical trials. Our study populations are representative of the regions from which they arise (greater Portland area and state of Hawaii) allowing for generalizations to similar populations. In contrast, the women in the Maternal Fetal Medicine Units (MFMU) Network and Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) trials had specific entry criteria including a requirement of pre-specified glucose levels of mild hyperglycemia [2,3]. The HAPO study, which used the 1-step screening to evaluate the dose-response linear relationship of glucose with maternal and perinatal outcomes allowed a wide range of hyperglycemia (excluding diabetes diagnosed before or during pregnancy); treatment was based on the diverse clinics' standard of care [4]. In contrast, the protocol for GDM monitoring and treatment in our study was consistent across all clinics.

Recently the United States Preventative Services Task Force (USPSTF) recommended, for the first time, GDM screening and treatment in pregnant women after 24 weeks gestation because evidence demonstrates identifying and treating women with GDM reduces risk for preeclampsia, fetal macrosomia and shoulder dystocia [12]. Yet, evidence about differences in outcomes based on GDM screening strategies is still lacking. Our study will help to fill in the gap on whether one of the two GDM screening strategies improves outcomes among pregnant women that receive the same treatment after GDM diagnosis.

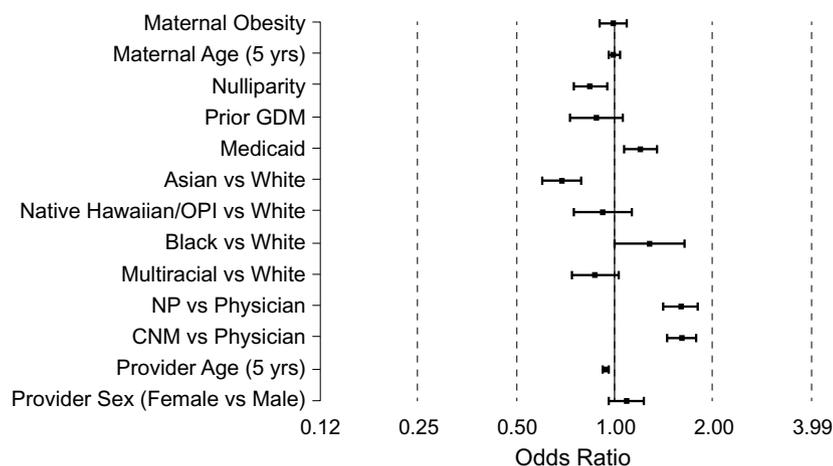
While we found several maternal and provider characteristics to be related to not receiving a test and/or to receiving the alternate study test or another acceptable test, it is not clear if these factors are directly related to non-adherence or if they are markers of other underlying reasons for not receiving the randomly assigned test. For example, women with obesity were more likely to receive an HbA1c or FPG test for Early screening, suggesting that the provider suspected undiagnosed pre-existing diabetes. Women who experienced GDM in a prior pregnancy were often monitored and treated as GDM without screening in subsequent pregnancies. Of those with prior GDM who were screened, they were more likely to receive the 1-step, possibly because providers assumed the risk of needing the second step (3 h, 100 g OGTT) is very high in women with prior GDM so it's most efficient to do the simpler 1-step (2-h 75 g OGTT).

While trying to address the greater degree of non-adherence for the 1-step, we received consistent feedback from providers in both regions on several practical considerations including the first step (50 g GCT) of the 2-step strategy can be ordered and conducted at the same OB visit when the provider orders the test, whereas the 1-step requires fasting and hence a subsequent visit. This is particularly relevant for Early testing, as women only see their providers monthly in the first trimester and providers wanted to be sure to not miss the opportunity for screening in these high-risk women. Consistent with other healthcare organizations who experienced a decrease in compliance with the 1-step [13], some of our providers reported that opting out of the 1-step improved the likelihood of completing at least one screening test. This is not an insignificant clinical issue as failure to adhere to any screening protocol could adversely impact outcomes more than the potential gain of using one strategy over another.

Additional anecdotal reasons for non-adherence to the 1-step randomization include women expressing their strong desire not to fast, the inconvenience of returning for lab testing for a second test, as well as other reasons (burden of committing 2+ hours for a test, aversion to multiple venous punctures, prior experience, etc.) that prompted the provider to order the 2-step test. Further, non-adherence rates varied greatly by provider (individual provider data not shown); a possible reason for this may be provider preference, including the familiarity with the 2-step strategy that had been used for many years prior to the randomization.

