GYNECOLOGY

Accuracy of self-assessment of gestational duration among people seeking abortion

Check for updates

Lauren J. Ralph, PhD, MPH; Katherine Ehrenreich, MSc; Rana Barar, MPH; M. Antonia Biggs, PhD; Natalie Morris, MPH; Kelly Blanchard, MSc; Nathalie Kapp, MD, MPH; Ghazaleh Moayedi, DO, MPH; Jamila Perritt, MD, MPH, FACOG; Elizabeth G. Raymond, MD, MPH; Kari White, PhD; Daniel Grossman, MD

BACKGROUND: Mifepristone, used together with misoprostol, is approved by the United States Food and Drug Administration for medication abortion through 10 weeks' gestation. Although in-person ultrasound is frequently used to establish medication abortion eligibility, previous research demonstrates that people seeking abortion early in pregnancy can accurately self-assess gestational duration using the date of their last menstrual period.

OBJECTIVE: In this study, we establish the screening performance of a broader set of questions for self-assessment of gestational duration among a sample of people seeking abortion at a wide range of gestations.

STUDY DESIGN: We surveyed patients seeking abortion at 7 facilities before ultrasound and compared self-assessments of gestational duration using 11 pregnancy dating questions with measurements on ultrasound. For individual pregnancy dating questions and combined questions, we established screening performance focusing on metrics of diagnostic accuracy, defined as the area under the receiver operating characteristic curve, sensitivity (or the proportion of ineligible participants who correctly screened as ineligible for medication abortion), and proportion of false negatives (ie, the proportion of all participants who erroneously screened as eligible for medication abortion). We tested for differences in sensitivity across individual and combined questions using McNemar's test, and for differences in accuracy using the area under the receiver operating curve and Sidak adjusted *P* values.

RESULTS: One-quarter (25%) of 1089 participants had a gestational duration of >70 days on ultrasound. Using the date of last menstrual period alone demonstrated 83.5% sensitivity (95% confidence interval, 78.4—87.9) in identifying participants with gestational durations of >70 days on ultrasound, with an area under the receiver operating characteristic curve of 0.82 (95% confidence interval, 0.79—0.85) and a proportion of false negatives of 4.0%. A composite measure of responses to questions on number of weeks pregnant, date of last menstrual period, and date they got pregnant demonstrated 89.1% sensitivity (95% confidence interval, 84.7—92.6) and an area under the receiver operating curve of 0.86 (95% confidence interval, 0.83—0.88), with 2.7% of false negatives. A simpler question set focused on being >10 weeks or >2 months pregnant or having missed 2 or more periods had comparable sensitivity (90.7%; 95% confidence interval, 86.6—93.9) and proportion of false negatives (2.3%), but with a slightly lower area under the receiver operating curve (0.82; 95% confidence interval, 0.79—0.84).

CONCLUSION: In a sample representative of people seeking abortion nationally, broadening the screening questions for assessing gestational duration beyond the date of the last menstrual period resulted in improved accuracy and sensitivity of self-assessment at the 70-day threshold for medication abortion. Ultrasound assessment for medication abortion may not be necessary, especially when requiring ultrasound could increase COVID-19 risk or healthcare costs, restrict access, or limit patient choice.

Key words: gestational age, gestational duration, last menstrual period, medication abortion, mifepristone, misoprostol, over-the-counter, pregnancy dating, pregnancy duration, self-selection, telehealth, ultrasound

Introduction

Medication abortion (MA) with the 2-drug regimen of mifepristone and misoprostol is currently approved by the US Food and Drug Administration (FDA) for use through 70 days' (or 10 weeks') gestation. Although in-person ultrasounds or physical exams are frequently used to establish gestational duration (GD) and eligibility for MA, there is growing adoption of models of care that forgo ultrasounds or physical exams and/or conduct eligibility screening virtually.[1−3] However, widespread uptake of these models is limited by state-level restrictions explicitly prohibiting telemedicine provision of abortion or requiring an in-person ultrasound or counseling visit.[4,5] Further, some providers and regulators remain concerned that most people seeking abortion need an ultrasound or physical exam to accurately determine the GD of their pregnancy and, by extension, their eligibility for MA.

Previous research suggests that people seeking abortion early in pregnancy are generally accurate in self-assessing GD using the date of their last menstrual period (LMP), particularly when the goal is minimizing the number of people who screen as eligible when they are not.[6] In the largest US-based study to date, 3.5% of participants who self-assessed their GD using the date of their LMP as ≤63 days (9 weeks) were determined to have a GD of >63 days on ultrasound.[7] Fewer participants (1.2%) had a GD of >70 days.[8] However, a 2014 review article found that the proportion of participants erroneously self-assessing their GD as making them eligible for MA ranged from 2.5% to 11.8% in different populations globally.[6] Further, larger studies were typically restricted to people seeking

Cite this article as: Ralph LJ, Ehrenreich K, Barar R, et al. Accuracy of self-assessment of gestational duration among people seeking abortion. Am J Obstet Gynecol 2022;226:710.e1-21.

0002-9378

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.1016/j.ajog.2021.11.1373

AJOG at a Glance

Why was this study conducted?

In-person ultrasound is typically used to establish gestational duration (GD) and eligibility for medication abortion (MA), but there is interest in expanding the use of telemedicine screening and removing ultrasound requirements. In this study, we examined whether patient's self-assessment of GD using the date of their last menstrual period (LMP) and other questions results in accurate assessment of MA eligibility.

Key findings

Responses to questions on date of LMP, date they got pregnant, and number of weeks pregnant, and statements about being >10 weeks pregnant, >2 months pregnant, or having missed more than 2 periods demonstrated very good sensitivity and accuracy in identifying people with GD of >70 days on ultrasound. When questions were used in combination, false negatives (ie, participants erroneously screening as having GD of ≤70 days) were reduced to as low as 2.3%, compared with the 4% of false negatives when using date of LMP alone.

What does this add to what is known?

Expanding the types of pregnancy dating questions allows for more comprehensive, accurate, and sensitive self-assessment of GD among people seeking abortion.

abortion in early pregnancy, and there is evidence in smaller samples that those later in gestation are more likely to underestimate GD using the date of their LMP.[9,10] Finally, reliance on the date of LMP alone may be unnecessarily restrictive. Studies of people seeking abortion suggest that 20% to 29% are not certain about the date of their LMP,[7,11,12] and up to 1 in 5 people report irregular periods or frequently spotting between periods,[11,13] both of which are associated with later presentation for abortion.[11,12,14]

In this study, we examined the diagnostic accuracy of a broad set of pregnancy dating questions—alone and in combination—in identifying people eligible for MA on the basis of GD. We examined questions presented as they would be on a label for an over-the-counter (OTC) MA product and in an electronic questionnaire that could be completed online, in a pharmacy or clinic waiting room, or during a telehealth consultation. This study offers new tools to support effective methods for self-assessment of GD and ultimately expanded access to abortion care via telemedicine in the present and in a future OTC context.

Materials and Methods

We recruited participants from 7 freestanding abortion facilities in Alabama, California, Florida, Illinois, North Dakota, Texas, and Washington, District of Columbia between October 2019 and March 2020. Sites providing care beyond the first trimester were selected to achieve geographic and policy diversity.

A trained research assistant approached all patients in facility waiting rooms. Those interested in the study completed an eligibility screening tool; eligibility was restricted to those aged ≥15, able to speak and read English or Spanish, seeking an abortion (medication or surgical), and who had not yet had an ultrasound at that facility. There were no inclusion criteria regarding type of abortion or estimated gestation. To align with state or facility policies, parental consent was obtained for 15- to 17-year-olds in North Dakota and Texas, and only people aged >17 were eligible in Alabama.

After research assistants obtained verbal consent, they gave participants a tablet to complete the survey and a study index card. They instructed participants to give the index card to their ultrasound technician. Participants received a $25 Amazon gift card as remuneration. All study activities were approved by the University of California, San Francisco Institutional Review Board.

Measures

GD in weeks and days on ultrasound was recorded by ultrasound technicians. Transvaginal or transabdominal ultrasound was performed in the standard fashion at each facility, generally using crown-rump length up to 14 weeks' gestation and biparietal diameter, head circumference, and/or femur length after 14 weeks' gestation. When GD could not be determined, technicians recorded the reason. Dichotomous variables were created for GD of >56, >63, >70, and >77 days on ultrasound. Despite measurement error, which increases as pregnancy progresses, ultrasound is considered the standard of care for assessing pregnancy duration and serves as the reference standard in this study.[15]

Survey questions were developed by the research team and refined on the basis of input from members of a community advisory board composed of medical and nonmedical experts and feedback from 11 cognitive interviews conducted with people seeking abortion. The survey included 2 modules assessing GD. The first included statements as they might appear on a hypothetical MA Drug Facts Label (DFL), with response options limited to "Yes," "No," or "Not sure." The second was structured as a self-administered questionnaire allowing for varied question formats and skip patterns (Table 1). The order in which participants saw the 2 modules was randomized.

