



Defining safe criteria to diagnose miscarriage: prospective observational multicentre study

¹ Queen Charlotte's and Chelsea Hospital, Imperial College, London W12 0HS, UK

² Early Pregnancy and Acute Gynaecology Unit, St Thomas' Hospital, London, UK

³ St Mary's Hospital, Imperial College NHS Trust, London, UK

⁴ Chelsea and Westminster Hospital, London, UK

⁵ Northwick Park Hospital, Harrow, Middlesex, UK

⁶ Department of Obstetrics and Gynaecology, University Hospitals Leuven, Leuven, Belgium

⁷ Princess Anne Hospital, Southampton, Hampshire UK

⁸ Department of Development and Regeneration, KU Leuven, Belgium

Correspondence to: T Bourne
tbourne@imperial.ac.uk

Cite this as: *BMJ* 2015;351:h4579
<http://dx.doi.org/10.1136/bmj.h4579>

Accepted: 11 August 2015

Published: 23 September 2015

Jessica Preisler,¹ Julia Kopeika,² Laura Ismail,^{1,3} Veluppillai Vathanan,⁴ Jessica Farren,¹ Yazan Abdallah,¹ Parijat Battacharjee,⁵ Caroline Van Holsbeke,⁶ Cecilia Bottomley,⁴ Deborah Gould,³ Susanne Johnson,⁷ Catriona Stalder,¹ Ben Van Calster,⁸ Judith Hamilton,² Dirk Timmerman,^{6,8} Tom Bourne^{1,6,8}

Objectives To validate recent guidance changes by establishing the performance of cut-off values for embryo crown-rump length and mean gestational sac diameter to diagnose miscarriage with high levels of certainty. Secondary aims were to examine the influence of gestational age on interpretation of mean gestational sac diameter and crown-rump length values, determine the optimal intervals between scans and findings on repeat scans that definitively diagnose pregnancy failure.)

Design Prospective multicentre observational trial. Setting Seven hospital based early pregnancy assessment units in the United Kingdom.

Participants 2845 women with intrauterine pregnancies of unknown viability included if transvaginal ultrasonography showed an intrauterine pregnancy of uncertain viability. In three hospitals this was initially defined as an empty gestational sac <20 mm mean diameter with or without a visible yolk sac but no embryo, or an embryo with crown-rump length <6 mm with no heartbeat. Following amended guidance in December 2011 this definition changed to a gestational sac size <25 mm or embryo crown-rump length <7 mm. At one unit the definition was extended throughout to include a mean gestational sac diameter <30 mm or embryo crown-rump length <8 mm.

Main outcome measures Mean gestational sac diameter, crown-rump length, and presence or absence of embryo heart activity at initial and repeat transvaginal ultrasonography around 7-14 days later. The final outcome was pregnancy viability at 11-14 weeks' gestation.

Results The following indicated a miscarriage at initial scan: mean gestational sac diameter ≥ 25 mm with an empty sac (364/364 specificity: 100%, 95% confidence interval 99.0% to 100%), embryo with crown-rump length ≥ 7 mm without visible embryo heart activity (110/110 specificity: 100%, 96.7% to 100%), mean gestational sac diameter ≥ 18 mm for gestational sacs without an embryo presenting after 70 days' gestation (907/907 specificity: 100%, 99.6% to 100%), embryo with crown-rump length ≥ 3 mm without visible heart activity presenting after 70 days' gestation (87/87 specificity: 100%, 95.8% to 100%). The following were indicative of miscarriage at a repeat scan: initial scan and repeat scan after seven days or more showing an embryo without visible heart activity (103/103 specificity: 100%, 96.5% to 100%), pregnancies without an embryo and mean gestational sac diameter <12 mm where the mean diameter has not doubled after 14 days or more (478/478 specificity: 100%, 99.2% to 100%), pregnancies without an embryo and mean gestational sac diameter ≥ 12 mm showing no embryo heartbeat after

seven days or more (150/150 specificity: 100%, 97.6% to 100%).

Conclusions Recently changed cut-off values of gestational sac and embryo size defining miscarriage are appropriate and not too conservative but do not take into account gestational age. Guidance on timing between scans and expected findings on repeat scans are still too liberal. Protocols for miscarriage diagnosis should be reviewed to account for this evidence to avoid misdiagnosis and the risk of terminating viable pregnancies.

Introduction

Bleeding or pain is the most common reason women seek medical advice in early pregnancy. It is axiomatic that diagnostic criteria for miscarriage should be watertight, as one decision women might expect clinicians to have complete certainty about is the viability of their pregnancy. Until relatively recently the American College of Radiology guidance to define miscarriage used a mean gestational sac diameter of ≥ 16 mm for an empty gestational sac or visualisation of an embryo with a crown-rump length of ≥ 5 mm and no heartbeat.¹ In contrast, the Royal College of Obstetricians and Gynaecologists used a mean gestational sac diameter of ≥ 20 mm or an embryo with a crown-rump length of ≥ 6 mm and no heartbeat.² For measurements below these cut-off values a repeat ultrasound scan after seven or more days was recommended.

v110.1136/bmj.h4579v1h5137Video abstract

In October 2011 a series of papers provided evidence that these guidelines were unsafe. A systematic review concluded they were based on inadequate information.³ One study⁴ found significant interobserver and intra-observer variability for measurements of mean gestational sac diameter and crown-rump length. A prospective multicentre observational study showing the false positive rate for miscarriage using cut-off values for mean gestational sac diameter and crown-rump length was unacceptably high (false positive rate for mean sac diameter ≥ 16 mm: 4.4%, 95% confidence interval 8.4% to 2.2%, and for embryo crown-rump length ≥ 5 mm and no heartbeat: 8.3%, 25.8% to 2.3%).^{5,6} However, these results had wide confidence intervals, gestational age was not considered, and the study did not address the ultrasound findings expected and optimal timing of repeat ultrasound scans to avoid misdiagnosis.

Following these publications, the Royal College of Obstetricians and Gynaecologists immediately changed its guidance to one based on an empty

gestational sac of mean gestational sac diameter ≥ 25 mm or embryo with a crown-rump length ≥ 7 mm and no heartbeat.⁷ In December 2012 the UK National Institute for Health and Care Excellence adopted the recommendations of the Royal College of Obstetricians and Gynaecologists.⁸ The American College of Radiology guidance then also changed in 2013 and adopted the same criteria as the Royal College of Obstetricians and Gynaecologists to define miscarriage.⁹ As part of its guideline, NICE stated that for empty gestational sacs and embryos below the cut-off values, if the sac was empty or no embryo heartbeat was visible on an ultrasound scan after 7-10 days, these were definitive criteria for miscarriage. NICE reported no consensus on timing and frequency of repeat ultrasound scans for initially inconclusive scans and identified this as a research priority.⁸ Later in 2013 a review in the *New England Journal of Medicine* based on a consensus meeting of the US Society of Radiologists in Ultrasound highlighted the risks of incorrectly diagnosing early pregnancy failure and recommended criteria to diagnose miscarriage.¹⁰ This narrative review stated that the absence of an embryo with a heartbeat ≥ 14 days after a scan showing an empty gestational sac or absence of an embryo heartbeat ≥ 11 days after a scan showing a gestational sac and yolk sac were both categorically a miscarriage. These further recommendations were based on expert opinion but were not evidence based.

We validated and updated recent guidance changes by establishing the performance of cut-off values for crown-rump length and mean gestational sac diameter to diagnose miscarriage with narrow confidence intervals for specificity. We also examined the influence of gestational age on interpretation of values for mean gestational sac diameter and crown-rump length, to determine the optimal intervals between ultrasound scans and to identify findings expected on repeat ultrasound examinations that definitively diagnose pregnancy failure.