Among women who had both Early and Usual screening, we evaluated if the results of the Early test could be related to adherence at the Usual screening; we found no evidence to support this (data not shown). In our formal analyses of the clinical trial, we will consider

a. Odds of receiving 2-step when randomized to 1-step



b. Odds of receiving 1-step when randomized to 2-step

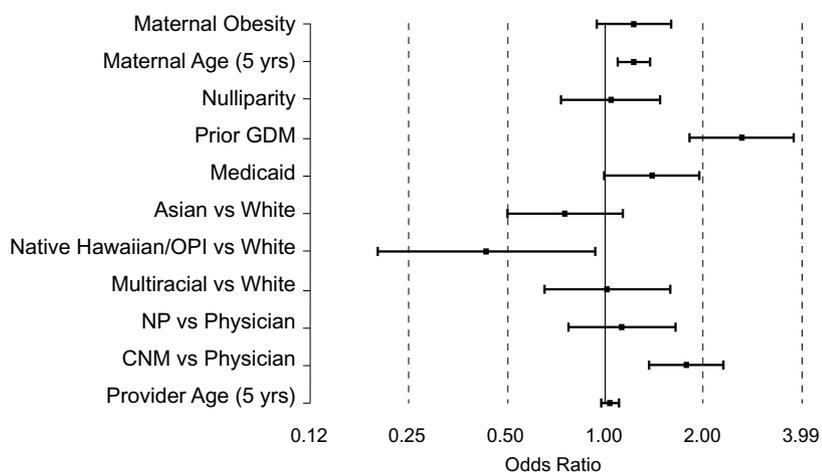


Fig. 1. Odds Ratios and 95% Confidence Intervals (horizontal bars) for factors related to receiving the alternate study test as the final diagnostic test among persons randomized to 1-step (Fig. 1a) and to 2-step (Fig. 1b) testing. A total of 11,922 patients were randomized to the 1-step; 3247 of these received the alternate study test (2-step). A total of 11,870 patients were randomized to the 2-step; 281 of these received the alternate study test (1-step). In Fig. 1b, estimates are not reported for Black vs White (due to small cell size) and provider sex (due to high correlation of CNM with gender). NP = Nurse Practitioner; CNM = Certified Nurse Midwife; OPI = Other Pacific Islander. Note: log scale is used for horizontal axis.

factors related to adherence at both Early and Usual screening as well as those related to the final diagnostic test. However, because the final diagnostic test is the one that likely impacts treatment and subsequent outcomes, we anticipate that adjustments to the intention-to-treat analyses will most likely involve factors associated with the final diagnostic test.

In conclusion, despite a balanced randomization scheme, the adherence to assigned testing protocol was not similar between the two randomized groups. Thus, the planned intention-to-treat analysis would likely be biased if alternative statistical methods are not also utilized to account for an imbalance in adherence. A number of analytical techniques have been proposed to adjust for non-adherence in randomized trials, including complier average causal effect (CACE), instrumental variable, principal score weighting and g-methods as well as other modifications to each [14–17]. Our updated analytical plan is to conduct and present results of several analyses to allow for fully-informed inference regarding these two GDM testing strategies.

Funding

This work was funded by the National Institutes of Health [grant number R01HD074794].

Declaration of Competing Interest

None

Acknowledgments

We thank Angela Paolucci for her excellent assistance with manuscript preparation and Lisa Fox for figure preparation.

Kaiser Permanente Northwest: Teresa Hillier, Kim Vesco, Jan VanMarter, Nancy Perrin, Suzanne Lubarsky, Ann MacFarlane, Sperry Robinson, Mike Krall, Kathryn Pedula.

Kaiser Permanente Hawaii: Teresa Hillier, Keith Ogasawara, Caryn Oshiro, Carmen Wong, Valentyna Pishchalenko, Yannica Theda Martinez, Amy Stone-Murai, Joanne Mor, Aurelio Candelario, Stacey Honda, Kathryn Pedula.

Data and Safety Monitoring Committee (thanks to all of members for their invaluable volunteered time): Jodi Lapidus Ph.D (chair), Aaron Caughey M.D., Ph.D, Jeanne-Marie Guise M.D., M.P.H., of Oregon Health and Sciences University. We'd also like to thank Dawn Peters, Ph.D for her service in DSMB charter planning.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cct.2019.105829>.