LMP-based GD was calculated by subtracting the date of the survey from the reported date of LMP. GD based on date of fertilization was calculated by subtracting the date of the survey from the reported date they got pregnant, and then adding 14 days. When these calculations resulted in nonsensical values (eg, negative or ≤14 days), we applied a set of recoding rules (Supplement).

For most pregnancy dating questions, we dichotomized responses as >70 days (or ≥10 weeks if the question was phrased in week format) vs earlier. For certain items in the Questionnaire

TABLE 1
Description of survey items and response options

Question type	Module 1: Drug facts label (Response options were *Yes, No, Not sure* for all questions)	Module 2: Self-administered questionnaire (Response options in *italics*)
Pregnancy dating questions		
	You are more than 2 months pregnant. You are more than 10 weeks pregnant. You had a positive pregnancy test for this pregnancy more than 7 weeks ago. Your last normal period started more than 8 weeks ago. Your last normal period started more than 10 weeks ago. Since you got pregnant, you have missed more than 2 periods.	How many weeks pregnant do you think you are today? (*Less than 4, 5, 6, 7, 8, 9, 10, More than 10*) Do you think you got pregnant in the last 2 months? (*Yes, No, Not sure*) Select the date when you think you got pregnant. (calendar) Approximately when was the first day of your last period? *(Less than 4 weeks ago, 5, 6, 7, 8, 9, 10, More than 10 weeks ago)* Using the calendar, please select the first day of your last period. (calendar) Did you have a positive pregnancy test before (date 49 days ago)? (*Yes, No, Not sure*)
Other questions		
Irregular periods	In the 6 months before this pregnancy, you were having periods about once a month.	How often do your periods usually come? (*Every 3 weeks or shorter, Every 4 weeks or about once a month, Every 5 weeks or longer, It's different every time, Not sure*)
Hormonal contraceptive use	You had a birth control shot (Depo) within the 4 months before you got pregnant. You were taking birth control pills when you got pregnant. You had a birth control implant in your arm when you got pregnant.	Around the time when you think you got pregnant, did you use any of the following methods of birth control? (*List of all methods; select all that apply*)
Discussion with healthcare providers about pregnancy duration	A doctor or nurse has told you how far along you are in this pregnancy.	Have you already had an ultrasound or seen a doctor or nurse at another clinic who told you how far along in this pregnancy you are? (*Yes, I had an ultrasound; Yes, a doctor or nurse told me how far along; No*)

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

Module, we also created dichotomous measures of >56 days or ≥8 weeks, >63 days or ≥9 weeks, and >77 days or ≥11 weeks (vs earlier). We recoded nonresponse (skipped the question or selected "Not sure") as GD above the threshold or ineligible for MA.

Both modules included items related to regularity of periods, hormonal contraceptive (HC) method use, and previous pregnancy dating (Table 1). Participants who selected "Yes" to the "periods about once a month" statement in the DFL Module or "every 4 weeks or about once a month" in the Questionnaire Module were classified as having regular periods; all other responses, including "Not sure," were considered irregular periods. In the DFL Module, participants who selected "Yes" to statements about injectable, pill, or implant use were classified as HC users; all others, including "Not sure," were considered nonusers. In the Questionnaire Module, participants who selected birth control pills, vaginal rings, patches, implants, or injectables were classified as HC users.

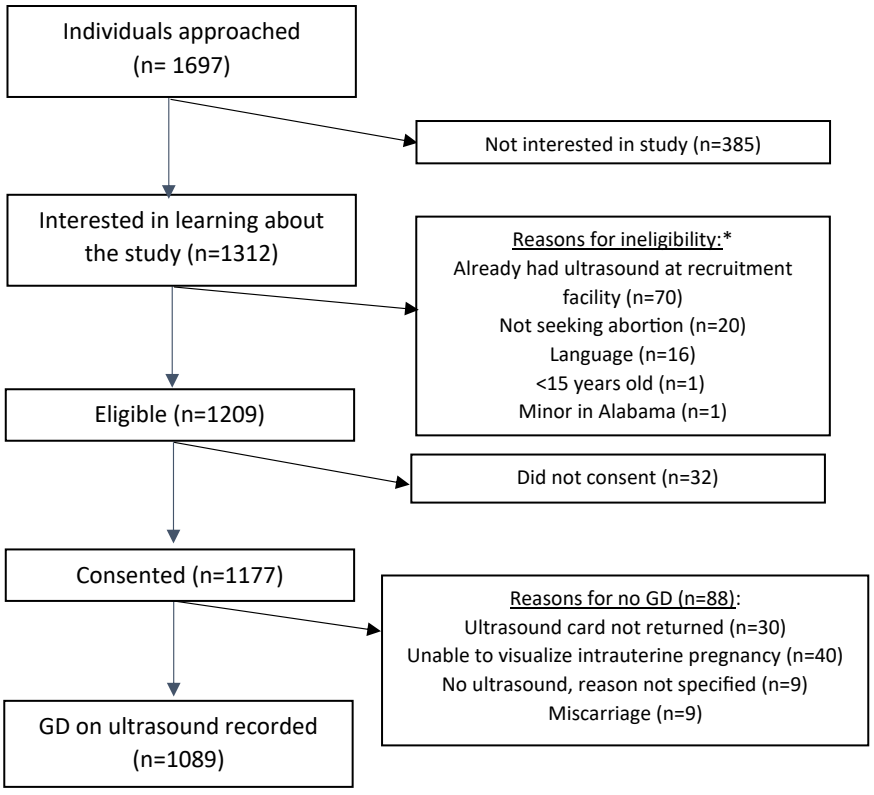
Demographics were assessed at the end of the survey and included age, race and ethnicity, place of birth (in or outside the United States), highest level of education, number of previous births, and household income and size. Household poverty level was calculated using 2019 thresholds.[16]

Analysis

We presented descriptive statistics and compared the sample to people seeking abortion nationally using the Guttmacher Institute's 2014 Abortion Patient Survey[17] (for socio-demographics) and 2017 Abortion Provider Census[18] (for pregnancy duration). We examined concordance between GD estimated by date of LMP and by calculating and summarizing the difference between these 2 values and generating 2-way scatterplots.

We described nonresponse to each pregnancy dating question and examined the screening performance of questions, alone and then in combination, according to 6 indicators:

FIGURE 1
Study flowchart



The *asterisk* represents participants could be ineligible for multiple reasons.

GD, gestational duration.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), area under the receiver operating characteristic curve (AUC), and proportion of false negatives in the total sample. Indicators were derived using the 2×2 table and formulas presented in the Supplement; a positive result was defined as greater than the specified threshold (most often 70 days). FDA guidance on self-selection considers accuracy, sensitivity, and NPV to be most important when determining whether someone can self-select to take a medication, for example in an OTC environment.[19] Thus, although most previous research on GD self-assessment has focused on the proportion of false negatives,[6] in this study we also prioritized performance regarding sensitivity and AUC, which captures overall accuracy or balance between sensitivity and specificity. AUC also has the advantage of not varying with outcome prevalence. For AUC, we considered values of 0.7 to 0.8 as good, 0.8 to 0.9 as very good, and 0.9 to 1.0 as excellent in terms of diagnostic accuracy.[20]

We compared AUC values of individual dating questions using Sidak adjusted *P* values to account for paired comparisons; Stata software's roccomp command was used for calculations.[21] We then combined responses using individual questions with AUC >0.75 or with a factor loading >0.60 in exploratory factor analysis (Supplement B provides details). We combined questions in 2 ways. First, in scaffolding, we layered responses according to question accuracy (from higher to lower), using responses to higher-accuracy questions unless they were missing, in which case we used responses to the next most accurate item. In the second, composite measure of responses, we recoded participants as ineligible if their response to any question classified them as such.

We tested the sensitivity and AUC of combinations of questions with the goal of arriving at 1 combination per module with the fewest items (ie, lowest participant burden) and highest sensitivity and AUC, and then contrasted performance between the best performing DFL and questionnaire combinations. Differences in sensitivity were assessed using McNemar's test for paired data. As mentioned previously, differences in AUC were assessed using receiver operating characteristic curves and Sidak adjusted *P* values.