Methods

This was a prospective multicentre observational study. We recruited women from early pregnancy assessment units in seven hospitals: four university hospitals in London (St George's, Queen Charlottes and Chelsea, St Thomas', and St Mary's), one university hospital outside London (Princess Anne, Southampton), and two London general hospitals with university affiliations (Chelsea and Westminster and Northwick Park). Data were collected in two waves. Firstly, women were recruited consecutively between September 2010 and March 2011 at St George's, Queen Charlottes and Chelsea, and Chelsea and Westminster. Results from these data were published previously.^{5,6} In the second wave, women were recruited consecutively between August 2011 and May 2013 at all hospitals except St George's. Experienced nurse practitioners, ultrasonographers, and doctors with an interest in the use of ultrasound in early pregnancy carried out the scans in the study.

Women attended because of vaginal bleeding or pain, or both, hyperemesis, and for reassurance after a previous miscarriage or ectopic pregnancy. They were included if transvaginal ultrasonography revealed a singleton intrauterine pregnancy of uncertain viability. In three hospitals this was initially defined according to Royal College of Obstetricians and Gynaecologists guidance at that time,² as an empty gestational sac < 20 mm mean diameter with or without a visible yolk sac but no embryo, or an embryo with a crown-rump length < 6 mm with no heartbeat. Following amended guidance from the Royal College of Obstetricians and Gynaecologists in December 2011¹⁰ this definition changed to a gestational sac size < 25 mm or an embryo with a crown-rump length < 7 mm. At Queen Charlottes and Chelsea hospital the definition was extended throughout to include mean gestational

sac diameters < 30 mm or embryo crown-rump lengths < 8 mm. To establish immediate viability, the women were asked to return for ultrasonography 7-14 days later. For intervals between scans we excluded women who returned for a second scan less than two days or more than 30 days later. In the event that a pregnancy was diagnosed as a miscarriage, management was according to the existing protocols in the individual units, which was either surgery, medical, or expectant. In general a second operator checked a diagnosis of miscarriage, and after guidance changes in December 2012⁸ this practice was introduced formally in all units.

For clarity the immediate outcome was the outcome that was used to establish if a pregnancy was viable and to clinically manage the patients. The final outcome and reference standard for the study was viability at 11-14 weeks to establish whether pregnancies of uncertain viability that were shown to be viable on repeat scans subsequently went on to fail in any event. We excluded women who were clinically unstable, required intervention, or underwent termination of pregnancy. Further follow-up visits were scheduled according to the clinical situation.

Demographic variables recorded included the indication for scanning, maternal age, ethnicity, and gestational age at first scan (calculated from last menstrual period or embryo transfer date after infertility treatment). To calculate gestational age we asked the women if they were certain of the date of the first day of their last menstrual period. We set a minimum plausible gestational age of 14 days and maximum of 84 days. Gestational ages outside this range were treated as missing.

We measured gestational sac diameters from the inner borders in three orthogonal planes and calculated the mean diameter (see supplementary figure). The presence of a yolk sac was documented. Embryo crown-rump length was recorded, and the presence or absence of a heartbeat. We measured the crown-rump length to the nearest millimetre, placing calipers at the outer side of the embryo's crown and rump (greatest length).

All women underwent transvaginal ultrasonography using Voluson E8 (GE Medical System, Zipf, Austria), Accuvix XG (Samsung Medison, Korea), or SSD 5000 (Aloka, Japan) machines. All machines used in the study to make a final diagnosis were less than five years old and subject to standard regular maintenance.

The study was registered as a clinical audit of the performance of cut-off values to diagnose miscarriage (audit registration number 1149); the project was considered by the national research ethics service in London and they advised that ethical review by a National Health Service research ethics committee was not required.

Patient involvement

There was no patient involvement in this study.

Statistical analysis

The data were used to validate existing guidelines and to suggest new or updated guidelines. Because we aimed to define criteria that definitively diagnose pregnancy failure rather than to determine diagnostic accuracy, the main statistics for this study were sensitivity, specificity, and positive predictive values. We defined sensitivity as the percentage of non-viable pregnancies that met the cut-off or criterion, specificity as the percentage of viable pregnancies that did not meet the cut-off or criterion, and positive predictive value as the percentage of non-viable pregnancies among all pregnancies that met the cut-off or criterion. We used Jeffreys method to compute 95% confidence intervals.¹¹ To suggest guidelines that depend on gestational age or that focus on the

appropriate interval between scans to exclude viability, we mainly relied on scatter plots. Four authors (JP, TB, DT, and BVC) carried out expert interpretation of the results to decide on all new or updated guidelines.

For missing data we did not replace missing values but performed each analysis with available cases. After internal investigation we strongly believe that missing data on gestational age, mean gestational sac diameter, crown-rump length, second scan, and interval between scans did not differ systematically from women with complete data, which would imply that the omission of women with missing data would not bias results. The only exceptions were cases lost to follow-up (see supplementary file for details of these pregnancies).

Results

Overall, 3192 pregnancies were enrolled in the study. The 11-14 week outcome was unknown in 337 (10.6%) because of loss to follow-up,

pregnancy termination, or the 11-14 week scan was scheduled after data collection closed. Of the remaining 2855 pregnancies, the category of intrauterine pregnancy of uncertain viability could not be determined in 10 women owing to incomplete data, leaving 2845 pregnancies (89.1%) for the evaluation of first scan data (table 1). Information on repeat scans was lacking in 94 pregnancies (mainly due to a logistical problem at St George's hospital), and repeat scan classification (miscarriage, empty sac, gestational sac+yolk sac, embryo without heartbeat, embryo with heartbeat) was unclear in two, leaving 2749 pregnancies (86.1%) with information on 11-14 week viability, pregnancy type, and repeat scan classification. Finally, intervals between scans were unknown in 125 pregnancies (117 were women at Queen Charlottes and Chelsea hospital who had miscarried by follow-up scan and hence would not be used for repeat scan analysis anyway) and 66 had an interval outside the 1-30 day range, leaving 2558 pregnancies (80.1%) to evaluate the interval between scans (fig 1).

Table 1 | Data summary by centre and diagnostic group at initial scan. Values are numbers (percentages) unless stated otherwise

Hospitals	No of women	Empty gestational sac	Gestational sacs with a yolk sac	Embryo present	Viable at 11-14 weeks
Queen Charlottes	791	243 (31)	320 (40)	228 (29)	285 (36)
St Thomas's	695	181 (26)	365 (53)	149 (21)	326 (47)
St George's	457	265 (58)	156 (34)	36 (8)	223 (49)
St Mary's	366	137 (37)	146 (40)	83 (23)	174 (48)
Chelsea and Westminster	283	67 (24)	132 (47)	84 (30)	138 (49)
Northwick Park	135	55 (41)	54 (40)	26 (19)	64 (47)
Princess Anne	118	16 (14)	59 (50)	43 (36)	55 (47)
All data	2845	964 (34)	1232 (43)	649 (23)	1265 (44)