References

[1] J.P. Vandersten, W.C. Dodson, M.A. Espeland, W.A. Grobman, J.M. Guise, B.M. Mercer, H.L. Minkoff, B. Poindexter, L.A. Prosser, G.F. Sawaya, J.R. Scott, R.M. Silver, L. Smith, A. Thomas, A.T. Tita, NIH consensus development conference:

- diagnosing gestational diabetes mellitus, NIH Consens State Sci Statements. 29 (1) (2013) 1–31.
- [2] M.B. Landon, C.Y. Spong, E. Thom, M.W. Carpenter, S.M. Ramin, B. Casey, R.J. Wapner, M.W. Varner, D.J. Rouse, J.M. Thorp Jr., A. Sciscione, P. Catalano, M. Harper, G. Saade, K.Y. Lain, Y. Sorokin, A.M. Peaceman, J.E. Tolosa, G.B. Anderson, A multicenter, randomized trial of treatment for mild gestational diabetes, *N. Engl. J. Med.* 361 (14) (2009) 1339–1348.
- [3] C.A. Crowther, J.E. Hiller, J.R. Moss, A.J. McPhee, W.S. Jeffries, J.S. Robinson, Group ACISIPWAT, Effect of treatment of gestational diabetes mellitus on pregnancy outcomes, *N. Engl. J. Med.* 325 (24) (2005) 2477–2486.
- [4] Group THSCR, Hyperglycemia and adverse pregnancy outcomes, *N. Engl. J. Med.* 358 (19) (2008) 1991–2002.
- [5] Standards of MEDICAL CARE IN Diabetes – 2013, *Diabetes Care* 36 (Suppl. 1) (2013) S11–S66.
- [6] American College of Obstetricians and Gynecologists, ACOG practice bulletin no. 190: gestational diabetes mellitus, *Obstet. Gynecol.* 131 (2) (2018) e49–e64.
- [7] World, Medical association declaration of Helsinki: ethical principles for medical research involving human subjects, *J. Am. Coll. Dent.* 81 (3) (2014) 14–18.
- [8] Panel IAoDaPSGC, B.E. Metzger, S.G. Gabbe, B. Persson, T.A. Buchanan, P.A. Catalano, P. Damm, A.R. Dyer, A. Leiva, M. Hod, J.L. Kitzmiller, L.P. Lowe, H.D. McIntyre, J.J. Oats, Y. Omori, M.I. Schmidt, International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy, *Diabetes Care* 33 (3) (2010) 676–682.
- [9] American Diabetes Association *Diabetes Care* 42 (2019), pp. S13–S28.
- [10] Committee on Practice Bulletins–Obstetrics, Practice bulletin no. 137: gestational diabetes mellitus, *Obstet. Gynecol.* 122 (2) (2013) 406–416 Pt 1.
- [11] D.J. Wexler, C.E. Powe, L.A. Barbour, T. Buchanan, D.R. Coustan, R. Corcoy, P. Damm, F. Dunne, D.S. Feig, A. Ferrara, L.M. Harper, M.B. Landon, S.J. Meltzer, B.E. Metzger, H. Roeder, J.A. Rowan, D.A. Sacks, D. Simmons, J.G. Umans, P.M. Catalano, Research gaps in gestational diabetes mellitus: executive summary of a National Institute of Diabetes and Digestive and Kidney Diseases Workshop, *Obstet. Gynecol.* 132 (2) (2018) 496–505.
- [12] L. Hartling, D.M. Dryden, A. Guthrie, M. Muise, B. Vandermeer, W.M. Aktary, D. Pasichnyk, J.C. Seida, L. Donovan, Screening and diagnosing gestational diabetes mellitus, *Evid. Rep. Technol. Assess (Full Rep)*. 210 (2012) 1–327.
- [13] K.P. Fuller, A.F. Borgida, Gestational diabetes mellitus screening using the one-step versus two-step method in a high-risk practice, *Clin. Diabetes* 32 (4) (2014) 148–150.
- [14] G.W. Imbens, D.B. Rubin, Bayesian inference for causal effects in randomized experiments with noncompliance, *Ann. Stat.* 25 (1997) 305–327.
- [15] C.E. Frangakis, D.B. Rubin, Principal stratification in causal inference, *Biometrics*. 58 (1) (2002 March) 21–29.
- [16] J.D. Angrist, G.W. Imbens, D.R. Rubin, Identification of causal effects using instrumental variables (with discussion), *J. Am. Statist. Ass.* 91 (1996) 444–455.
- [17] P. Ding, J. Lu, Principal stratification analysis using principal scores, *J. R. Stat. Soc. B* 79 (2017) 757–777.