Once we identified the optimal combination by module, we tested whether sensitivity and AUC improved with the addition of ineligibility criteria of HC use and/or irregular periods. We also tested whether sensitivity and AUC improved by shifting self-reported GD to 56 and 63 days, keeping ultrasound at >70 days. In descriptive analysis, we calculated the number of false negative cases that were appropriately reclassified by moving from reliance on date of LMP alone to the 3-item composite of questionnaire items.

Finally, in sensitivity analyses, we examined whether performance was similar: (1) when shifting the threshold to 77 days; (2) among participants who reported no previous ultrasound; and (3) on the basis of the order in which participants saw the modules.

We sought to estimate whether the rate at which people falsely identify themselves as eligible (proportion of false negatives) for MA is reduced by at least 1.5% with the use of a combined question; a sample size of 1000 provides 86% power to detect this difference assuming, consistent with previous research,[7] that the proportion of false negatives using date of LMP alone is 3.5%. We used Stata, version 15 for all analyses (version 15; StataCorp, College Station, TX); L.J.R., N.M., and- M.A.B. conducted analyses between March 20, 2020 and December 11, 2020. This manuscript was prepared in accordance with Standards for the Reporting of Diagnostic Accuracy Studies (STARD) guidelines.[22]

Results

Of 1697 individuals approached, 1312 (77%) were interested in learning about the study. Most (n=1209) were eligible, and 1177 provided consent and initiated the survey. The final analytical sample included 1089 participants for whom GD measured by ultrasound was available; the most common reason for no reported GD was no visualization of intrauterine pregnancy (n=40) (Figure 1).

Most (n=1039, 95.4%) participants reached the end of the survey. Median age was 26 years (interquartile range, 22–30). Approximately one-third of participants were Black (35.4%), another third White (31.0%), and 1 in 5 (18.0%) Hispanic or Latina/Latinx. Four in 10 participants (42.2%) were nulliparous (Table 2). The participants' demographic profiles and GD measured by ultrasound mirrored those of people seeking abortion nationally (Supplemental Table 1). Mean GD on ultrasound was 61.7 days, and median GD was 53 days, ranging from 28 to 224 days. One-quarter of participants (25%) had a GD >70 days.

Missingness and accuracy of Drug Facts Label Module items

Of the 5 DFL questions, having "missed >2 periods" produced the lowest nonresponse (6% vs 12%–23% for other questions). Nonresponse was highest for the statement of being ">2 months pregnant" (23%) (Table 3).

The DFL statement with highest AUC was ">10 weeks" (0.82; 95% confidence interval [CI], 0.79–0.85), which performed significantly better than ">2 months" (0.75; 95% CI, 0.73–0.78), "positive pregnancy test 7 weeks ago" (0.69; 95% CI, 0.66–0.72), and "last normal period started 10 weeks ago" (0.74; 95% CI, 0.71–0.77) at $P<.001$. AUC for "missed >2 periods" was also high (0.79; 95% CI, 0.77–0.82) and similar to ">10 weeks" ($P=.374$) (Table 3).

Missingness and accuracy of Questionnaire Module items

Of the 6 questions, "number of weeks pregnant" had the lowest nonresponse (3%). Questions focused on dates of LMP and date they got pregnant had the highest nonresponse at 14%.

TABLE 2
Sociodemographic and pregnancy characteristics of the study sample (N = 1089)

Characteristics	No.	Percent
Age (y)		
15–17	26	2.4
18–19	84	7.7
20–24	331	30.4
25–29	305	28.0
30–34	187	17.2
35–39	73	6.7
>40	34	3.1
Missing	49	4.5
Race/ethnicity		
Non-Hispanic White	337	31.0
Non-Hispanic Black	385	35.4
Hispanic/Latina/Latinx	196	18.0
Non-Hispanic Asian or Pacific Islander	50	4.6
Other/Multiple	81	7.4
Missing	40	3.7
Nativity		
US-born	913	83.8
Born outside the United States	136	12.5
Missing	40	3.7
Highest level of education completed		
<High school	74	6.8
High school diploma or equivalent	314	28.8
Some college/associate's degree	436	40.0
College graduate or higher	225	20.7
Missing	40	3.7
No. of previous births		
0	459	42.2
1	239	22.0
≥ 2	354	32.5
Missing	37	3.4
Household income as % of FPL		
<100% FPL	375	34.4
100%–199% FPL	211	19.4
≥200% FPL	217	19.9
Missing	286	26.3
Language of survey		
English	1070	98.3
Spanish	19	1.7

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022. *(continued)*

TABLE 2
Sociodemographic and pregnancy characteristics of the study sample (N = 1089) *(continued)*

Characteristics	No.	Percent
Reported previous pregnancy dating		
Ultrasound at another facility	198	18.1
Discussed with a doctor or nurse	141	13.0
GD on ultrasound (d)		
<36	42	3.8
36–42	125	11.5
43–49	255	23.4
50–56	176	16.2
57–63	129	11.9
64–70	88	8.1
71–77	64	5.9
78–84	49	4.5
85–91	34	3.1
92–98	27	2.5
>98	100	9.2

FPL, federal poverty level; *GD*, gestational duration.
Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

The question with the highest AUC was "number of weeks pregnant" (0.87; 95% CI, 0.84–0.89); this was higher than the AUC of the date of LMP question (0.82; 95% CI, 0.79–0.85) at P=.007. The questions on date they got pregnant and number of weeks since LMP performed similarly to the question on date of LMP on AUC (Table 3).

Supplemental Figure 2, A and B depict general concordance between GD estimated by date of LMP/date they got pregnant and GD measured by ultrasound. The mean difference between GD estimated by date of LMP and GD measured by ultrasound was −0.10 (95% CI, −1.03 to 0.83); the mean difference between GD estimated by date they got pregnant and GD measured by ultrasound was −4.0 (95% CI, −4.90 to −3.00).

Accuracy and sensitivity of Drug Facts Label Module question combinations

A composite measure of DFL Module questions ">10 weeks," "missed >2 periods," or ">2 months" demonstrated 90.8% sensitivity (95% CI, 86.7–93.9) and AUC of 0.82 (95% CI, 0.79–0.84) in identifying participants with GD of >70 days on ultrasound. This sensitivity was greater than that of using just 2 questions (">10 weeks," "missing >2 periods") at P=.024. AUC increased slightly if the 3 responses were scaffolded (0.84; 95% CI, 0.81–0.87; P=.043) (Module 1/DFL item combinations in Table 4); however, scaffolding resulted in significantly lower sensitivity (78.6; 95% CI, 73.2–83.3) than that of the 3-item composite (P<.001, not shown) (Module 1/DFL item combinations in Table 4).

Accuracy and sensitivity of Questionnaire Module question combinations

A composite measure of questions on number of weeks pregnant, date of LMP, and date they got pregnant demonstrated 89.3% sensitivity (95% CI, 85.0%–92.7%) and AUC of 0.86 (95% CI, 0.83–0.88) in identifying participants with GD of >70 days on ultrasound. AUC was comparable when these 3 questions were scaffolded (0.87; 95% CI, 0.84–0.89); however, scaffolding resulted in significantly lower sensitivity (P<.001, not shown) (Module 2/Questionnaire item combinations in Table 4). At a 77-day GD threshold, the 3-item scaffolded composite had sensitivity of 86.0% (95% CI, 80.5–90.4) and AUC of 0.87 (95% CI, 0.84–0.90) (Supplemental Table 2).

Question combinations of the Drug Facts Label Module vs Questionnaire Module

The 3-item composite measure prioritizing an ineligible response to the Questionnaire Module items of number of weeks pregnant, date of LMP, or date they got pregnant performed similarly in sensitivity to the composite measure of DFL Module items of ">10 weeks pregnant," "missed >2 periods," or ">2 months pregnant" (89.1 vs 90.7; P=.327). However, AUC of the Questionnaire Module composite was significantly higher (0.86 vs 0.82; P=.002).

Adding hormonal contraceptive use or irregular periods

For both DFL and Questionnaire Module composite measures, the addition of HC use and irregular periods significantly increased sensitivity and reduced AUC (Combinations with hormonal contraceptive use and irregular period as additional layers in Table 4).

Lowering the threshold of self-reported pregnancy duration

For the composite questionnaire measures, shifting self-reported GD to 56 or 63 days significantly improved sensitivity and reduced AUC. Using the 56-day self-reported cutoff resulted in the lowest false-negative proportion (0.65%) but also the lowest specificity (49.3%) and PPV (38.8%) (Lower dating threshold with 70-d gestational age on ultrasound (questionnaire only) in Table 4).

