

Ultrasound as a primary screening tool for detecting low birthweight newborns

A meta-analysis

Eita Goto, MD, MPH*

Abstract

Background: As low birthweight (i.e., birthweight < 2500 g) is a major determinant of neonatal mortality and morbidity, the pre-delivery detection of low birthweight is clinically advantageous. This study was performed to determine whether ultrasound is suitable for use in primary screening to detect low birthweight newborns.

Methods: The primary outcomes included sensitivity, specificity, and positive and negative likelihood ratios of ultrasound detection of low birthweight newborns. Ten databases, including PubMed, were searched. All English language studies that provided true- and false-positive and true- and false-negative results regarding the pre-delivery ultrasound detection of low birthweight newborns were eligible for inclusion in the analysis. Study quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies. Bivariate diagnostic meta-analysis was performed and hierarchical summary receiver operating characteristic curves were constructed.

Results: Studies of relatively good quality were included in the analysis to evaluate crown-rump length ($n = 12$); femur length ($n = 5$); formulas of Campbell, Hadlock, and Shepard ($n = 9$); and uterine artery blood flow ($n = 7$). All showed low sensitivity ($=0.24$ – 0.58) regardless of specificity ($=0.60$ – 0.96). The formulas of Campbell, Hadlock, and Shepard were usable for a confirmation strategy only (positive and negative likelihood ratios = 14.8 and 0.44, respectively), but crown-rump or femur length, and uterine artery blood flow were not usable for an exclusion or confirmation strategy (positive and negative likelihood ratios = 1.4–2.8 and 0.71–0.85, respectively).

Conclusions: Primary screening does not have to confirm low birthweight, but should almost always categorize low birthweight as a positive result and exclude normal birthweight. Therefore, ultrasound is not suitable as a primary screening tool to detect low birthweight newborns.

Abbreviations: AUC = area under the curve, DOR = diagnostic odds ratio, HSROC = hierarchical summary receiver operating characteristic, LR = likelihood ratio, QUADAS = Quality Assessment of Diagnostic Accuracy Studies.

Keywords: low birth weight, meta-analysis, sensitivity and specificity, ultrasonography

1. Introduction

Low birthweight (i.e., birthweight < 2500 g) is one of the major determinants of neonatal mortality and morbidity.^[1] Therefore, early detection of low birthweight newborns may be necessary to ensure the provision of immediate and appropriate care. Maternal anthropometric measurements and symphysis-fundal

height were shown not to be useful for the detection of low birthweight newborns.^[2–4] However, neonatal chest and arm circumferences were demonstrated to be useful parameters for this purpose, especially in developing countries.^[5] However, low birthweight newborns should ideally be detected before rather than after delivery. Recently, ultrasound has become more widely available at hospitals in developing and developed countries. Use of ultrasound is not limited to confirmation of pregnancy but is also used to check for multiple pregnancy, determination of the baby's sex, screening for intrauterine deaths and congenital abnormalities, monitoring of fetal growth and position, prevention of maternal complications, and estimation of gestational age and delivery date. Therefore, ultrasound may be useful for detecting low birthweight newborns prior to delivery.

Here, bivariate diagnostic meta-analysis was performed and hierarchical summary receiver operating characteristic (HSROC) curves were constructed^[6] to determine whether ultrasound is suitable as a primary screening tool for detecting low birthweight newborns prior to delivery.

2. Methods

2.1. Primary outcomes and inclusion criteria

The primary outcomes were sensitivity and specificity, positive and negative likelihood ratios (LRs), diagnostic odds ratio (DOR), and area under the curve (AUC) of ultrasound detection of low birthweight newborns. The inclusion criteria were all the

Editor: Jing Liu.

EG designed the study, acquired, analyzed, and interpreted the data, and drafted the manuscript.

The authors have no funding and conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Department of Medicine and Public Health, Nagoya Medical Science Research Institute, Nagoya, Japan.

* Correspondence: Eita Goto, Department of Medicine and Public Health, Nagoya Medical Science Research Institute, 1-118 Kamenoi, Meitoku, Nagoya 465-0094, Japan (e-mail: egoto1@nifty.com).

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

Medicine (2016) 95:35(e4750)

Received: 13 May 2016 / Received in final form: 12 July 2016 / Accepted: 8 August 2016

<http://dx.doi.org/10.1097/MD.0000000000004750>

English language studies that provided true- and false-positive and true- and false-negative results regarding predelivery ultrasound detection of low birthweight newborns. Studies, in which missing result(s), if any, could be calculated from other data (e.g., number of subjects, prevalence of low birthweight newborns, and sensitivity and specificity) were included in the analysis. The objectives of the studies included in the analysis were not limited to evaluation of the diagnostic performance of ultrasound for detecting low birthweight newborns.