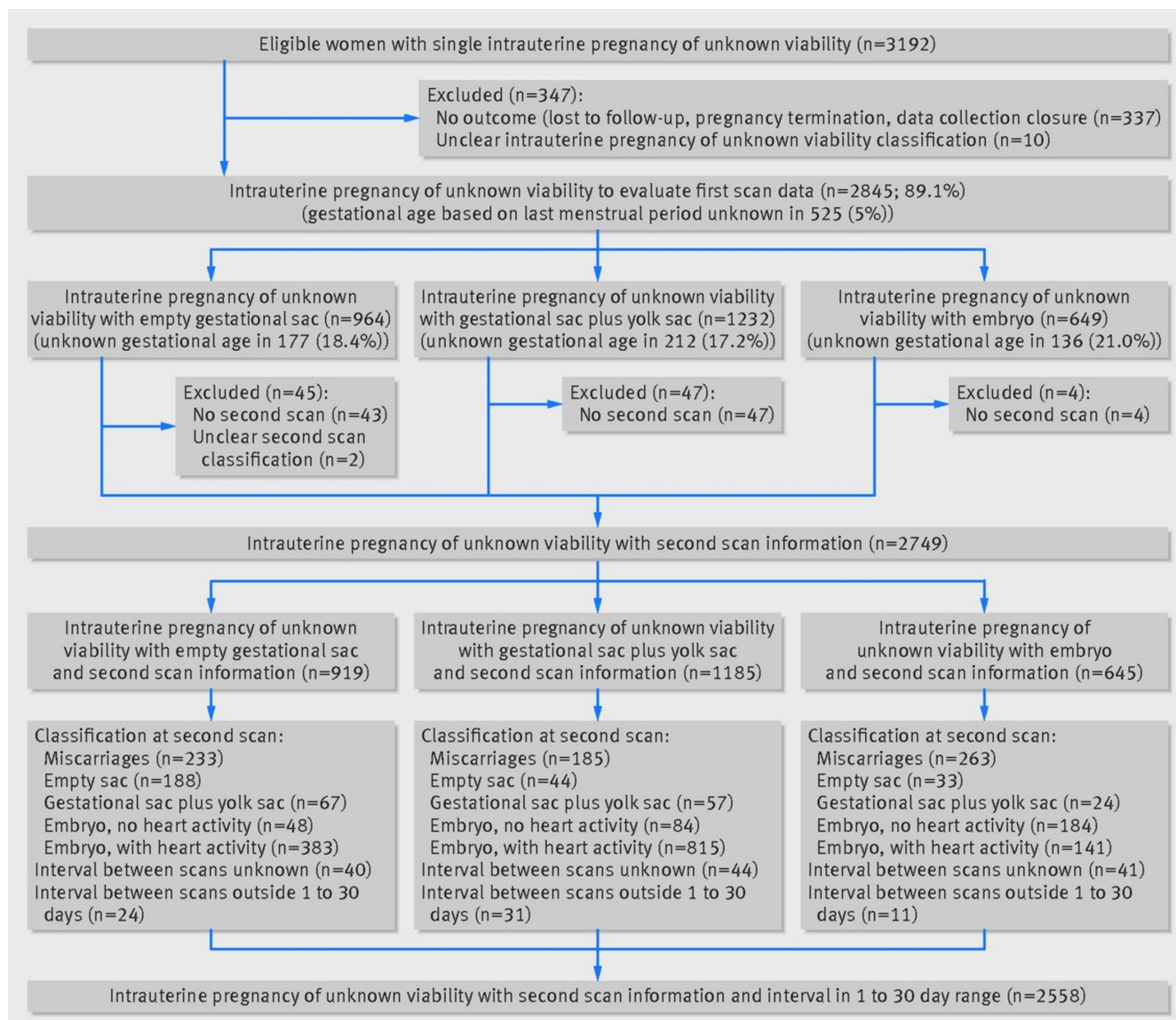


Fig 1 | Fig 1 Flow chart showing number of different types of miscarriage at both initial and follow-up scans, as well as exclusions

First scan results

Descriptive statistics

Women with pregnancies where the final outcome was non-viable presented at later gestations with higher mean gestational sac diameter and crown-rump lengths (if an embryo was visualised) than women with pregnancies that were subsequently viable (see supplementary table S1). Women with pregnancies of uncertain viability with an embryo presented at later gestations and with higher mean gestational sac diameters than others. Gestational age based on last menstrual period was missing in 525/2845 pregnancies (18.5%), and this did not differ by diagnostic group (fig 1).

Mean gestational sac diameter cut-off values

Overall, 49/947 (5.2%) pregnancies with an empty gestational sac had a mean sac diameter of ≥ 20 mm: 47 of the 583 (8.1%) non-viable

pregnancies and two of the 364 (0.5%) viable pregnancies (table 2 and see supplementary fig S1). Twelve of 583 (2.1%) non-viable pregnancies had a mean gestational sac diameter of ≥ 25 mm, whereas the maximum mean sac diameter for pregnancies with an empty gestational sac with a viable outcome was 20.3 mm. The performance of mean gestational sac diameter cut-off values for gestational sacs with a visible yolk sac but no embryo was similar to pregnancies with an empty gestational sac only (see supplementary table S2 and fig S2). The main difference was slightly higher mean gestational sac diameter values—5/764 (0.7%) viable pregnancies had a mean gestational sac diameter of ≥ 20 mm and 1/764 (0.1%) a mean gestational sac diameter of ≥ 25 mm. The maximum mean gestational sac diameter for a viable pregnancy that presented with a visible yolk sac but without an embryo was 25.7 mm.

Table 2 | Diagnostic performance of measurements of mean gestational sac diameter to predict miscarriage in pregnancies with an empty gestational sac. Sensitivity is based on 583 non-viable pregnancies, specificity on 364 viable pregnancies

Mean sac diameter cut-off (mm)	Sensitivity* (n=583)		Specificity† (n=364)		Positive predictive value‡ (%)
	No of women	% (95% CI)	No of women	% (95% CI)	
8	298	51.1 (47.1 to 55.2)	254	69.8 (64.9 to 74.3)	73
10	233	40.0 (36.1 to 44.0)	301	82.7 (78.5 to 86.2)	79
12	181	31.0 (27.4 to 34.9)	327	89.8 (86.3 to 92.5)	83
14	140	24.0 (20.7 to 27.6)	345	94.8 (92.0 to 96.6)	88
16	103	17.7 (14.8 to 21.0)	352	96.7 (94.3 to 98.1)	90
18	70	12.0 (9.6 to 14.9)	360	98.9 (97.2 to 99.6)	95
20	47	8.1 (6.1 to 10.6)	362	99.5 (98.0 to 99.9)	96
21	38	6.5 (4.8 to 8.8)	364	100 (99.0 to 100)	100
22	29	5.0 (3.5 to 7.1)	364	100 (99.0 to 100)	100
23	20	3.4 (2.2 to 5.2)	364	100 (99.0 to 100)	100
24	12	2.1 (1.2 to 3.6)	364	100 (99.0 to 100)	100
25	12	2.1 (1.2 to 3.6)	364	100 (99.0 to 100)	100
26	9	1.5 (0.8 to 2.9)	364	100 (99.0 to 100)	100
27	6	1.0 (0.5 to 2.2)	364	100 (99.0 to 100)	100
28	4	0.7 (0.3 to 1.8)	364	100 (99.0 to 100)	100
29	2	0.3 (0.1 to 1.2)	364	100 (99.0 to 100)	100
30	1	0.2 (0.03 to 1.0)	364	100 (99.0 to 100)	100

*Percentage of non-viable pregnancies that met (≥) cut-off.

†Percentage of viable pregnancies that did not meet (<) cut-off.

‡Percentage of non-viable pregnancies among all pregnancies that met cut-off.