False negatives

Using the date of LMP alone and a 70-day threshold, there were 42 false negatives. One-half (23 of 42) of these were appropriately reclassified as ineligible when using the composite questionnaire measure of number of weeks pregnant, date they got pregnant, and date of LMP (Figure 2).

TABLE 3
Missingness and screening performance of individual pregnancy dating items

	No. not sure or skipped /No. saw question	Percent not sure or skipped	Sensitivity: % ineligible who screened as ineligible (95% CI)	Specificity: % eligible who screened as eligible) (95% CI)	Negative predictive value (95% CI)	Positive predictive value (95% CI)	AUC (95% CI)	Sidak *P* value comparing AUC values	Percent of false negatives	No. of false negatives/ total N
Module 1/DFL items										
Missed >2 periods since conception	62/1075	5.8	84.4 (79.5–88.5)	73.9 (70.8–76.9)	93.4 (91.2–95.2)	51.9 (47.1–56.7)	0.79 (0.77–0.82)	.374	3.91	42/1075
Positive pregnancy test >7 wk ago	132/1078	12.2	69.6 (63.8–75.1)	68.7 (65.4–71.9)	87.1 (84.3–89.6)	42.6 (38.0–47.4)	0.69 (0.66–0.72)	<.001	7.61	82/1078
Last normal period started >10 wk ago	173/1075	16.1	80.3 (75.0–84.9)	68.0 (64.6–71.2)	91.2 (88.6–93.3)	45.6 (41.0–50.2)	0.74 (0.71–0.77)	<.001	4.93	53/1075
>10 wk pregnant	200/1081	18.5	84.1 (79.2–88.3)	79.9 (76.9–82.6)	93.8 (91.7–95.5)	58.3 (53.2–63.2)	0.82 (0.79–0.85)	ref	3.98	43/1081
>2 mo pregnant	247/1081	22.8	85.6 (80.9–89.6)	65.2 (61.8–68.5)	93.1 (90.7–95.1)	45.1 (40.8–49.6)	0.75 (0.73–0.78)	<.001	3.61	39/1081
Module 2/Questionnaire items										
Number of weeks pregnant	37/1079	3.4	81.9 (76.8–86.3)	91.2 (89.0–93.1)	93.8 (91.8–95.3)	75.8 (70.4–80.6)	0.87 (0.84–0.89)	.007	4.55	49/1079
Got pregnant within the past 2 mo	73/1079	6.8	45.8 (39.7–51.9)	86.3 (83.7–88.6)	82.6 (79.9–85.1)	52.8 (46.2–59.3)	0.66 (0.63–0.69)	<.001	13.62	147/1079
Number of weeks since first day of LMP[a]	104/1041	10.0	81.6 (76.3–86.1)	79.0 (76.0–81.8)	93.0 (90.8–94.8)	55.8 (50.6–60.9)	0.80 (0.78–0.83)	.213	4.51	47/1041
Positive pregnancy test >7 wk ago[a]	104/1012	10.3	60.2 (53.9–66.4)	82.7 (79.8–85.3)	86.4 (83.7–88.8)	53.2 (47.2–59.1)	0.71 (0.68–0.75)	<.001	9.78	99/1012
Date of LMP[a]	147/1041	14.1	83.5 (78.4–87.9)	80.2 (77.2–82.9)	93.8 (91.6–95.5)	57.7 (52.5–62.8)	0.82 (0.79–0.85)	ref	4.03	42/1041
Reported date they got pregnant	155/1079	14.4	88.5 (84.0–92.0)	74.2 (71.0–77.2)	95.0 (93.0–96.6)	53.5 (48.7–58.2)	0.81 (0.79–0.84)	.405	2.90	31/1079

For all questions, respondents who selected "Not sure" or skipped the question were coded as ineligible for medication abortion in calculating screening performance measures.

AUC, area under the receiver operating characteristic curve; *CI*, confidence interval; *DFL*, Drug Facts Label; *LMP*, last menstrual period; *ref*, reference group in significance testing.

[a] Sample sizes lower because of a skip pattern in the survey; if a respondent reported no period in the last year, they did not see date of LMP questions. Similarly, if a respondent reported not taking a pregnancy test, they skipped the pregnancy test timing question.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

TABLE 4
Screening performance of combinations of pregnancy dating items, overall and with inclusion of additional questions for hormonal contraceptive use and irregular periods and shifting of gestational duration thresholds

	Sensitivity (95% CI)	McNemar's *P* value comparing sensitivities	Specificity (95% CI)	Negative predictive value (95% CI)	Positive predictive value (95% CI)	AUC (95%CI)	Sidak *P* value comparing AUC values	Percent of false negative	No. of false negatives/ total N
Module 1/DFL item combinations									
Scaffold									
>10 wk > Missed 2 periods >> 2 mo	78.6 (73.2–83.3)	ref	89.8 (87.5–91.8)	92.6 (90.6–94.3)	72.0 (66.5–77.0)	0.84 (0.82–0.87)	ref	5.34	58/1081
>10 wk > Missed 2 periods	79.0 (73.6–83.7)	>.999	89.4 (87.1–91.4)	92.7 (90.6–94.4)	71.3 (65.9–76.4)	0.84 (0.81–0.87)	.997	5.28	57/1081
Composite									
>10 wk, missed 2 periods, or >2 mo preg[a]	90.8 (86.7–93.9)	ref	72.6 (69.4–75.6)	95.9 (94.0–97.3)	52.6 (47.9–57.2)	0.82 (0.79–0.84)	ref	2.31	25/1081
>10 wk or missed 2 periods	86.0 (81.3–89.9)	.024	75.1 (71.9–78.0)	94.1 (92.0–95.8)	53.6 (48.8–58.3)	0.81 (0.78–0.83)	.333	3.51	38/1081
Module 2/Questionnaire item combinations									
Scaffold									
No. of weeks pregnant > date of LMP > date got pregnant > weeks since LMP	80.4 (75.2–85.0)	ref	92.7 (90.7–94.4)	93.4 (91.5–95.0)	78.7 (73.4–83.4)	0.87 (0.84–0.89)	ref	4.91	53/1079
No. of weeks pregnant > date of LMP > date got pregnant	80.4 (75.2–85.0)	>.999	92.7 (90.7–94.4)	93.4 (91.5–95.0)	78.7 (73.4–83.4)	0.87 (0.84–0.89)	>.999	4.91	53/1079
No. of weeks pregnant > date of LMP	80.8 (75.6–85.3)	>.999	92.9 (91.0–94.6)	93.5 (91.6–95.1)	79.3 (74.1–84.0)	0.87 (0.84–0.89)	.777	4.81	52/1079
Composite									
No. of weeks pregnant, date of LMP, date got pregnant, or weeks since LMP	91.1 (87.1–94.2)	ref	79.1 (76.1–81.8)	96.4 (94.7–97.7)	59.4 (54.5–64.1)	0.85 (0.83–0.87)	ref	2.22	24/1079

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

(continued)

TABLE 4
Screening performance of combinations of pregnancy dating items, overall and with inclusion of additional questions for hormonal contraceptive use and irregular periods and shifting of gestational duration thresholds *(continued)*

	Sensitivity (95% CI)	McNemar's *P* value comparing sensitivities	Specificity (95% CI)	Negative predictive value (95% CI)	Positive predictive value (95% CI)	AUC (95%CI)	Sidak *P* value comparing AUC values	Percent of false negative	No. of false negatives/ total N
No. of weeks pregnant, date of LMP, date got pregnant[a]	89.3 (85.0–92.7)	.063	81.9 (79.1–84.5)	95.8 (94.0–97.2)	62.4(57.3–67.2)	0.86 (0.83–0.88)	.547	2.68	29/1079
No. of weeks pregnant, date of LMP	83.8 (78.8–87.9)	<.001	88.9 (86.5–90.9)	94.2 (92.3–95.8)	71.6 (66.3–76.5)	0.86 (0.84–0.89)	.404	4.08	44/1079
Combinations with hormonal contraceptive use and irregular period as additional layers									
Hormonal contraceptive use		McNemar's *P* value comparing sensitivities (reference = without layer)[a]					Sidak *P* value comparing AUCs (reference = without layer)[a]		
DFL composite: >10 wk, missed 2 periods, or >2 mo pregnant	93.0 (89.3–95.7)	.031	64.2 (60.8–67.5)	96.5 (94.5–97.9)	46.5 (42.2–50.8)	0.79 (0.76–0.81)	<.001	1.76	19/1081
Questionnaire composite: No. of weeks pregnant, date of LMP, date got pregnant	93.4 (89.7–96.0)	.001	72.6 (69.4–75.7)	97.0 (95.3–98.2)	53.4 (48.8–57.9)	0.83 (0.81–0.85)	.001	1.67	18/1079
Irregular periods									
DFL composite: >10 wk, missed 2 periods, or >2 mo pregnant	93.4 (89.7–96.0)	.016	64.4 (61.0–67.7)	96.7 (94.8–98.0)	46.8 (42.5–51.1)	0.79 (0.77–0.81)	<.001	1.67	18/1081
Questionnaire composite: No. of weeks pregnant, date of LMP, date got pregnant	93.4 (89.7–96.0)	.001	66.6 (63.2–69.8)	96.8 (94.9–98.1)	48.4 (44.0–52.8)	0.80 (0.78–0.82)	<.001	1.67	18/1079
Hormonal contraceptive use and irregular periods									
Questionnaire composite: No. of weeks pregnant, date of LMP, date got pregnant	95.9 (92.9–98.0)	.001	59.8 (56.3–63.2)	97.8 (96.1–98.9)	44.4 (40.4–48.6)	0.78 (0.76–0.80)	<.001	1.02	11/1079