2.2. Search strategies, study selection, and data extraction

PubMed/MEDLINE (i.e., Medical Literature Analysis and Retrieval System Online) was searched (April 13, 2016) using search terms generated by adding the key words to Falck-Ytter filter,^[7] as described in the Online Supplementary Methods, <http://links.lww.com/MD/B235>. There was no limitation of publication date. Articles that were determined to be unrelated to the purpose of the study by scanning the titles and abstracts were excluded. The remaining articles were subjected to full-text retrieval. Articles that were determined to be unrelated by retrieving the full texts were also excluded. Those that remained were potentially eligible articles. Articles that were reviews or that did not provide all of the true- and false-positive and true- and false-negative results by full-text retrieval were also excluded. The remaining articles were finally eligible for inclusion in the analysis. Attempts were made to identify additional eligible articles by investigating: the PubMed-related citations shown by clicking “See all . . .” on the right sides of the PubMed screens displaying potentially eligible articles and the bibliographic references of the potentially eligible articles. Nine other databases were searched, CINAHL (i.e., Cumulative Index of Nursing and Allied Health Literature), PsycInfo (i.e., Psychology Information), Wiley Online Library, ProQuest Central (e.g., ProQuest Health and Medical Complete and ProQuest Nursing & Allied Health Source), ProQuest Dissertations & Theses Global, the entire Cochrane Library (e.g., Cochrane Central Register of Controlled Trials), Web of Knowledge, Google Scholar, and Scopus. The selection process was repeated periodically. The data extracted were as follows: the first author’s name; publication year; country; ultrasound methods used to detect low birthweight newborns; cutoff points; true- and false-positive and true- and false-negative results; Quality Assessment of Diagnostic Accuracy Studies (QUADAS) score (see “Study quality assessment” below);^[8] and the presence or absence of 1 of the 3 major sources of bias in a study included in diagnostic meta-analysis, that is, disregard for the use of the same reference test regardless of the result of the index test, cohort study rather than case-control study, and prospective study rather than retrospective study.^[9,10]

2.3. Study quality assessment

The QUADAS, which is a tool consisting of 14 question items devised to assess quality of studies included in diagnostic meta-analysis,^[8] was used. Study quality was assessed 5 times, and the most frequent responses were considered to be the most appropriate. The QUADAS score was defined as the number of “yes” responses to the question items. For statistical analysis, a value of “1” was assigned to a “yes” response to each question item, and a value of “0” was assigned to a “no” or “unclear” response.

2.4. Statistical analysis

Stata/MP (i.e., multiple processor) 13.1 (Stata Corp LP [i.e., limited partnership], College Station, TX) and R version 3.0.1 (The R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analyses.^[11–14] Attempts were made to detect outlier(s) by model checking^[11,13] using the spike plot of Cook distance for each study and scatter plot of the standardized residuals of healthy (x-axis) and diseased (y-axis) populations for each study. The cutoff point for Cook distance was calculated as 4 times the number of parameters divided by the number of studies. The cutoff point for standardized residual was the standardized 2-level residual. Studies located outside the cutoff points in both of these plots were classified as potential outliers. The potential outliers were omitted as true outliers, if their study designs and materials were different from those of any other studies that were included in the final analysis.

I^2 was used to assess whether the data in the studies were heterogeneous (i.e., $I^2 \geq 50\%$) or homogenous (i.e., $I^2 < 50\%$).^[13,14] An attempt was made to reach homogeneity from heterogeneous data by limiting the studies based on Africa, Asia, Europe, Latin America, the Middle East, North America, or Oceania, versus other regions; developing versus developed countries; QUADAS score ≥ 8 versus < 8 ; and the presence versus absence of 1 of the 3 major sources of bias in a study included in diagnostic meta-analysis (investigation of heterogeneity sources).