Crown-rump length cut-off values

Overall, 121/637 (19.0%) pregnancies with an embryo on initial scan had a crown-rump length of ≥ 5 mm: 117 of the 527 (22.2%) non-viable pregnancies and four of the 110 (3.6%) viable pregnancies. Seventeen

of 527 (3.2%) non-viable pregnancies had a crown-rump length of ≥ 7 mm, whereas the maximum crown-rump length for pregnancies that ended as viable was 6.0 mm (table 3 and see supplementary fig S3).

Table 3 | Diagnostic performance of crown-rump length to predict miscarriage in pregnancies where embryo with no heart activity has been visualised. Sensitivity is based on 527 non-viable pregnancies; specificity on 110 viable pregnancies

Crown-rump cut-off (mm)	Sensitivity* (n=527)		Specificity† (n=110)		Positive predictive value‡ (%)
	No of women	% (95% CI)	No of women	% (95% CI)	
3.0	380	72.1 (68.1 to 75.8)	86	78.2 (69.6 to 84.9)	94
3.5	308	58.4 (54.2 to 62.6)	95	86.4 (78.7 to 91.6)	95
4.0	237	45.0 (40.8 to 49.2)	99	90.0 (83.0 to 94.3)	96
4.5	172	32.6 (28.8 to 36.8)	104	94.5 (88.6 to 97.5)	97
5.0	117	22.2 (18.9 to 25.9)	106	96.4 (91.0 to 98.6)	97
5.5	80	15.2 (12.4 to 18.5)	107	97.3 (92.3 to 99.1)	96
6.0	56	10.6 (8.3 to 13.6)	109	99.1 (95.0 to 99.8)	98
6.2	38	7.2 (5.3 to 9.7)	110	100 (96.6 to 100)	100
6.4	32	6.1 (4.3 to 8.5)	110	100 (96.6 to 100)	100
6.6	26	4.9 (3.4 to 7.1)	110	100 (96.6 to 100)	100
6.8	22	4.2 (2.8 to 6.2)	110	100 (96.6 to 100)	100
7.0	17	3.2 (2.0 to 5.1)	110	100 (96.6 to 100)	100
7.2	9	1.7 (0.9 to 3.2)	110	100 (96.6 to 100)	100
7.4	8	1.5 (0.8 to 3.0)	110	100 (96.6 to 100)	100
7.6	5	0.9 (0.4 to 2.2)	110	100 (96.6 to 100)	100
7.8	5	0.9 (0.4 to 2.2)	110	100 (96.6 to 100)	100
8.0	4	0.8 (0.3 to 1.9)	110	100 (96.6 to 100)	100

*Percentage of non-viable pregnancies that met (≥) cut-off.

†Percentage of viable pregnancies that did not meet (<) cut-off.

‡Percentage of non-viable pregnancies among all pregnancies that met cut-off.

Gestational age

Viability was reduced when pregnancies of uncertain viability presented at later gestations (fig 2). For pregnancies that ended as viable, the mean gestational sac diameter or crown-rump length measurements were lower as the gestational age at presentation increased. In total, 260 of 2320 (11.2%) pregnancies of uncertain viability with known gestational age presented at 70 days or more gestation: 15 of the 1019 (1.5%) pregnancies that ended as viable and 245 of the 1301 (18.8%) that ended as non-viable. Among these 260 pregnancies, 78 presented with an empty sac (six ended as viable, with a maximum mean gestational sac diameter of 12.7 mm), 49 presented with a visible yolk sac without an embryo (eight ended

as viable, with a maximum mean gestational sac diameter of 12.3 mm), and 133 presented with an embryo (one ended as viable, with a crown-rump length measurement of 1.6 mm). Fifty two pregnancies presented after 70 days or more gestation without an embryo and with a mean gestational sac diameter of ≥18 mm, and all of these were non-viable. A further 87 pregnancies presented with a visible embryo with a crown-rump length of ≥3 mm, after 70 days or more gestation, and all these were non-viable. For a detailed breakdown of diagnostic performance see supplementary table S4 (mean gestational sac diameter measurements) and table S5 (crown-rump length measurements) in relation to gestational age to predict miscarriage.

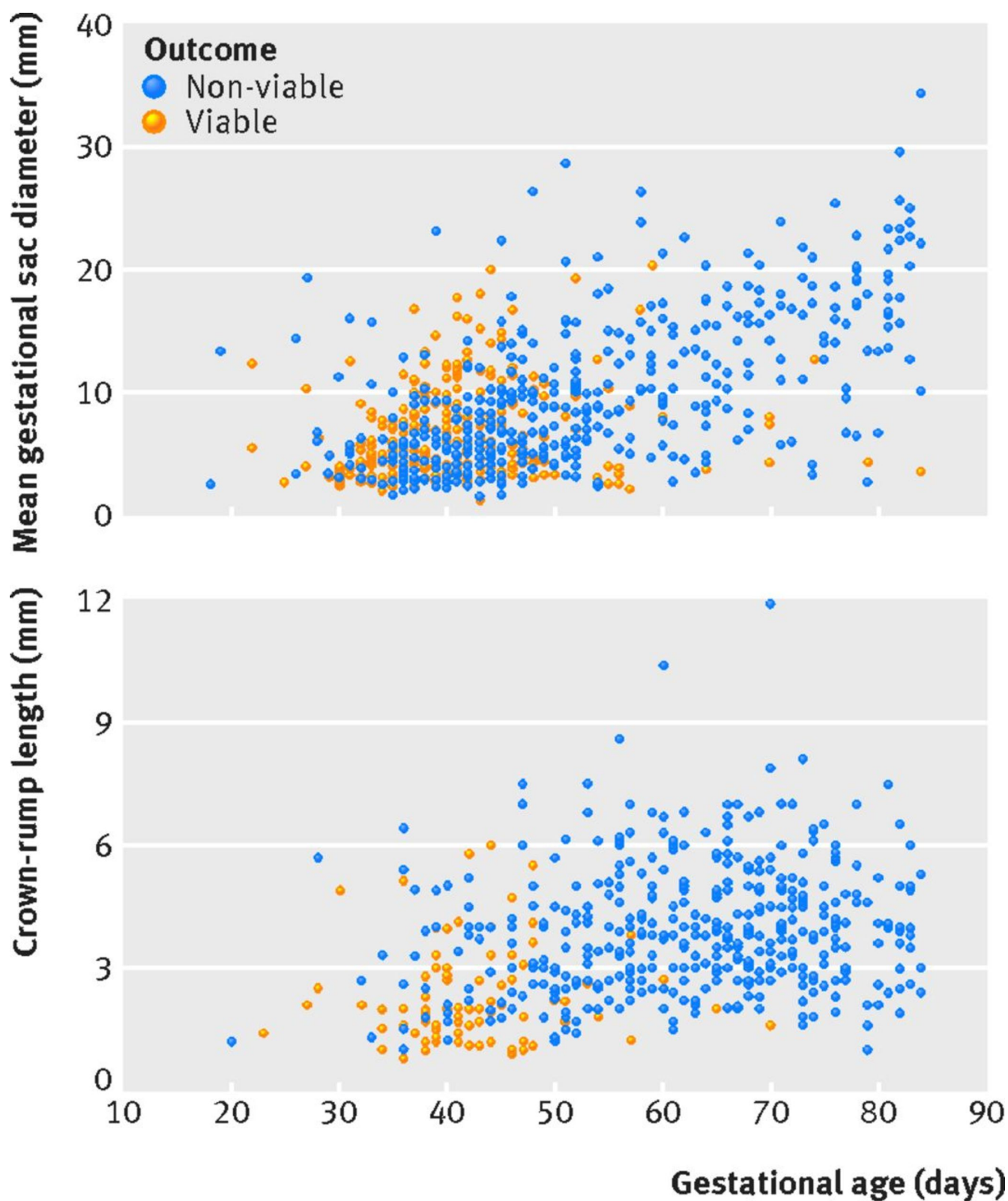


Fig 2 | Fig 2 Scatter plots of gestational age and mean gestational sac diameter for pregnancies with an empty gestational sac (top) and of gestational age and crown-rump length for pregnancies where an embryo was seen with no visible heart activity (bottom)