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022. *(continued)*

TABLE 4
Screening performance of combinations of pregnancy dating items, overall and with inclusion of additional questions for hormonal contraceptive use and irregular periods and shifting of gestational duration thresholds *(continued)*

	Sensitivity (95% CI)	McNemar's *P* value comparing sensitivities	Specificity (95% CI)	Negative predictive value (95% CI)	Positive predictive value (95% CI)	AUC (95%CI)	Sidak *P* value comparing AUC values	Percent of false negative	No. of false negatives/ total N
Label composite: >10 wk, missed 2 periods, or >2 mo pregnant	95.2 (91.9–97.4)	<.001	57.9 (54.4–61.3)	97.3 (95.4–98.6)	43.1 (39.1–47.1)	0.77 (0.74–0.79)	<.001	1.20	13/1081
Lower dating threshold with 70-d GA on ultrasound (questionnaire only)									
		(reference=70 d)[a]					(reference=70 d)[a]		
Questionnaire composite: No. of weeks pregnant, date of LMP, date got pregnant as 63 d	95.2 (91.9–97.4)	<.001	67.3 (64.0–70.6)	97.7 (96.0–98.8)	49.4 (45.1–53.8)	0.81 (0.79–0.83)	<.001	1.20	13/1079
Questionnaire composite: No. of weeks pregnant, date of LMP, date got pregnant as 56 d	97.4 (94.8–99.0)	<.001	49.0 (45.5–52.5)	98.3 (96.5–99.3)	39.1 (35.4–42.8)	0.73 (0.71–0.75)	<.001	0.64	7/1079

*Scaffold" and "Composite" refer to how question responses were combined into 1 dichotomous variable (eligible vs ineligible). In scaffolding, responses were layered from highest to lowest-accuracy questions. In composite, an ineligible response to any question resulted in coding the participant as ineligible, regardless of their responses to other questions. For all combinations, respondents who selected "Not sure" or skipped all of the questions were coded as ineligible.

AUC, area under the receiver operating characteristic curve; *CI*, confidence interval; *DFL*, Drug Facts Label; *GA*, gestational age; *LMP*, last menstrual period; *ref*, reference group in significance testing.

[a] In significance testing for adding in hormonal contraceptive use and irregular periods as additional layers or lowering the dating threshold, the marked variables are the reference group.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

The average ultrasound-measured GD of participants with false negatives (n=25) to the DFL Module composite items of ">10 weeks," "missed >2 periods," or ">2 months pregnant" was 83.3 days and ranged from 72 to 117 days. Two-thirds (64%, 16/25) of participants had a GD <84 days (not shown).

The average ultrasound-measured GD of participants with false negatives (n=29) to the Questionnaire Module composite items of number of weeks pregnant, date of LMP, or date they got pregnant was 82.6 days and ranged from 71 to 126 days. Most (72%, 21/29) participants had a GD <84 days (not shown).

Sensitivity analysis

In subgroup analyses restricted only to participants who reported no previous ultrasound (n=867, 81%), the screening performance of individual items and the composite measures was largely comparable (Supplemental Table 3).

There were no statistically significant differences in AUC by module randomization order (Questionnaire vs DFL), with p-values ranging from 0.196 to 0.986 (Supplemental Table 4).

Comment

Principal findings and results in context

In this sample of people seeking abortion at facilities across the United States, multiple questions were effective in enabling patients to self-assess the duration of their pregnancy. Asking for a "Yes"/"No" response to a single statement—"You are >10 weeks pregnant"—correctly identified 84% of participants with GD >70 days and would result in 4% screening as eligible for MA when they were >10 weeks' gestation on ultrasound. Adding 3 additional statements for evaluating whether the participants were >2 months pregnant, missed >2 periods, or had irregular periods before pregnancy, correctly identified 93% of participants with GD of >70 days on ultrasound, and would result in 1.7% incorrectly screening as eligible. This represents a reduction in the proportion of participants incorrectly screened as eligible compared with previous studies, where

this proportion was 3.3% or higher when using date of LMP alone.[7,23–26]

Notably, these screening questions did not rely on date of LMP, the standard approach to pregnancy dating,[15] the focus of previous research,[6] and a measure that past studies have demonstrated as harder to recall for some, including young people and those with irregular periods.[13,27] Our study indicates that a set of 3 non-LMP-based questions can be used to self-assess pregnancy duration, particularly when the goal is minimizing the frequency of erroneous self-assessment of GD as <70 days (ie, when sensitivity is of paramount importance). These 3 questions could appear on a DFL for a future OTC MA product and enable very good self-assessment of GD.

Still, our study findings do not suggest that LMP or date-based questions should be abandoned. If access to facility-based abortion care is restricted, whether because of COVID-19[28] or restrictive policies,[29–32] screening questions may need to prioritize balancing sensitivity and specificity, or overall accuracy, rather than sensitivity alone. Our study revealed that a composite measure of responses to 3 questions on date of LMP, date they think they got pregnant, and number of weeks pregnant had an AUC value of 0.86, reflecting very good sensitivity (89%) and specificity (82%), with a 70-day threshold. Notably, performance was comparable with a 77-day threshold, with AUC of 0.87, 86% sensitivity, and 89% specificity. Thus, adding just 2 additional questions beyond date of LMP expands the number of people who can respond, and with very good accuracy. Therefore, these questions could be well-suited for an online tool for self-screening of eligibility for an OTC MA product, and could be effective for most people seeking abortion.

Clinical implications

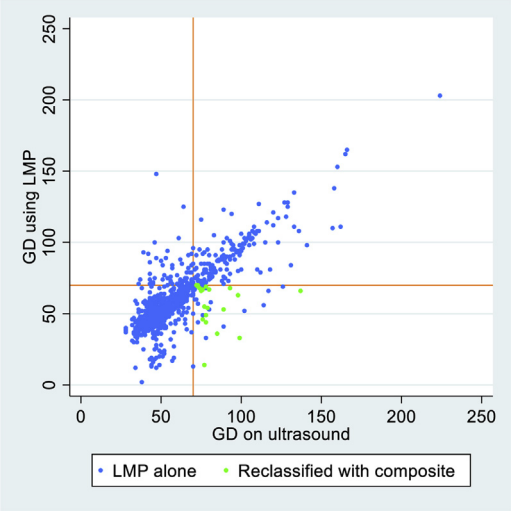
Like most screening questions, there was no 1 question or combination of questions that resulted in perfect eligibility classification. Thus, the decision to forgo ultrasound partly depends on the levels of risk tolerance of people in need of abortion care, providers, and policymakers regarding the use of MA at >70 or >77 days' gestation, taking into account that such use is rare. In this study, >70% of the 29 patients who erroneously screened as eligible but were actually >70 days on ultrasound had a GD of less than 84 days (12 weeks) gestation, where existing evidence indicates that MA efficacy in ending a pregnancy remains high and that people receive appropriate follow-up care when needed, including for ectopic pregnancy.[33] However, there will be rare cases where people screen as eligible and take the medications later in gestation and may require additional counseling or treatment. Levels of risk tolerance can depend on other factors such as availability and, importantly, accessibility of facility-based abortion care to all patients.[32] In the context of COVID-19, another factor may be the desire to minimize risk of exposure to the virus. Of paramount importance is that informed consent and counseling practices evolve to center people seeking abortion in the decision-making on proceeding without ultrasound.