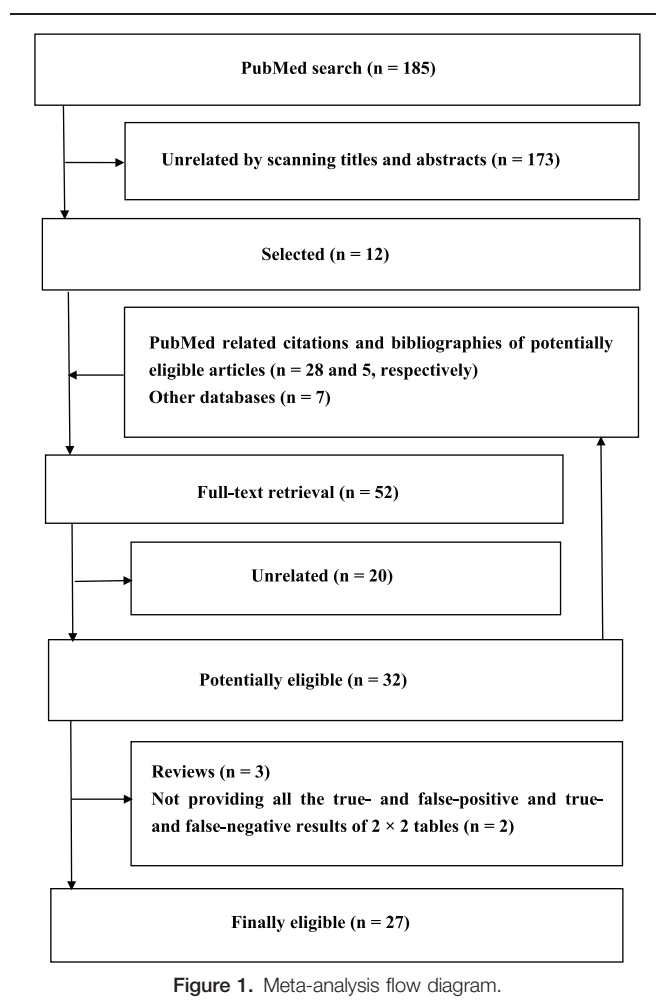
Diagnostic bivariate meta-analysis was performed to summarize sensitivity and specificity, positive and negative LRs, and DOR.^[6,11–14] Informational usability was categorized as exclusion and confirmation strategies, that is, positive LR > 10 and negative LR < 0.1 ; confirmation strategy only, that is, positive LR > 10 and negative LR > 0.1 ; exclusion strategy only, that is, positive LR < 10 and negative LR < 0.1 ; and no exclusion or confirmation strategy, that is, positive LR < 10 and negative LR > 0.1 .^[13] HSROC curves were also constructed to provide AUC and summary points of sensitivity and specificity, 95% confidence regions and prediction regions.^[6,11–14] Prediction region is the region in which sensitivity and specificity of future studies will be plotted with a certain probability (e.g., 95%).

The data were summarized separately in the same way as described for the investigation of heterogeneity sources to limit the studies (subgroup analysis). Subgroup analysis excluded “other regions” because some of these “other regions” may have been located far away from others included in the same “other regions.” Meta-regression was also performed to evaluate the statistical significance of differences in sensitivity and specificity between categories subjected to investigation of heterogeneity sources versus their counterparts.^[13] Publication bias was assessed using Deeks funnel plot asymmetry test.^[15] Cutoff points were proposed using the Youden index, that is, the point located on the HSROC curve that is the most distant from the straight line connecting the origin with the point at the upper right angle.^[16] As all of the data were extracted from the published literature, there was no requirement for ethical approval or informed consent.

3. Results

3.1. Systematic review

Twenty-seven articles were finally deemed to be eligible for the analysis (Fig. 1). However, 1 article evaluating cervical length, 1 article evaluating placental grade, and 1 article evaluating umbilical artery were excluded^[17–19] because the findings based



on 1 data source were not generalizable. Four articles evaluating biparietal diameter were also excluded, because fetuses with shorter biparietal diameter than the lower limit of the predetermined range and with longer biparietal diameter than its upper limit were categorized as a group (positive result), and fetuses with biparietal diameter falling into this range were categorized as another group (negative result) in 3 of these 4 articles.^[20–22] As a result, 7 articles evaluating crown–rump length, 4 articles evaluating femur length, 4 articles evaluating the formulas of Campbell, Hadlock, and Shepard, and 5 articles evaluating uterine artery blood flow were finally included in this meta-analysis (Table 1 and Online Supplementary Results, <http://links.lww.com/MD/B235>). The formulas of Campbell, Hadlock, and Shepard incorporated at least 1 measurement of biparietal diameter, head and abdominal circumferences, and femur length, as described in the Online Supplementary Results, <http://links.lww.com/MD/B235>.

Two or more studies were extracted from some of the included articles that used 2 or more cutoff points or formulas (Table 1). Therefore, 12 studies with 26,493 women evaluating crown–rump length; 5 studies with 8033 women evaluating femur length; 9 studies with 2675 women evaluating the formulas of Campbell, Hadlock, and Shepard; and 7 studies with 1400 women evaluating uterine artery blood flow were finally included in this meta-analysis (Table 2). Studies were conducted in 5 developing and 11 developed countries in Asia, Europe, Latin

America, the Middle East, North America, and Oceania. The prevalence of low birthweight newborns ($=2.0\%–24.9\%$) varied depending on the study setting. On the other hand, longer black and gray bars versus shorter white bars in Fig. 2 indicated overall good quality of the studies included in the analysis. The 3 major sources of bias in a study included in diagnostic meta-analysis were relatively well controlled. That is, the same reference test regardless of the result of the index test, the cohort design, and prospective data collection were used in all of the studies, almost all of the studies ($=91\%$), and almost half of the studies ($=45\%$), respectively.

