Peer Review File

Article information: https://dx.doi.org/10.21037/gs-24-135

Reviewer A

1-First, the term “diagnostic value” in the title and elsewhere is vague, I suggest the authors to correct it as “diagnostic accuracy”.

Response：Thank you for your suggestion. After reviewing the article, we believe that updating the "diagnostic value" in the title to "diagnostic accuracy" can better summarize the outcome indicators. The updated title is as follows：

Diagnostic accuracy of automated breast volume scanning, hand-held ultrasound and molybdenum-target mammography for breast lesions: a meta-analysis

2- Second, the abstract needs some revisions. The background needs to specify the controversy regarding the diagnostic accuracy of the three screening methods, the methods need to describe the inclusion criteria according to the PICOS principle, clinical research, and data extraction of included studies, the results need to describe the total sample sizes and quality of included studies, and the conclusion needs comments for the clinical implications of the findings.

Response: Thank you for your suggestions. In this update, we have systematically enriched the content of each part of the abstract. The updated abstract is as follows:

Abstract

Background: *Given the high incidence and increasing burden of breast cancer, more approaches are needed to improve the early diagnosis of breast cancer. The three mainstream diagnostic methods, automated breast volume scanning (ABVS), hand-held ultrasound (HHUS) and mammography, are still controversial in their diagnostic accuracy. The aim of this study is to systematically evaluate the accuracy of three diagnostic methods.*

Methods: *PubMed, EMbase, Web of Science, The Cochrane Library, Wanfang Data, China National Knowledge Infrastructure (CNKI), VIP and SinoMed databases were searched by computer. Studies on the accuracy of ABVS, HHUS and mammography in the diagnosis of benign and malignant breast lesions were collected, and the search*

time limit was from the establishment of the database to August 2022. The Chi-square test was then performed using Meta-Disc software, and the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used for bias and quality assessment.

Results: *A total of 31 studies involving 8107 benign or malignant lesion were included in the meta-analysis. The pooled sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio, and area under curve for HHUS were 0.86 (0.84, 0.87), 0.80 (0.78, 0.81), 4.20 (3.53, 4.99), 0.20 (0.16, 0.24), 22.88 (16.84, 31.08) and 0.898, respectively. And those for ABVS were 0.90 (0.89, 0.91), 0.87 (0.86, 0.88), 7.93 (5.05, 12.45), 0.11 (0.09, 0.15), 74.63 (45.37, 122.76) and 0.956, respectively. And those for molybdenum-target mammography were 0.81 (0.78, 0.84), 0.90 (0.88, 0.91), 6.94 (4.32, 11.17), 0.23 (0.18, 0.29), 31.41 (17.01, 57.98) and 0.887, respectively. Indicators related to patient selection and reference standards suggested a high risk of bias in several included studies.* ***Conclusion:*** *Meta-analysis found a higher diagnostic accuracy of ABVS in benign and malignant breast lesions. This results provide a reference for clinical practitioners in the selection of diagnostic methods, but considering the possible bias of the included studies, the results need to be treated with caution and further verified.*

3-Third, in the introduction, the authors need to explain why the diagnostic accuracy of the three screening methods could be compared and there is a need for the comparisons. As indicated by the authors, the three methods have different indications for the ascertainment of BC. In other words, they are not interchangeable, so the comparisons are not feasible.

Response：Thanks for your suggestion, we have reviewed the contents of the introduction section, rewritten the expressions that may have ambiguities, and supplemented some supporting research evidence as the basis for the comparison of the necessity of the three diagnostic methods. In this part, we hope to express that the existing diagnostic methods have their own advantages and disadvantages, but in the current situation of the declining ability of traditional mammography as the gold standard of breast cancer screening to reduce mortality, the evalution of the overall accuracy of various screening methods is necessary. Some absolute statements in the introduction section of the original manuscript may easily lead reviewers and readers

to misunderstand that the target population of these diagnostic methods has been identified. However, our intention was to explore the overall effectiveness of the three diagnostic methods as a screening tool, and we emphasize this point in the updated text. The updated statement is as follows：

Moreover, the efficacy of molybdenum-target mammography screening in reducing mortality in women younger than 40 years of age has declined because of concerns about radiation and generally thicker breasts.(6)