Strengths and limitations

This study has several notable strengths. First, because we included nearly all people seeking abortion at recruitment facilities, the sample was racially and socioeconomically diverse and generalizable to people seeking facility-based abortion with ultrasound in the United States. Further, our sample included much more people seeking abortion around the 70-day threshold for MA than those in previous research. Given that one-quarter of the sample had a GD of >10 weeks on ultrasound, this study offers evidence of the real-world performance of screening questions among people in need of abortion. Importantly, this study also establishes the screening performance of questions beyond date of LMP, offering clinicians and researchers an expanded set of tools for assessing duration of pregnancy.

This study has several limitations. First, we developed many of the dating questions and examined their performance around a 70-day threshold given current FDA mifepristone labeling. Further evaluation would be necessary to ensure that some items, particularly in the DFL Module, perform similarly at higher GD thresholds. Notably, in sensitivity analyses, a composite measure of date of LMP, date they got pregnant,

FIGURE 2
Scatterplot of GD estimated using date of LMP vs GD on ultrasound



Highlighted participants (*green dots*) were reclassified appropriately as ineligible for MA with composite questionnaire combination of date they got pregnant, date of LMP, and number of weeks pregnant compared with using date of LMP alone (*blue dots*).

GD, gestational duration; *LMP*, last menstrual period; *MA*, medication abortion.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

and number of weeks pregnant performed similarly with a 77-day threshold, which is important because some providers have already started using this threshold.[1] Second, although the comparison of our study population data with Guttmacher's national survey data suggest that our study findings are generalizable to people seeking facility-based abortion nationally, additional research would be useful in ensuring that questions be appropriate for people with low literacy levels, which is prioritized by the FDA but was not assessed in this study. Similarly, <2% of participants chose to complete the study in Spanish, and thus we were unable to establish screening performance among Spanish speakers; a future study among Spanish speakers would ensure equitable access to expanded screening models.

Conclusions

This study provides robust evidence that people seeking abortion can self-assess their GD, particularly around the 70 or 77-day threshold. Importantly, beyond relying on date of LMP primarily, our study offers alternative pregnancy dating questions that produce more accurate self-assessments and lower nonresponse. Our findings suggest that policies requiring in-person ultrasound or dispensing of medications such as the Risk Evaluation and Mitigation Strategy for mifepristone are not universally necessary to establish gestation-based eligibility for MA. ■

Acknowledgments

The authors express deep gratitude to members of this study's Community Advisory Board for their significant input in improving study design and interpretation. Members include Claudie Kiti Bustamante, Sung Yeon Choimorrow, Debra Hauser, Yamani Hernandez, Tammi Kromenaker, and Ena Valladares. We are also grateful to Isabel Munoz, MPH, Elizabeth Gonzalez, BA, and Miriam Parra, BA for support in reviewing and translating study materials and to Tanvi Gurazada, MS, and Jessica Navarro, MPH for study support activities.

References

1. Raymond EG, Grossman D, Mark A, et al. Commentary: no-test medication abortion: a sample protocol for increasing access during a pandemic and beyond. Contraception 2020;101:361–6.
2. Chong E, Shochet T, Raymond E, et al. Expansion of a direct-to-patient telemedicine abortion service in the United States and experience during the COVID-19 pandemic. Contraception 2021;104:43–8.
3. Upadhyay UD, Koenig LR, Meckstroth KR. Safety and efficacy of telehealth medication abortions in the US during the COVID-19 pandemic. JAMA Netw Open 2021;4:e2122320.
4. Ramaswamy A, Weigel G, Sobel L, Salganicoff A. Medication abortion and telemedicine: innovations and barriers during the COVID-19 emergency. 2020. Available at: https://www.kff.org/policy-watch/medication-abortion-telemedicine-innovations-and-barriers-during-the-covid-19-emergency/. Accessed April 26, 2021.
5. Mifeprex REMS Study Group, Raymond EG, Blanchard K, et al. Sixteen years of overregulation: time to unburden Mifeprex. N Engl J Med 2017;376:790–4.
6. Schonberg D, Wang LF, Bennett AH, Gold M, Jackson E. The accuracy of using last menstrual period to determine gestational age for first trimester medication abortion: a systematic review. Contraception 2014;90:480–7.
7. Bracken H, Clark W, Lichtenberg ES, et al. Alternatives to routine ultrasound for eligibility assessment prior to early termination of pregnancy with mifepristone-misoprostol. BJOG 2011;118:17–23.
8. Raymond EG, Bracken H. Early medical abortion without prior ultrasound. Contraception 2015;92:212–4.
9. Janiak E, Kawachi I, Goldberg A, Gottlieb B. Abortion barriers and perceptions of gestational age among women seeking abortion care in the latter half of the second trimester. Contraception 2014;89:322–7.
10. Norman WV, Bergunder J, Eccles L. Accuracy of gestational age estimated by menstrual dating in women seeking abortion beyond nine weeks. J Obstet Gynaecol Can 2011;33:252–7.
11. Drey EA, Foster DG, Jackson RA, Lee SJ, Cardenas LH, Darney PD. Risk factors associated with presenting for abortion in the second trimester. Obstet Gynecol 2006;107:128–35.
12. McCarthy M, Upadhyay U, Biggs MA, Anthony R, Holl J, Roberts SC. Predictors of timing of pregnancy discovery. Contraception 2018;97:303–8.
13. Swanson M, Karasek D, Drey E, Foster DG. Delayed pregnancy testing and second-trimester abortion: can public health interventions assist with earlier detection of unintended pregnancy? Contraception 2014;89:400–6.
14. Foster DG, Jackson RA, Cosby K, Weitz TA, Darney PD, Drey EA. Predictors of delay in each step leading to an abortion. Contraception 2008;77:289–93.
15. Committee Opinion No 700: methods for estimating the due date. Obstet Gynecol 2017;129:e150–4.
16. United States Census Bureau. Poverty thresholds. 2021. Available at: https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html. Accessed February 25, 2021.
17. Guttmacher Institute. 2014 Abortion Patient Survey (APS). 2020. Available at: https://www.guttmacher.org/population-center/dataset/2014-abortion-patient-survey-aps. Accessed March 3, 2021.
18. Jones RK, Witwer E, Jerman J. Abortion incidence and service availability in the United States 2017. 2019. Available at: https://www.guttmacher.org/report/abortion-incidence-service-availability-us-2017. Accessed April 1, 2020.
19. US Food and Drug Administration. Self-selection studies for nonprescription drug products. 2020. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/self-selection-studies-nonprescription-drug-products. Accessed February 25, 2021.
20. Šimundić AM. Measures of diagnostic accuracy: basic definitions. EJIFCC 2009;19:203–11.
21. Cleves MA. From the help desk: comparing areas under receiver operating characteristic curves from two or more probit or logit models. The Stata Journal 2002;2:301–13.
22. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6:e012799.
23. McGalliard C, Gaudoin M. Routine ultrasound for pregnancy termination requests increases women's choice and reduces inappropriate treatments. BJOG 2004;111:79–82.
24. Saavedra-Avendano B, Schiavon R, Sanhueza P, Rios-Polanco R, Garcia-Martinez L, Darney BG. Early termination of pregnancy: differences in gestational age estimation using last menstrual period and ultrasound in Mexico. Reprod Health 2020;17:89.
25. Blanchard K, Cooper D, Dickson K, et al. A comparison of women's, providers' and ultrasound assessments of pregnancy duration among termination of pregnancy clients in South Africa. BJOG 2007;114:569–75.
26. Constant D, Harries J, Moodley J, Myer L. Accuracy of gestational age estimation from last menstrual period among women seeking abortion in South Africa, with a view to task sharing: a mixed methods study. Reprod Health 2017;14:100.
27. Finer LB, Frohwirth LF, Dauphinee LA, Singh S, Moore AM. Timing of steps and reasons for delays in obtaining abortions in the United States. Contraception 2006;74:334–44.
28. White K, Kumar B, Goyal V, Wallace R, Roberts SCM, Grossman D. Changes in abortion in Texas following an executive order ban during the coronavirus pandemic. JAMA 2021;325:691–3.
29. Gerdts C, Fuentes L, Grossman D, et al. Impact of clinic closures on women obtaining abortion services after implementation of a

restrictive law in Texas. Am J Public Health 2016;106:857–64.
30. Myers C, Ladd D. Did parental involvement laws grow teeth? The effects of state restrictions on minors' access to abortion. J Health Econ 2020;71:102302.
31. Grossman D, Baum S, Fuentes L, et al. Change in abortion services after implementation of a restrictive law in Texas. Contraception 2014;90:496–501.
32. Perritt J, Grossman D. The health consequences of restrictive abortion laws. JAMA Intern Med 2021;181:713–4.
33. Raymond EG, Tan YL, Comendant R, et al. Simplified medical abortion screening: a demonstration project. Contraception 2018;97:292–6.