3.2. Outlier detection and investigation of heterogeneity sources

There were no true outliers in this meta-analysis. The data regarding crown–rump and femur lengths, and uterine artery blood flow were markedly heterogeneous ($I^2=97\%–100\%$, Table 2). This heterogeneity may have been due to the cutoff points, which could be both values corresponding to birthweight of 2500 g and those not corresponding to birthweight of 2500 g (Table 1). The data regarding femur length were homogenous ($I^2=46\%$) by limiting to retrospective studies (Table S1, <http://links.lww.com/MD/B235>) that exclusively used a cutoff point of the 5th percentile. The data regarding the formulas of Campbell, Hadlock, and Shepard were homogenous ($I^2=0\%$, Table 2). This homogeneity may have been due to the cutoff points, which could be only values corresponding to birthweight of 2500 g. The formulas of Campbell, Hadlock, and Shepard also had narrower 95% confidence and prediction regions around the summary point than crown–rump length, femur length, or uterine artery blood flow (Fig. 3). This was consistent with the relatively high proportion of heterogeneity, likely due to the threshold effect ($=0.61–1.00$).

3.3. Meta-analysis and subgroup analysis

On primary screening, almost all low birthweight newborns should be categorized as positive results, but not all normal birthweight newborns have to be categorized as negative results. Therefore, sensitivity is more important than specificity. However, all of the methods examined in this meta-analysis showed low sensitivity, regardless of specificity (Table 2 and Fig. 3). Primary screening should also exclude normal birthweight newborns but does not have to confirm low birthweight newborns. Therefore, an exclusion strategy is more important than a confirmation strategy. However, the informational usability of crown–rump length, femur length, or uterine artery blood flow was categorized as no exclusion or confirmation strategy (Table 2). The informational usability of the formulas of Campbell, Hadlock, and Shepard was categorized as confirmation strategy only. This was the case for the results in all groups subjected to subgroup analysis (Table S1, <http://links.lww.com/MD/B235>).

3.4. Meta-regression analysis

Meta-regression analysis showed a number of possible confounders as follows: Asia versus other regions had an effect on specificity of the formulas of Campbell, Hadlock, and Shepard ($P=0.00$) and sensitivity of uterine artery blood flow ($P=0.00$); Europe versus other regions had an effect on specificity of the formulas of Campbell, Hadlock, and Shepard ($P=0.00$) and

Table 1**Characteristics of studies included in the meta-analysis.**

Author	Country	Design	Cutoff point and/or formula	Number of participants	QUADAS
Crown–rump length					
Kang	South Korea	Cohort	=26.5 mm on day 67	63	5
Mercer	USA	Cohort	1 day smaller than expected	4404	9
Mercer	USA	Cohort	8 days smaller than expected	4405	9
Mook-Kanamori	The Netherlands	Cohort	=20th percentile	1618	8
Pederson	Denmark	Case–control	7 days smaller than expected	403	7
Smith	UK	Cohort	2 days smaller than expected	3237	8
Smith	UK	Cohort	1 day larger than expected	3237	8
Smith	UK	Cohort	2 days smaller than expected; LBW at ≥ 37 wks	3332	8
Smith	UK	Cohort	1 day larger than expected; LBW at ≥ 37 wks	3332	8
Ustunyurt	Turkey	Cohort	2 days smaller than expected	992	6
Ustunyurt	Turkey	Cohort	1 day smaller than expected	992	6
Vafaei	Iran	Cohort	1 day smaller than expected	478	9
Femur length					
Mallath-Pokorny	Austria	Case–control	*Isolated femur length = 5th percentile for GA	317	9
Özli	USA	Cohort	*Isolated femur length = 10th percentile	4992	7
Ventura	Peru	Cohort	*Isolated femur length = 5th percentile for GA	244	8
Weisz	UK	Cohort	=5th percentile for GA	1240	8
Weisz	UK	Cohort	*Isolated femur length = 5th percentile for GA	1240	8
Formulas of Campbell, Hadlock, and Shepard					
Kayem	France/Belgium	Cohort	=292 mm for abdominal circumference in the formula of Campbell	1689	6
Khani	Iran	Cohort	Formula of Hadlock IV	174	9
Kumara	Sri Lanka	Cohort	Formula of Campbell	86	9
Kumara	Sri Lanka	Cohort	Formula of Hadlock I	86	9
Kumara	Sri Lanka	Cohort	Formula of Hadlock II	86	9
Kumara	Sri Lanka	Cohort	Formula of Hadlock III	86	9
Kumara	Sri Lanka	Cohort	Formula of Hadlock IV	86	9
Kumara	Sri Lanka	Cohort	Formula of Shepard	86	9
Prechapanich	Thailand	Cohort	Formula of Shepard	296	9
Uterine artery blood flow					
Asnafi	Taiwan	Cohort	Notch	70	8
Contro	Italy	Case–control	PI = 95 percentile for GA	262	10
Cooper	Canada	Cohort	PI = 1.45 in 18 weeks	229	6
Cooper	Canada	Cohort	PI = 1.45 in 22 weeks	229	6
Dane	Turkey	Cohort	Bilateral notches with RI = 0.52 or unilateral notch with RI = 0.66	404	6
Konchak	USA	Cohort	Notch	103	8
Konchak	USA	Cohort	RI = 0.7 (considered as 95th percentile)	103	8