Repeat scan results

Appearance of structures

Pregnancies presenting at earlier gestations had more favourable outcomes on repeat scans. Median gestation at presentation was around 40 days where heart activity was seen but more than 50 days if the pregnancy had miscarried according to repeat scan findings (see supplementary table S3). Overall, the absence of an embryo with heart activity on a repeat scan was indicative of a poor final outcome, as only 5.8% were viable (42/729) compared with 86.9% when an embryo with a heartbeat was seen on a repeat scan (1163/1339). The viability rate for empty gestational sacs that were still empty on a repeat scan was 4.3% (8/188) but increased to 26.9% (18/67) if a yolk sac was seen and to 12.5% (6/48) for embryos with no heartbeat. If a yolk sac was seen initially, 8.8% (5/57) of embryos were viable when a yolk sac was the only structure seen on repeat scan and 4.8% (4/84) if an embryo with no heartbeat was seen on repeat scan. If an embryo without heart activity was seen on an initial scan, the viability rate was 0% when there was still no heartbeat visible on repeat scanning (0/184).

Interval between scans

Pregnancies with an empty gestational sac and mean sac diameter of ≥ 9 mm and still of uncertain viability on repeat scanning never ended as viable, irrespective of the timing of a repeat scan (fig 3). One pregnancy with an empty gestational sac on initial scan that was still empty on a repeat scan after 14 days ended as viable. The mean gestational sac diameter on initial scan was 3.3 mm but quadrupled in size over 14 days, to 14.1 mm. When a yolk sac was seen on an initial scan, results were comparable to pregnancies with an empty gestational sac (see supplementary fig S4). If the initial mean gestational sac diameter was ≥ 11 mm and viability remained uncertain on a repeat scan, the final outcome was never viable. All pregnancies where there was uncertainty after 10 days or more ended as non-viable. Finally, for embryos with no heartbeat on an initial scan, these never ended viable if the repeat scan showed no heartbeat (fig 4). For more details on diagnostic performance in relation to the interval between scans see supplementary tables S6 and S7.

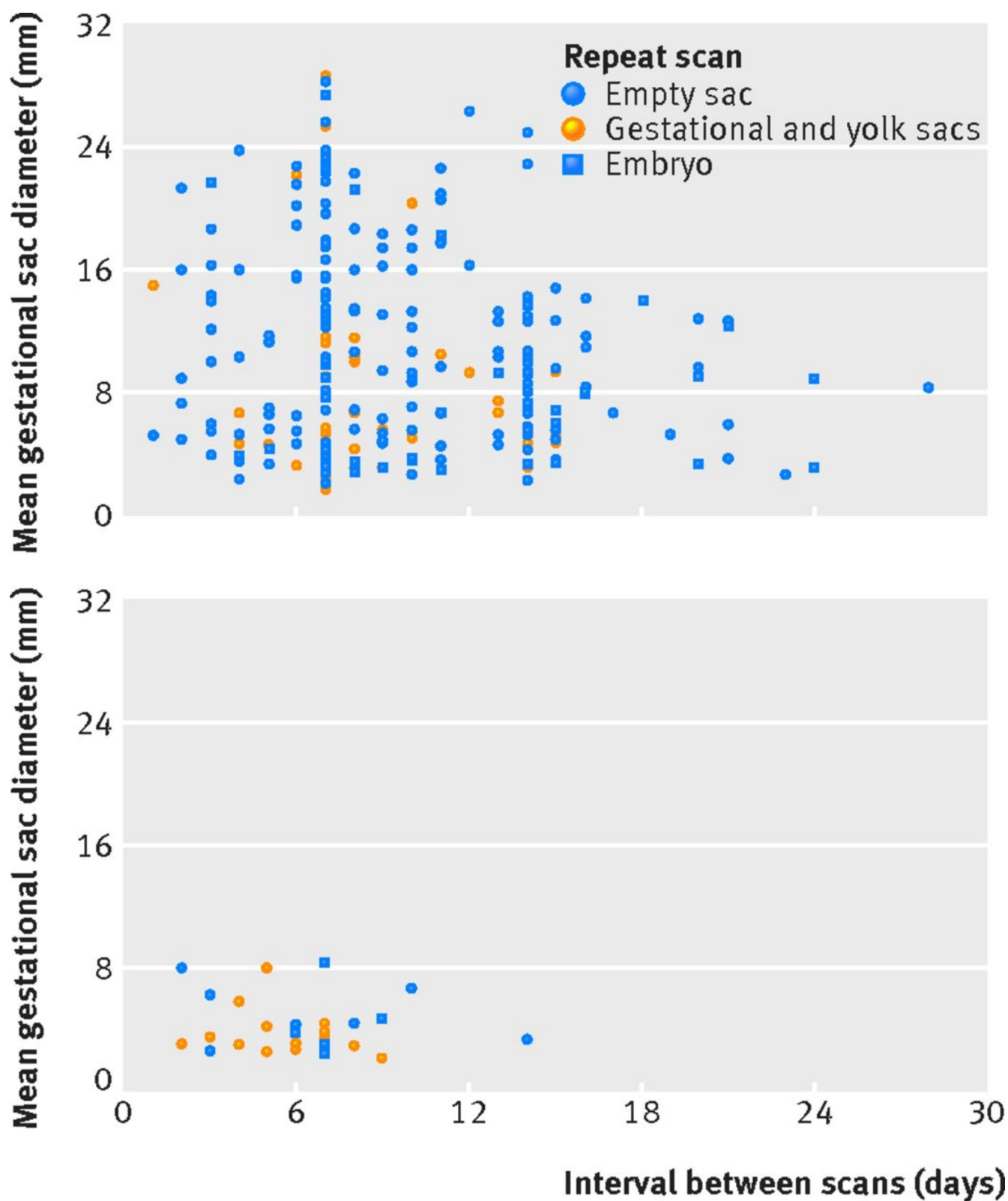


Fig 3 | Fig 3 Scatter plots of initial mean gestational sac diameters and intervals between scans for pregnancies with empty gestational sacs on initial scan but still intrauterine pregnancies of unknown viability on repeat scanning. Data are stratified by repeat scan outcome. (Top) Non-viable pregnancies at 11-14 weeks' outcome. (Bottom) Viable pregnancies at 11-14 weeks' outcome