Author and article information

From the Advancing New Standards in Reproductive Health research group, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, Oakland, CA (Dr Ralph, Mses Ehrenreich and Barar, Dr Biggs, Ms Morris, and Dr Grossman); Ibis Reproductive Health, Cambridge, MA (Ms Blanchard); Ipas, Chapel Hill, NC (Dr Kapp); Pegasus Health Justice Center, Dallas, TX (Dr Moayedi); Physicians for Reproductive Health, New York, NY (Dr Perritt); Gynuity Health Projects, Tacoma, WA (Dr Raymond); and Steve Hicks School of Social Work and Department of Sociology, The University of Texas at Austin, Austin, TX (Dr White).

Received July 30, 2021; revised Nov. 15, 2021; accepted Nov. 24, 2021.

The authors report no conflict of interest.

This research was supported by a grant from the Society of Family Planning Research Fund (SFPRF12-MA9). At the time of the study, L.J.R. was supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Office of Research on Women's Health, Building Interdisciplinary Research Careers in Women's Health grant (2K12 HD052163). The funders had no role in the study design; data collection, analysis, and interpretation; or decision to submit the work for publication. The views and opinions expressed are those of the authors and do not necessarily represent the views and opinions of the funders.

Corresponding author: Lauren J. Ralph, PhD, MPH. Lauren.Ralph@ucsf.edu

Supplemental Materials

Accuracy of self-assessment of gestational duration among people seeking abortion.

Technical details

Formulas for screening indicators

The performance of individual questions and combinations of questions was assessed using screening measures derived from the following 2×2 table. All measures were estimated using Stata's diagti command.

SUPPLEMENTAL FIGURE 1
2×2 table

		Ultrasound result	
		Ineligible (typically > 70 days or ≥ 10 weeks)	Eligible (typically ≤ 70 days or < 10 weeks)
Screening question(s)	+ Self-assess as ineligible for MA (typically > 70 days or ≥ 10 weeks gestation)	A True Positive	B False positive
	- Self-assess as eligible for MA (typically ≤ 70 days or < 10 weeks gestation)	C False negative	D True Negative

MA, medication abortion.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

Sensitivity=A / [A+C]

Specificity=D / [B+D]

Positive Predictive Value (PPV)=A / [A+B]

Negative Predictive Value (NPV)=D / [C+D]

Proportion of false negatives=C / [A+B+C+D]

Area Under the Receiver Operating Characteristic (ROC) Curve (AUC)=(Sensitivity + Specificity)/2[1]

The US Food and Drug Administration guidance on self-selection considers accuracy, sensitivity, and NPV to be most important when determining whether someone can self-select to take a medication, for example in an over-the-counter environment.[2] Sensitivity reflects correct self-selection of those who cannot take the medication (true positive rate), whereas NPV captures correct self-selection of those who select to take the medication.

AUC is a summary measure of overall accuracy reflecting both sensitivity and specificity; AUC values of 1.0 represent a test that perfectly discriminates people eligible vs ineligible to take a medication, whereas an AUC of 0.5 reflects a test with no discriminatory ability.[3,4]

Additional details on study methods

Rules for recoding gestational duration variables calculated on the basis of dates

To calculate GD using the date of LMP, we subtracted the date of the survey from LMP date. This resulted in some nonsensical values, defined as ≤14 days.

Values of 12 to 14 days (n=7) were left as is and gestational duration calculated accordingly; values between −19 and 12 were recoded to missing (n=10). Large, negative numbers (n=8) were reviewed in detail and determined to be the result of misreporting of year in the calendar; the programming of the survey defaulted to current calendar year and our survey period spanned 2019 to 2020. For these 8 cases, we replaced year with 2019, generating a plausible value of pregnancy duration ≤280 days.

A similar recoding approach was used for GD calculated using the date they got pregnant, defined as the number of days between the date they got pregnant and the survey date plus 14. Nonsensical values were defined as ≤28 days. Values of 20 to 28 (n=8) were left as is; values between −8 and 20 (n=8) were recoded to missing; large and negative values that likely resulted from mis-entry of year (n=14) were recoded using calendar year 2019.

Details on exploratory factor analysis

To reduce the number of items considered for combination and to confirm that the items were measuring only 1 construct, we ran 2 separate exploratory factor analyses using the iterated principal-factor method. The first model contained 5 candidate items from the Drug Facts Label module, whereas the second model contained 6 items from the Questionnaire module. Factor 1 eigenvalues for the Label and Questionnaire items were 2.75 and 3.15, respectively, each followed by a precipitous drop in eigenvalues (0.27 and 0.13, respectively), suggesting that they measure only 1 construct. We removed items with low factor loading (<0.60), preserving a 1-factor solution. The retained items that were then considered further in combination had factor loadings ranging from 0.63 to 0.75 for the Drug Facts Label (n=4) and 0.70 to 0.85 for the Questionnaire items (n=4).

Supplemental References

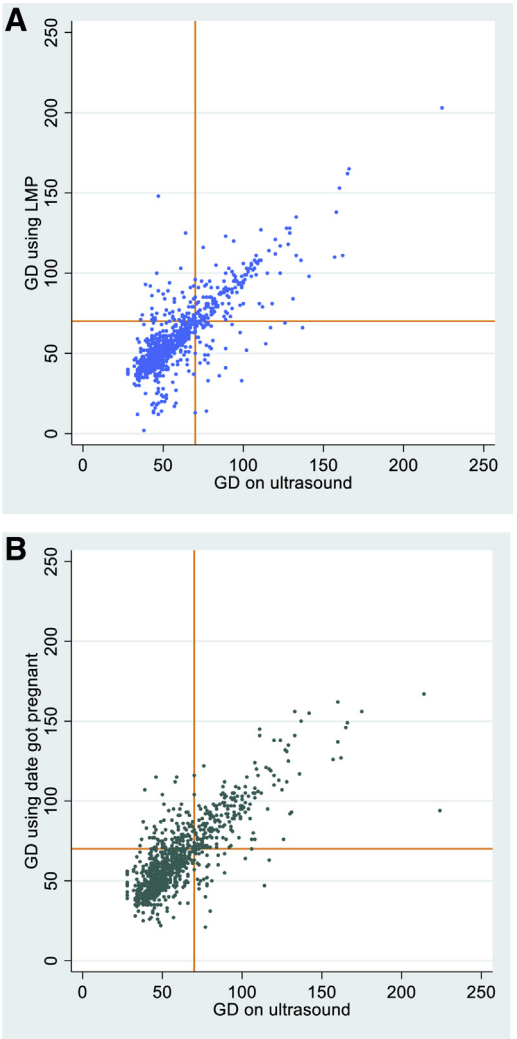
1. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–84.

2. US Food and Drug Administration. Center for Drug Evaluation and Research. Self-Selection Studies for Nonprescription Drug Products. 2020. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/selfselection-studies-nonprescription-drug-products. Accessed February 25, 2021.

3. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration. 2010. Available from, http://srdta.cochrane.org/. Accessed April 20, 2020.

4. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford, England: Oxford University Press; 2003.

SUPPLEMENTAL FIGURE 2
Scatterplot



Scatterplot of GD estimated by participant using (**A**) date of LMP and on ultrasound; (**B**) date they got pregnant and on ultrasound.