GA = gestational age, LBW = low birthweight, PI = pulsatility index, QUADAS = Quality Assessment of Diagnostic Accuracy Studies, RI = resistance index.

*Isolated short femur length is defined as femur length <10th or 5th percentile in a fetus with normal abdominal circumference (abdominal circumference ≥ 10 th or 5th percentile).

uterine artery blood flow ($P=0.04$); North America versus other regions had an effect on sensitivity of uterine artery blood flow ($P=0.01$); developing versus developed countries had an effect on specificity of the formulas of Campbell, Hadlock, and Shepard ($P=0.00$); QUADAS score ≥ 8 versus <8 had an effect on the specificity of formulas of Campbell, Hadlock, and Shepard ($P=0.00$); and cohort versus case–control study had an effect on specificity of uterine artery blood flow ($P=0.04$). Within the

limits of availability of P values, no other variables were shown to be confounders that affected sensitivity or specificity ($P=0.06$ – 0.94 or 0.05 – 0.91 , respectively).

3.5. Publication bias and proposed cutoff points

Deeks funnel plot asymmetry test showed publication bias among the data of crown–rump length ($P=0.04$) but not among the data

Table 2**Results of meta-analysis.**

Number of studies	Number of participants	AUC	Sensitivity (95% CI)	Specificity (95% CI)	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)	I^2 (%)
Crown–rump length ($n=12$)	26493	0.60	0.56 (0.39–0.72)	0.60 (0.42–0.76)	1.4 (1.2–1.6)	0.73 (0.64–0.84)	2 (2–2)	100
Femur length ($n=5$)	8033	0.69	0.24 (0.10–0.46)	0.90 (0.79–0.95)	2.3 (1.9–2.9)	0.85 (0.73–0.99)	3 (2–4)	99
Formulas of Campbell, Hadlock, and Shepard ($n=9$)	2675	0.92	0.58 (0.50–0.66)	0.96 (0.95–0.97)	14.8 (10.0–21.9)	0.44 (0.36–0.53)	34 (20–58)	0
Uterine artery blood flow ($n=7$)	1400	0.67	0.39 (0.24–0.58)	0.86 (0.72–0.94)	2.8 (1.6–4.9)	0.71 (0.56–0.89)	4 (2–8)	97

AUC = area under the curve, CI = confidence interval, DOR = diagnostic odds ratio, NLR = negative likelihood ratio, PLR = positive likelihood ratio.