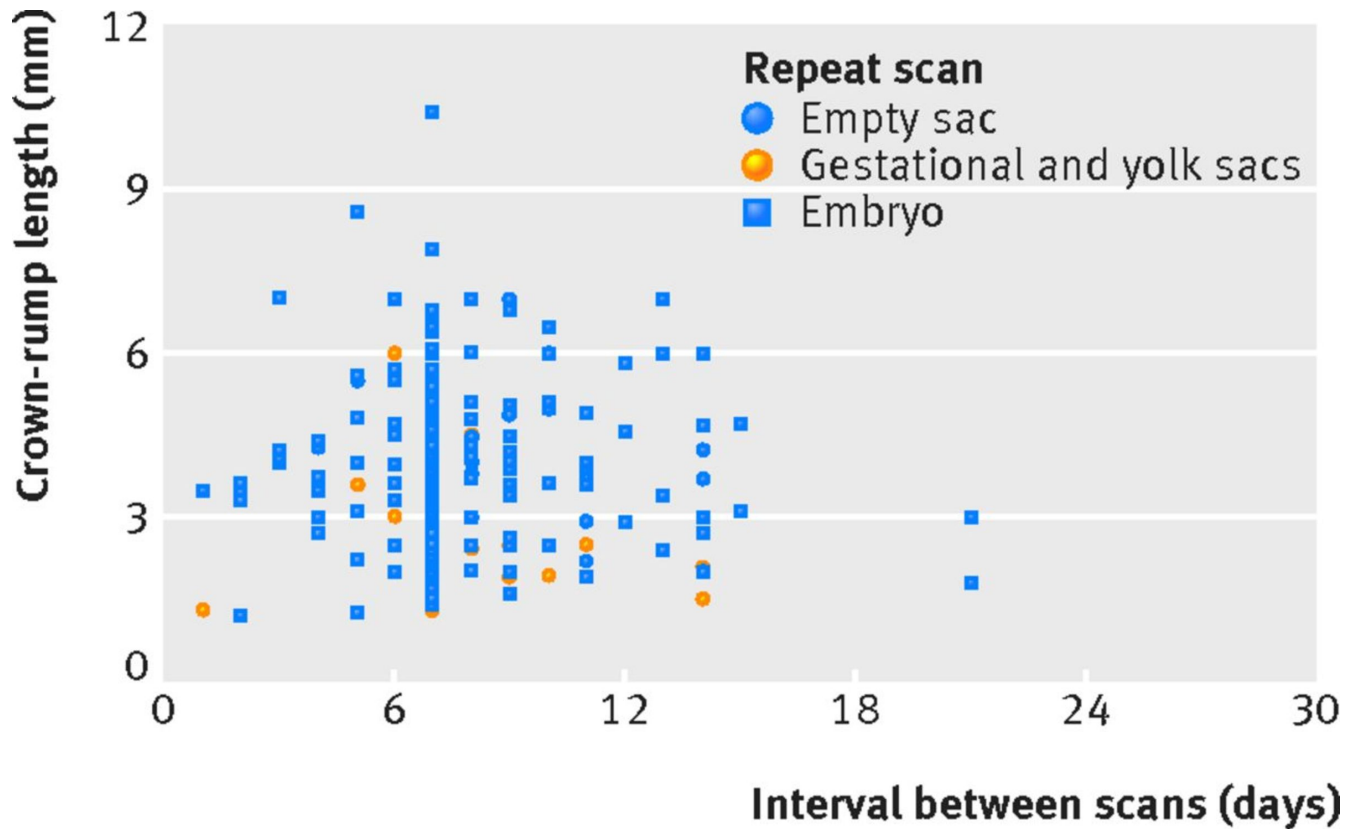


Fig 4 | Fig 4 Scatter plot of initial crown-rump length and intervals between scans for pregnancies with an embryo visualised on initial scan but still intrauterine pregnancies of unknown viability on repeat scanning. Data are stratified by repeat scan outcome. There were no viable pregnancies in this group

Mean gestational sac diameter growth

Where no embryo was seen on initial scan and viability was still uncertain on repeat scanning seven days later, all pregnancies were non-viable if the mean gestational sac diameter increased by less than 50% (0/64) (see supplementary fig S5). Following a repeat scan after 14 days, 1/72 pregnancies still of uncertain viability was viable (the pregnancy where the mean gestational sac diameter quadrupled). The mean gestational sac diameter did not double in 43/72 (60%) cases. Supplementary table S8 shows the diagnostic performance in the absence of embryonic heart activity of mean gestational sac diameter growth in relation to the initial mean sac diameter and the interval between scans to predict miscarriage in pregnancies without an embryo (with or without a yolk sac).

Evaluation of current guidelines

We evaluated selected current guidelines for miscarriage and those suggested by Doubilet et al⁹ (table 4). Diagnostic criteria using mean gestational sac diameter measurements of ≥ 25 mm for empty gestational sacs or crown-rump lengths ≥ 7 mm for embryos without heart activity were reliable. For repeat scans, for embryos with a crown-rump length < 7 mm and no heartbeat on initial scan, no visible heartbeat after seven days is reliable evidence of miscarriage. Guidance relating to time between scans is not robust and may be associated with false positive diagnoses. No guidance considers the presence of embryonic structures on repeat scans. Based on our data we have proposed reliable diagnostic criteria for miscarriage (table 5).

Table 4 | Evaluation of current guidelines using data from study

Source of guidance	Criteria proposed to definitively diagnose miscarriage	Positive predictive value (%; 95% CI)	Specificity (%; 95% CI)
Initial scan criteria:			
ACR, NICE, RCOG, RANZCOG, SRU (Doubilet et al)	Presenting without an embryo or yolk sac, and mean gestational sac diameter ≥ 25 mm	12/12 (100, 73.5 to 100)	364/364 (100, 99.0 to 100)
ACR, NICE, RCOG, RANZCOG, SRU (Doubilet et al)	Presenting with an embryo without heart activity, and crown-rump length ≥ 7 mm	17/17 (100, 80.5 to 100)	110/110 (100, 96.7 to 100)
Repeat scan criteria:			
NICE	Presenting as an empty gestational sac with mean sac diameter < 25 mm or with embryo (no heart activity) with crown-rump length < 7 mm, and returning after at least seven days: no definitive diagnosis (that is, no embryo heart activity visible)	530/549 (96.5, 94.8 to 97.8)	1074/1093 (98.3, 97.4 to 98.9)
ACR	Presenting as an empty gestational sac, and gestational sac still empty 7-10 days later	75/77 (97.4, 91.9 to 99.5)	86/88 (97.7, 92.9 to 99.5)
NICE	Presenting with an embryo (no heart activity) with crown-rump length < 7 mm, and still no embryo heart activity visible after at least seven days	140/140 (100, 97.4 to 100)	103/103 (100, 96.5 to 100)
RCOG	In event of doubt repeat scan in at least one week	Not applicable	
SRU (Doubilet et al)	Presenting with an empty gestational sac, and no embryo heart activity visible after at least 14 days	71/72 (98.6, 93.7 to 99.9)	180/181 (99.5, 97.5 to 99.9)
SRU (Doubilet et al)	Presenting with a gestational sac and yolk sac, and no embryo heart activity visible after at least 11 days	85/85 (100, 95.8 to 100)	537/537 (100, 99.3 to 100)

ACR=American College of Radiology; RCOG=Royal College of Obstetricians and Gynaecologists; NICE=National Institute for Health and Care Excellence; RANZCOG=Royal Australia and New Zealand College of Obstetricians and Gynaecologists; SRU=Society of Radiologists in Ultrasound.

Modified Jeffreys method used for confidence intervals when percentages equalled 100% (or 0%). Standard Jeffreys methods used otherwise.