GD, gestational duration; *LMP*, last menstrual period.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

SUPPLEMENTAL TABLE 1

Comparison of the sociodemographic and pregnancy characteristics of the study sample with Guttmacher surveys of abortion patients and providers

	Gestational duration dating study (n=1089)		2014 Guttmacher Abortion Patient Survey (unweighted N=8380)
	No.	Percent	Percent
Age (y)			
<15	0		0.2
15–17	26	2.4	3.4
18–19	84	7.7	8.2
20–24	331	30.4	33.6
25–29	305	28.0	26.5
30–34	187	17.2	15.9
35–39	73	6.7	9.1
≥40	34	3.1	3.1
Missing	49	4.5	n/a
Race/ethnicity			
Non-Hispanic White	337	31.0	38.7
Non-Hispanic Black	385	35.4	27.6
Hispanic/Latinx	196	18.0	24.8
Non-Hispanic Asian or Pacific Islander	50	4.6	5.5
Other/Multiple	81	7.4	3.4
Missing	40	3.7	n/a
Nativity			
US-born	913	83.8	83.9
Born outside the United States	136	12.5	16.1
Missing	40	3.7	n/a
Highest level of education completed			
<high school	74	6.8	8.9
High school diploma or GED	314	28.8	27.0
Some college/associate's degree	436	40.0	40.9
College graduate	225	20.7	23.1
Missing	40	3.7	n/a
No. of previous births			
0	459	42.2	40.7
1	239	22.0	26.2
≥2	354	32.5	33.1
Missing	37	3.4	n/a
Household income as % of federal poverty level			
<100	375	34.4	49.3
100–199	211	19.4	25.7
≥200	217	19.9	25.0
Missing	286	26.3	n/a

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022. *(continued)*

SUPPLEMENTAL TABLE 1

Comparison of the sociodemographic and pregnancy characteristics of the study sample with Guttmacher surveys of abortion patients and providers *(continued)*

	Gestational duration dating study (n=1089)		2014 Guttmacher Abortion Patient Survey (unweighted N=8380)
	No.	Percent	Percent
Gestational duration on ultrasound (wk)			2017 Guttmacher Abortion Provider Survey (2017)
≤8	705	64.7	65.4
9–10	166	15.2	14.7
11–12	87	8.0	8.2
13–15	78	7.2	6.3
16–20	39	3.6	4.1
≥21	14	1.3	1.3

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

SUPPLEMENTAL TABLE 2
Screening performance of select pregnancy dating questions at 77-day threshold

Questions	N	Sensitivity (95% CI)	Specificity (95% CI)	Negative predictive value (95% CI)	Positive predictive value (95% CI)	AUC (95% CI)	Proportion (%) of false negatives	No. of false negatives/ total N
Date of LMP alone[a]	1041	83.7 (77.7–88.6)	82.5 (79.8–85.0)	95.6 (93.9–97.0)	52.6 (46.9–58.2)	0.83 (0.80–0.86)	3.07	32/1041
Questionnaire composite: No. of weeks pregnant, date of LMP, or date they got pregnant	1079	86.0 (80.5–90.4)	88.0 (85.6–90.0)	96.4 (94.8–97.5)	62.9 (57.0–68.5)	0.87 (0.84–0.90)	2.66	29/1079
Lower dating threshold								
No. of weeks pregnant, date of LMP, or date they got pregnant screened as ≤70 d	1079	93.2 (88.9–96.3)	77.6 (74.4–80.4)	98.0 (96.6–98.9)	49.7 (44.7–54.8)	0.85 (0.83–0.88)	1.30	14/1079
No. of weeks pregnant, date of LMP, or date they got pregnant screened as ≤63 d	1079	96.6 (93.2–98.6)	63.1 (59.8–66.3)	98.7 (97.4–99.5)	38.3 (34.1–42.6)	0.80 (0.78–0.82)	0.65	7/1079
No. of weeks pregnant, date of LMP, or date they got pregnant screened as ≤56 d	1079	98.1 (95.1–99.5)	45.8 (42.4–49.1)	99.0 (97.5–99.7)	30.0 (26.6–33.6)	0.72 (0.70–0.74)	0.37	4/1079

AUC, area under the receiver operating characteristic curve; *CI*, confidence interval; *LMP*, last menstrual period.

[a] Respondents who selected "Not sure" or skipped the question were coded as ineligible.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

SUPPLEMENTAL TABLE 3

Screening performance of individual and combination pregnancy dating questions among those with no previous ultrasound (n = 867)

	No.	Sensitivity: % ineligible who screened as ineligible) (95% CI)	Specificity: % eligible who screened as eligible) (95% CI)	Negative predictive value (95% CI)	Positive predictive value (95% CI)	AUC (95% CI)	Proportion (%) of false negatives
Module 1/DFL items							
Missed >2 periods since got pregnant	862	80.4 (73.5–86.1)	75.4 (72.0–78.5)	94.1 (91.8–95.9)	44.1 (38.5–49.9)	0.78 (0.74–0.81)	3.83
Positive pregnancy test >7 wk ago	863	66.1 (58.4–73.2)	69.5 (65.9–72.9)	89.4 (86.5–91.9)	34.4 (29.2–39.8)	0.68 (0.64–0.72)	6.66
Last normal period started >10 wk ago	862	77.4 (70.3–83.5)	67.3 (63.7–70.8)	92.5 (89.8–94.6)	36.4 (31.4–41.6)	0.72 (0.69–0.76)	4.41
>10 wk pregnant	865	79.9 (73.0–85.6)	77.7 (74.5–80.8)	94.1 (91.8–95.9)	46.6 (40.7–52.5)	0.79 (0.75–0.82)	3.93
>2 mo pregnant	865	81.7 (75.0–87.2)	65.4 (61.7–68.9)	93.6 (91.1–95.6)	36.4 (31.6–41.5)	0.74 (0.70–0.77)	3.58
Composite: >10 wk, >2 mo, or missed > 2 periods	864	85.7 (79.5–90.6)	73.9 (70.4–77.1)	95.5 (93.4–97.1)	44.2 (38.7–49.7)	0.80 (0.77–0.83)	2.77
Module 2/Questionnaire Items							
Number of weeks pregnant	866	74.1 (66.9–80.5)	91.4 (89.0–93.4)	93.5 (91.4–95.3)	67.7 (60.5–74.4)	0.83 (0.79–0.86)	5.08
Got pregnant within past 2 mo	867	41.8 (34.3–49.6)	85.9 (83.1–88.4)	85.8 (83.0–88.3)	42.0 (34.5–49.8)	0.64 (0.60–0.68)	11.4
Number of weeks since first day of LMP	843	77.0 (69.8–83.2)	80.2 (77.0–83.2)	93.5 (91.1–95.3)	48.7 (42.4–54.9)	0.79 (0.75–0.82)	4.51
Positive pregnancy test >7 wk ago	822	57.1 (48.9–64.9)	82.9 (79.8–85.7)	89.2 (86.5–91.5)	43.8 (36.9–51.0)	0.70 (0.66–0.74)	8.15
Date of LMP	843	79.4 (72.4–85.3)	81.4 (78.3–84.3)	94.2 (92.0–95.9)	51.0 (44.7–57.2)	0.80 (0.77–0.84)	4.03
Reported date they got pregnant	867	84.7 (78.4–89.8)	74.7 (71.4–77.9)	95.2 (93.1–96.9)	45.0 (39.5–50.6)	0.80 (0.77–0.83)	3.00
Composite: Number of weeks pregnant, date of LMP, or date they got pregnant	858	85.3 (79.1–90.3)	83.5 (80.5–86.2)	95.9 (94.0–97.3)	55.8 (49.5–61.9)	0.84 (0.81–0.87)	2.88

For all questions, respondents who selected "Not sure" or skipped the question were coded as ineligible for calculating screening performance measures. Individuals who had already had an ultrasound at the recruitment facility before recruitment were ineligible to participate in the study. Individuals who had an ultrasound at another facility remained eligible to participate. Therefore, in this sensitivity analysis, we removed participants who reported having an ultrasound at another facility before seeking abortion care at the recruitment facility.

AUC, area under the receiver operating characteristic curve; *CI*, confidence interval; *DFL*, Drug Facts Label; *LMP*, last menstrual period.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.

SUPPLEMENTAL TABLE 4

Sensitivity analysis comparing area under the receiver operating characteristic curve values (with 95% confidence intervals) of individual items and combination measures by order in which participants saw the module

	Randomization order		*P* value[a]
Module 1/Drug Facts Label Items	First	Second	
Missed >2 periods since got pregnant	0.80 (0.76–0.84)	0.78 (0.75–0.82)	.411
>10 wk pregnant	0.81 (0.78–0.85)	0.83 (0.79–0.86)	.202
>2 mo pregnant	0.74 (0.70–0.78)	0.77 (0.73–0.80)	.555
Composite: "Yes" to missed >2 periods, >10 wk, or >2 mo pregnant	0.81 (0.78–0.85)	0.82 (0.79–0.85)	.870
Module 2/Questionnaire Items			
Number of weeks pregnant	0.84 (0.80–0.87)	0.80 (0.76–0.84)	.212
Date of LMP	0.85 (0.82–0.89)	0.88 (0.85–0.91)	.196
Reported date they got pregnant	0.81 (0.77–0.84)	0.82 (0.79–0.85)	.871
Composite: number of weeks pregnant, date of LMP, date they got pregnant	0.85 (0.82–0.89)	0.86 (0.83–0.89)	.986

AUC, area under the receiver operating characteristic curve; *LMP*, last menstrual period.

[a] *P* value from X^2 test comparing AUC values.

Ralph et al. Abortion patients' self-assessment of gestational duration. Am J Obstet Gynecol 2022.