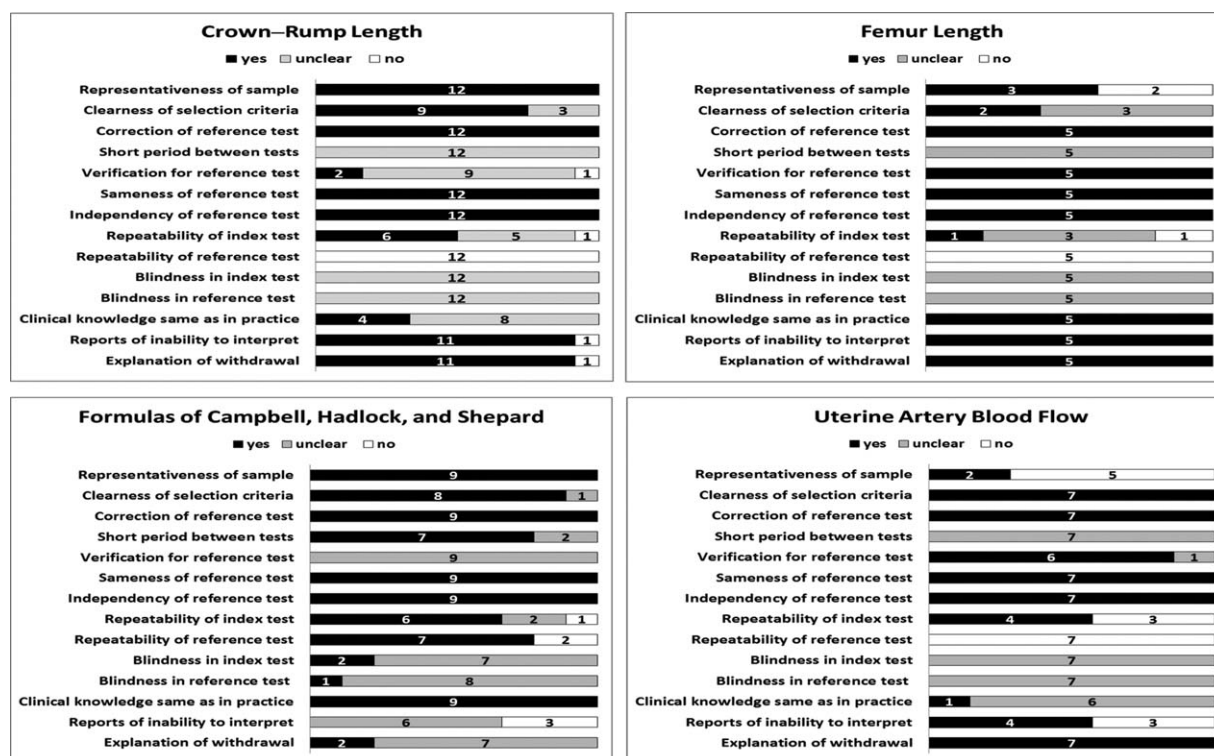


Figure 2. Results of study quality assessment according to the QUADAS. QUADAS=Quality Assessment of Diagnostic Accuracy Studies.

of femur length, the formulas of Campbell, Hadlock, and Shepard, or uterine artery blood flow ($P=0.18$, 0.15 , or 0.53 , respectively) (Figure S1, <http://links.lww.com/MD/B235>).

The Youden index could not propose a cutoff point of crown-rump or femur length, or uterine artery blood flow (Fig. 3), because there were no studies using similar values of the cutoff point around the Youden index. The cutoff points of the formulas of Campbell, Hadlock, and Shepard were determined in advance to correspond to a birthweight of 2500 g.

4. Discussion

4.1. Main findings

Based on the literature search described in the Methods section, the present study is the first meta-analysis to evaluate the diagnostic performance of ultrasound used to detect low birthweight newborns. There is no evidence that ultrasound is suitable as a primary screening tool for detecting low birthweight newborns. This meta-analysis involved 38,601 participants in 16 counties in Asia, Europe, Latin America, the Middle East, North America, and Oceania by including 7, 4, 4, and 5 articles evaluating crown-rump and femur lengths; the formulas of Campbell, Hadlock, and Shepard; and uterine artery blood flow, respectively (Table 1 and Online Supplementary Results, <http://links.lww.com/MD/B235>). Therefore, the findings in the total population (Table 2 and Fig. 3) are relatively generalizable (i.e., external validity). This meta-analysis also included good quality studies, as suggested by more “yes” and “unclear” responses versus fewer “no” responses to the QUADAS question items (Fig. 2). Prospective versus retrospective data collection used by only about half of the studies did not alter the interpretation of results (subgroup analysis and meta-regression analysis), and 2

other major sources of bias in a study included in diagnostic meta-analysis were almost always controlled. Therefore, the findings in the total population (Table 2 and Fig. 3) are not seriously affected by bias due to poor quality of included studies (i.e., internal validity).

4.2. Interpretation

Heterogeneity, possible confounders, or publication bias did not alter the interpretation of the results. Due to the high proportion of heterogeneity likely due to the threshold effect, as mentioned in the Results section, homogeneity ($I^2=0-38\%$) was achieved from all of the heterogeneous data in the total population by excluding the portions of heterogeneity likely due to threshold effects, supporting the justification in summarizing the data. Even by adjusting for almost all of the possible confounders, sensitivity ($=0.23-0.62$) was not sufficiently high and negative LRs ($=0.39-1.00$) were not sufficiently low for exclusion strategy, that is, negative LR < 0.1 . The exception was sensitivity of uterine artery blood flow in Asia ($=0.90$), but this finding was not generalizable, because it was based on only 1 study with small sample size ($=70$). As statistical significance in differences of sensitivity or specificity depending on possible cofounders may have been due to a threshold effect (i.e., ecological fallacy), it is also unclear whether possible cofounders were truly confounders. Based on the slope of funnel plots (Figure S1, <http://links.lww.com/MD/B235>), an increase in effective sample size of the data regarding crown-rump length, among which there is publication bias, contributes to a decrease of DOR, that is, lower levels of diagnostic performance.