Table 5 | Proposals for diagnostic criteria for miscarriage based on this study

Our recommendations to definitively diagnose miscarriage	Positive predictive value (%; 95% CI)	Specificity (%; 95% CI)
Agreement with current criteria:		
Presenting with no visible embryo or yolk sac, and mean gestational sac diameter ≥ 25 mm	12/12 (100, 73.5 to 100)	364/364 (100, 99.0 to 100)
Presenting with an embryo with no heart activity, and crown-rump length ≥ 7 mm	17/17 (100, 80.5 to 100)	110/110 (100, 96.7 to 100)
Suggested additional new criteria		
Initial scan criteria:		
Presenting with an embryo with crown-rump length ≥ 3 mm, and gestational age ≥ 70 days	102/102 (100, 96.4 to 100)	87/87 (100, 95.8 to 100)
Presenting with no visible embryo: mean gestational sac diameter ≥ 18 mm and gestational age ≥ 70 days (10 weeks) from date of known last menstrual period	52/52 (100, 93.2 to 100)	907/907 (100, 99.6 to 100)
Repeat scan criteria:		
Presenting with no visible embryo (with or without visible yolk sac) with mean gestational sac diameter ≥ 12 mm and returning after at least seven days: no embryo with embryo heart activity visible	130/130 (100, 97.2 to 100)	150/150 (100, 97.6 to 100)
Presenting without an embryo (with or without visible yolk sac) with mean gestational sac diameter < 12 mm and returning after at least 14 days: no embryo heart activity and mean gestational sac diameter has not doubled	41/41 (100, 91.4 to 100)	478/478 (100, 99.2 to 100)
Presenting with an embryo (irrespective of crown-rump length) without heart activity, and still no heart activity visible after at least seven days	191/191 (100, 98.1 to 100)	103/103 (100, 96.5 to 100)

Modified Jeffreys method used for confidence intervals when percentages equalled 100% (or 0%). Standard Jeffreys methods used otherwise.

Discussion

We have confirmed that previous cut-off values for mean gestational sac diameter of ≥ 16 mm and ≥ 20 mm were not clinically safe and were associated with possible false positive diagnoses of miscarriage. Previous criteria using crown-rump length measurements ≥ 5 mm also carried considerable risk. Our results show that the revised cut-off values recommended by the Royal College of Obstetricians and Gynaecologists, National Institute for Health and Care Excellence, and American College of Radiology for mean gestational sac diameter and crown-rump length to diagnose miscarriage are associated with 100% specificity. We can state this with relatively narrow confidence intervals. Our data suggest that although gestational age is a factor when interpreting ultrasound findings, even with certain menstrual dates, viable pregnancies occur with small gestational sacs and small embryos at relatively late gestations.

Our findings suggest current guidance on repeat scanning in the event of uncertainty should be reviewed. The timing of second scans to determine viability relates to gestational sac size at initial scan. Hence if the mean gestational sac diameter is < 10 mm, repeat scanning should be performed more than 14 days later. This is consistent with Doubilet et al¹⁰ but not with the American College of Radiology which recommends repeat scanning in 7-10 days without considering initial gestational sac size.⁹ Adhering to this approach may be associated with misdiagnosis. Once an embryo is identified, interpretation is more straightforward. We found no pregnancies with an initial crown-rump length measurement of < 7 mm and no heartbeat both initially and on a repeat scan that were viable at 11-14 weeks irrespective of the interval between scans. In these circumstances repeating scans after seven days would seem safe practice.

An important omission from all guidance relates to what examiners should expect on ultrasound scans repeated at an interval. Our data support the assertion by Doubilet et al¹⁰ that once a yolk sac is visualised, if there is no embryo with a heartbeat after 11 days the pregnancy is not viable. For initially empty sacs, if a yolk sac (with no embryo) is visualised on repeat scan the likelihood of viability is substantially increased. Accordingly, scans should be repeated in these circumstances. We found it possible to initially have an empty gestational sac and on repeat scan visualise an embryo of crown-rump length < 7 mm with no heartbeat, resulting in a viable pregnancy at 11-14 weeks. The requirement to repeat scans will inevitably leave women in a state of uncertainty that may be distressing. Evidence shows that giving women information about the likely outcome can manage expectations and reduce anxiety,¹² and scoring systems and models now exist that give an accurate prediction regarding the likelihood of miscarriage.¹³

Strengths and weaknesses of this study

A strength of our study is that it is prospective and large enough to define cut-off values for mean gestational sac diameter and crown-rump length with narrow confidence intervals for specificity. The multicentre design means results should have general applicability. A further strength is the inclusion of information on gestational age and appearance of structures on repeat scans, although 20% of women could not recall the date of their last menstrual period.

The endpoint and reference standard for our study was viability at 11-14 weeks. This is important as false positive diagnoses of miscarriage in this study involve pregnancies that if managed surgically or medically would terminate pregnancies that would be unlikely to miscarry.¹⁴

This analysis contains the 1060 patients from the study by Abdallah et al.⁵ The current study extends patient numbers by 1865, enabling us to calculate cut-off performance with greater confidence. This paper describes the largest prospective study examining criteria to diagnose miscarriage, and recruiting more numbers would not substantially improve knowledge of test performance in relation to mean gestational sac diameter and crown-rump length cut-off values. Notwithstanding the large numbers, a weakness of the study is that around certain decision boundaries, numbers are too small to be definitive. It is clear that there is a relation between the size of a pregnancy on initial scan, time required between scans, and viability on a later scan. We have drawn what we believe are reasonable conclusions about these problems, but it is important to remember that there are few data points in some areas. We did not impute missing data but analysed the available data for every research question because we believe that missing data in this trial is unsystematic with respect to the investigated relations. In general, because the cut-offs investigated lead to near perfect specificity to avoid harm to viable pregnancies, we also believed that it was most appropriate to analyse observed data. Around 10% of women were lost to follow-up. Most of these booked for delivery and had their 11-14 weeks scans carried out elsewhere and we could not trace them. These women had a yolk sac without an embryo visualised more often on the first scan than did those women with complete follow-up, this may suggest that women lost to follow-up were more likely to have a viable pregnancy at 11-14 weeks (see supplementary file for more detailed information about these pregnancies).

Other studies have also illustrated problems associated with previous guidance. In a retrospective study, Rowling et al¹⁵ reported 8% of pregnancies found with an empty gestational sac of 16 mm mean diameter were subsequently found to be viable. A further important problem relates to quality of scanning. For this study we did not introduce additional quality control measures beyond what was routine within the units, which included regular case review meetings, supervision by consultants, and follow-up of outcomes. For the study we carefully reviewed scans around the decision boundaries and carried out data cleaning and checking. Accordingly we believe the trial was pragmatic and the results reflect real clinical practice, where several factors including equipment and training can impact on the interpretation of ultrasound findings. Particular care must be taken when the uterus is distorted by fibroids or a congenital abnormality. For these reasons we feel the advice given by NICE that when ultrasound findings indicate a miscarriage, the findings should be checked by a second examiner is a sensible safeguard.⁸ Finally we have limited information on the small number of women who reached up to 70 days amenorrhoea yet still had a pregnancy of unknown viability. In these cases the timing of the 11-14 week scan was pushed later and the pregnancies formally redated at the time of that scan in line with national policy in the United Kingdom.

Misdiagnosis of miscarriage is not a theoretical problem. Sensitive home pregnancy tests inform women they are pregnant before their missed period.¹⁶ The potential impact of this was illustrated by Bottomley et al,¹⁷ who showed that the likelihood of a scan showing an intrauterine pregnancy of uncertain viability was 26% at 28-34 days' gestation, 60% at 35-41 days, and 29% at 42-48 days. It is likely that if women attend for an ultrasound scan soon after a positive pregnancy test result, a viable pregnancy will not be confirmed, leaving potential for errors about possible miscarriage. This potential is compounded by the fact that diagnosing miscarriage with certainty is problematic as it is reliant on negative findings—that is, the absence of a structure or a heartbeat. For women who are found to have gestational sac or embryo measurements below or

around the cut-off values that we have shown define miscarriage, a definitive diagnosis of miscarriage will only be made after a second scan has been carried out at an interval. In a recent publication, Hu et al¹⁸ found that 126 of 1013 early pregnancies threatening to miscarry (12%) fell into the more conservative zones defined by the size criteria we have proposed for a single scan to define miscarriage than would have been the case had previous US guidance been used (crown-rump length, previously 5 mm now 7 mm; mean sac diameter, previously 16 mm now 25 mm). Before changes in guidance some of these 126 pregnancies would have been at risk of inadvertent termination. The downside now is that these women will undergo a second scan before a diagnosis is made. This does not mean that all women will require a second scan to make a diagnosis, as many will attend for assessment where the embryo or gestational sac size will be considerably above the decision boundaries used and so allow a definitive diagnosis on the basis of a single visit.