Secondary screening should categorize almost all normal birthweight newborns as negative results and confirm low birthweight newborns. The formulas of Campbell, Hadlock, and

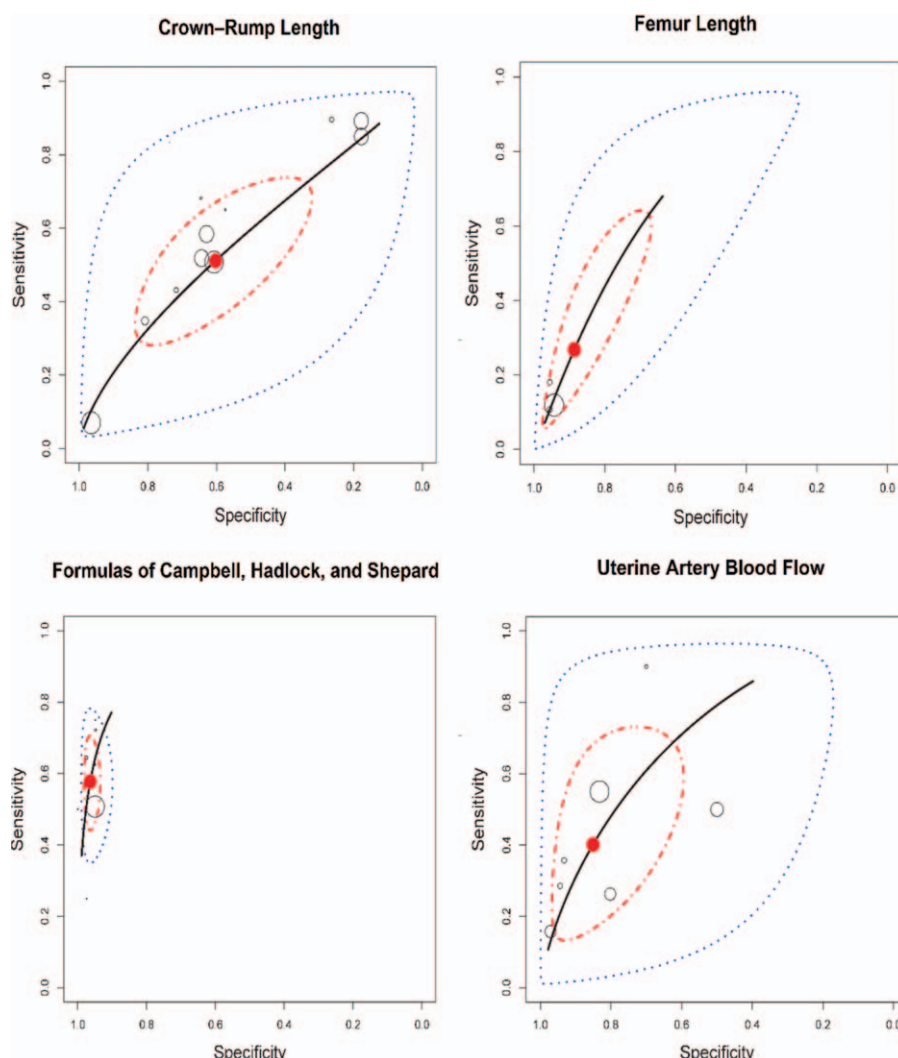


Figure 3. HSROC curves, black circles represent observational studies, red circles represent summary points, black lines represent HSROC curves, red dashed dotted lines represent 95% confidence region, and blue-dotted lines represent 95% prediction region. HSROC=hierarchical summary receiver operating characteristic.

Shepard may be used in secondary screening. In addition to a high degree of accuracy (i.e., $0.9 \leq \text{AUC} < 1.0$),^[23] the formulas of Campbell, Hadlock, and Shepard showed high specificity ($=0.96$), the informational usability was categorized as confirmation strategy only (Table 2 and Fig. 3), and 95% confidence and prediction regions were very narrow (Fig. 3). Despite the differences in formulas, the homogeneity among the data regarding the formulas of Campbell, Hadlock, and Shepard may provide the rationale for jointly summarizing the data.

4.3. Strengths and weaknesses of the study

The first strength of the present study was the accordance between the meta-analysis and procedural guidance to conduct meta-analysis.^[24] The second strength was the use of bivariate meta-analysis to incorporate the negative relationship between sensitivity and specificity^[6,11–14] and Deeks funnel plot asymmetry test to limit the inflation of type I error.^[15] The third strength was the internal and external validity, based on the inclusion of 33 good quality studies with 38,601 participants extracted from 20 data sources (Tables 1 and 2 and Fig. 2).

However, the present meta-analysis had some limitations, including extrapolation of the results to groups that were not subjected to subgroup or meta-regression analysis, for example, Africa versus other regions, males versus females, full-term versus preterm births, and intrauterine growth retardation versus all newborns except those with intrauterine growth retardation (Table S1, <http://links.lww.com/MD/B235>). Non-English language studies were excluded, and only 1 person selected and reviewed the studies. Finally, it was impossible to clarify which of the possible confounders were true confounders.

5. Conclusions

In summary, the results of the present meta-analysis are clinically important for preconceptional strategies to reduce neonatal mortality and morbidity. There is no evidence that ultrasound is suitable for primary screening to detect low birthweight newborns.

Acknowledgments

The author is grateful to the staff of the Medical Library, the Japan Medical Association (Tokyo, Japan), for help in retrieving

the full texts of the articles. English language usage was checked by Dolphin Corporation.

References

- [1] World Health Organization. Feto-maternal nutrition and low birth weight. 2016. http://www.who.int/nutrition/topics/feto_maternal/en/. Accessed August 18, 2016.
- [2] Goto E. Diagnostic value of maternal anthropometric measurements for predicating low birthweight in developing countries: a meta-analysis. *Asia Pac J Clin Nutr* 2015;24:260–72.
- [3] Goto E. Maternal anthropometric measurements as predictors of low birth weight in developing and developed countries. *Arch Gynecol Obstet* 2015;292:829–42.
- [4] Goto E. Prediction of low birthweight and small for gestational age from symphysis-fundal height mainly in developing countries: a meta-analysis. *J Epidemiol Community Health* 2013;67:999–1005.
- [5] Goto E. Meta-analysis: identification of low birthweight by other anthropometric measurements at birth in developing countries. *J Epidemiol* 2011;21:354–62.
- [6] Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:9829–90.
- [7] Falck-Ytter Y, Motschall E. New search filter for diagnostic studies: Ovid and PubMed versions not the same. *BMJ* 2004;328:1040.
- [8] Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- [9] Lijmer JG, Mol BM, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
- [10] Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469–76.
- [11] Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;9:211–29.
- [12] Harbord RM. Metandi postestimation – postestimation tools for metandi. METANDI: Stata module to perform meta-analysis of diagnostic accuracy. 2008. http://fmwww.bc.edu/repec/bocode/m/metandi_postestimation.html. Accessed August 18, 2016.
- [13] Dwamena BA, Sylvester R, Carlos RC. Midas: meta-analysis of diagnostic accuracy studies. 2009. http://fmwww.bc.edu/repec/bocode/m/mi_das.pdf. Accessed August 18, 2016.
- [14] Doebler F. Mada: meta-analysis of diagnostic accuracy. 2015. <https://cran.r-project.org/web/packages/mada/mada.pdf>. Accessed August 18, 2016.
- [15] Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic accuracy was assessed. *J Clin Epidemiol* 2005;58:882–93.
- [16] Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr* 2007;96:644–7.
- [17] Crane JM, Hutchens D. Use of transvaginal ultrasonography to predict preterm birth in women with a history of preterm birth. *Ultrasound Obstet Gynecol* 2008;32:640–5.
- [18] Dudley NJ, Fagan DG, Lamb MP. Ultrasonographic placental grade and thickness: associations with early delivery and low birthweight. *Br J Radiol* 1993;66:175–7.
- [19] Torres PJ, Gratacós E, Alonso PL. Umbilical artery Doppler ultrasound predicts low birth weight and fetal death in hypertensive pregnancies. *Acta Obstet Gynecol Scand* 1995;74:352–5.
- [20] Nakling J, Backe B. Adverse obstetric outcome in fetuses that are smaller than expected at second trimester routine ultrasound examination. *Acta Obstet Gynecol Scand* 2002;81:846–51.
- [21] Nguyen T, Larsen T, Engholm G, et al. A discrepancy between gestational age estimated by last menstrual period and biparietal diameter may indicate an increased risk of fetal death and adverse pregnancy outcome. *BJOG* 2000;107:1122–9.
- [22] Tunón K, Eik-Nes SH, Grøttum P. Fetal outcome when the ultrasound estimate of the day of delivery is more than 14 days later than the last menstrual period estimate. *Ultrasound Obstet Gynecol* 1999;14:17–22.
- [23] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–93.
- [24] Egger M, Smith GD, Altman DG. *Systematic Reviews in Healthcare: Meta-Analysis in Context*. 2nd ed. London:BMJ; 2001.