Implications for clinical practice

The data presented in this paper offer robust evidence that recent changes in guidance in relation to cut-off values for mean gestational sac diameter and crown-rump length are not too conservative. We have also shown that aspects of guidance to make a diagnosis of miscarriage currently provided by national bodies may be associated with misdiagnosis of miscarriage in some cases. This applies to the presence or absence of embryonic structures that on repeat ultrasound scans are associated with miscarriage and the interval that should be used before making a definitive diagnosis of miscarriage. We have proposed criteria for miscarriage based on our data in table 4, and in table 5 we describe the performance of currently used guidance when applied to our data to illustrate where there are limitations.

What is already known on this topic

- Mean gestational sac diameter (MSD) and crown-rump length (CRL) are used to diagnose miscarriage
- Cut-off values changed in 2011, yet data supporting this change had wide confidence intervals
- Currently, guidance is given for when to repeat scans in the event of uncertain viability on an initial scan and what to find on such scans, but this is not evidence based
- No advice exists on how to relate gestational age to the findings on a scan and a possible diagnosis of miscarriage

What this study adds

- On an initial scan an empty gestational sac of mean diameter ≥ 25 mm was 100% specific for miscarriage, as was an embryo with no heart activity and a CRL ≥ 7 mm
- Beyond 70 days' gestation, an MSD ≥ 18 mm with no embryo was 100% specific for miscarriage as was an embryo with CRL ≥ 3 mm with no heart activity
- For repeat scans, a pregnancy with an embryo with no heart activity on initial scan and a repeat scan ≥ 7 days later was 100% specific for miscarriage, as was a pregnancy with no embryo and an MSD < 12 mm if sac size had not doubled after ≥ 14 days, and pregnancies with no embryo and MSD ≥ 12 mm with no embryo heart activity after ≥ 7 days

Contributors: TB conceived and designed the study. JP, YK, LI, FS, YA, VV, PB, CS, DG, CB, SJ, JH, and CS enrolled patients and acquired data. BVC, CVH, JP, JF, and TB were involved in data cleaning. BVC analysed the data. TB, BVC, and DT were involved in data interpretation. They are the guarantors. TB and BVC wrote the first draft of the manuscript, which was then critically reviewed and revised by the co-authors. All authors approved the final version of the manuscript for submission. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding: TB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The views

expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. DT is a fundamental clinical researcher at the Research Foundation–Flanders. The sponsors had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the work for publication. The researchers performed this work independently of the funding sources.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study was registered as a clinical audit of the performance of cut-off values to diagnose miscarriage (audit registration No 1149) using anonymous data with no change to the usual clinical management of the patients; the project was considered by the national research ethics service in London and they advised that ethical review by a NHS research ethics committee was not required.

Data sharing: No additional data available.

Transparency: The lead authors (TB, BVC, and DT) affirm that that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

- 1 Horrow MM, Andreotti RF, Lee SI, et al. Expert Panel on Women's Imaging. ACR appropriateness criteria® first trimester bleeding. *Am Coll Radiol* 2009;6:235-41. www.excaliburmed.com/Portals/139802/docs/First%20Trimester%20Bleeding%206-12.pdf.
- 2 Royal College of Obstetricians and Gynaecologists. The management of early pregnancy loss. Green-Top Guideline No 25. RCOG Press, 2006.
- 3 Jeve Y, Rhana R, Bhide A, Thangaratnam S. Accuracy of first-trimester ultrasound in the diagnosis of early embryonic demise: a systematic review. *Ultrasound Obstet Gynecol* 2011;38:489-96.
- 4 Pexsters A, Luts J, Van Schoubroeck D, et al. Clinical implications of intra- and interobserver reproducibility of transvaginal sonographic measurement of gestational sac and crown-rump length at 6-9 weeks' gestation. *Ultrasound Obstet Gynecol* 2011;38:510-5.
- 5 Abdallah Y, Daemen A, Kirk E, et al. Limitations of current definitions of miscarriage using mean gestational sac diameter and crown-rump length measurements: a multicenter observational study. *Ultrasound Obstet Gynecol* 2011;38:497-502.
- 6 Abdallah Y, Daemen A, Guha S, et al. Gestational sac and embryonic growth are not useful as criteria to define miscarriage: a multicenter observational study. *Ultrasound Obstet Gynecol* 2011;38:503-9.
- 7 Royal College of Obstetricians and Gynaecologists. Addendum to GTG No 25 (Oct). The management of early pregnancy loss. RCOG Press; 2011.
- 8 Newbatt E, Beckles Z, Ullman R, Lumsden MA, on behalf of the Guideline Development Group. Ectopic pregnancy and miscarriage: summary of NICE guidance. *BMJ* 2012;345:e8136.
- 9 Lane BF, Wong-You-Cheong JJ, Javitt MC, et al. ACR appropriateness criteria: first trimester bleeding. *Ultrasound Q* 2013;29:91-6.
- 10 Doubilet PM, Benson CB, Bourne T, Blaivas M. Diagnostic criteria for nonviable pregnancy early in the first trimester. Society of Radiologists in Ultrasound Multispecialty Panel on Early First Trimester Diagnosis of Miscarriage and Exclusion of a Viable Intrauterine Pregnancy. *N Engl J Med* 2013;369:1443-51.
- 11 Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci* 2001;16:101-17.
- 12 Davison AZ, Appiah A, Sana Y, Johns J, Ross JA. The psychological effects and patient acceptability of a test to predict viability in early pregnancy: a prospective randomised study. *Eur J Obstet Gynecol Reprod Biol* 2014;178:95-9.
- 13 Bottomley C, Van Belle V, Pexsters A, et al. A model and scoring system to predict outcome of intrauterine pregnancies of uncertain viability. *Ultrasound Obstet Gynecol* 2011;37:588-95.
- 14 Westin M, Källén K, Saltvedt S, Almström H, Grunewald C, Valentin L. Miscarriage after a normal scan at 12-14 gestational weeks in women at low risk of carrying a fetus with chromosomal anomaly according to nuchal translucency screening. *Ultrasound Obstet Gynecol* 2007;30:728-36.
- 15 Rowling SE, Coleman BG, Langer JE, Arger PH, Nisenbaum HL, Horii SC. First-trimester US parameters of failed pregnancy. *Radiology* 1997;203:211-7.
- 16 Fletcher JL Jr. Update on pregnancy testing. *Prim Care* 1986;13:667-77.
- 17 Bottomley C, Van Belle V, Mukri F et al. The optimal timing of an ultrasound scan to assess the location and viability of an early pregnancy. *Hum Reprod* 2009;24(8):1811-7.
- 18 Hu M, Poder L, Filly RA. Impact of new society of radiologists in ultrasound early first-trimester diagnostic criteria for nonviable pregnancy. *Ultrasound Med J* 2014;33:1585-8.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.