And the diagnostic efficacy of ABVS for breast cancer has been examined in previous studies, but ABVS is only been recommended as a promising auxiliary diagnostic method for mammography in current studies. Considering the need for simple, rapid, and cost-effective screening methods, evaluating the accuracy of a single screening tool can provide the necessary data to support clinical decision making. However, the current research results have not reached a unanimous conclusion on this issue. Some studies showed that in the diagnosis of breast cancer, ABVS was superior to conventional HHUS. While other studies did not provide the evidence which showed the difference between ABVS and other methods in the diagnostic performance for breast cancer.(9) Given the inconsistent results, we aimed to systematically review the efficacy of ABVS in the diagnosis of breast cancer.

4-Fourth, in the methodology, the authors need to specify the inclusion criteria according to the PICOS, details of the quality assessment and how to incorporate the results of quality assessment into the statistical analysis, as well as the benign BC, how the golden diagnosis of BC was made. In statistics, please specify the test of heterogeneity and how the sources of heterogeneity were identified. Please indicate the P value for statistical significance.

Response: Thanks for your careful review and suggestions, we have re-clarified the Selection Criteria based on PICOS in this update. Details of the QUADAS-2 evaluation are provided. In the inclusion criteria section, it was clarified that the gold diagnostic criterion for BC was biopsy results. At the same time, the description and threshold of heterogeneity test have been updated, as well as the threshold of P value. The updated statement is as follows:

The inclusion criteria were listed as follows:

(1) Population: studies targeted on excisional biopsy confirmed benign or malignant breast lesions patients;

(2) Intervention and comparison: One of the following diagnostic method should be included, ABVS, HHUS or mammography;

(3) Outcomes: Data on diagnostic accuracy need to be provided, including true positive, false negative, positive predictive value and negative predictive value;

(4) Study types: diagnostic test.

For the quality assessment, the QUADAS-2 tool was used which involved four domains including 'index test', 'reference standard', 'patient selection', and 'flow and timing'. Risk of bias in each domain is assessed, and the first 3 domains are also assessed in terms of concerns regarding applicability (10). Signalling questions are included to help judge risk of bias.

The heterogeneity between the results of the study was analyzed by the Cochran Q-test, and the I2 index (I2>50% and Q-test P>0.10 indicated high heterogeneity).

P-values <0.05 were considered statistically significant.

At present, the sources of heterogeneity analysis were based on funnel plot and subgroup analysis. The subgroup analysis showed that DOR values varied in studies with different years or different sample sizes, but it did not change the conclusion that ABVS had the best diagnostic efficacy. In the discussion section, we also mentioned the problem of insufficient analysis of the sources of heterogeneity due to the amount of evidence.

5-Finally, please consider to cite several related papers: 1. Li JM, Shao YH, Sun XM, Shi J. Ultrasonic features of automated breast volume scanner (ABVS) and handheld ultrasound (HHUS) combined with molecular biomarkers in predicting axillary lymph node metastasis of clinical T1–T2 breast cancer. Quant Imaging Med Surg 2024;14(2):1359-1368. doi: 10.21037/qims-23-956. 2. Li W, Zheng Y, Liu H, Tai Z, Zhu H, Li Z, Gu Q, Li Y. Multimodal ultrasound imaging for diagnostic differentiation of sclerosing adenosis from invasive ductal carcinoma. Quant Imaging Med Surg 2024;14(1):877-887. doi: 10.21037/qims-23-524. 3. Tang YC, Cheung YC. Contrast-enhanced mammography-guided biopsy: technique and initial outcomes.

Quant Imaging Med Surg 2023;13(8):5349-5354. doi: 10.21037/qims-23-137.

Response：Many thanks for your suggestion, we carefully studied the contents of the above two literatures and took their findings as evidence of the diagnostic value of ABVS for breast cancer in the introduction section. The relevant statements for the addition are as follows：

Previous studies have confirmed that tumor size diagnosed by ABVS is associated with the diagnostic of clinical T1-T2 breast cancer (OR =1.033; P=0.002) and sclerosing adenosis (sensitivity, specificity, and accuracy: 75%, 86.76%, 73.53%).

Reviewer B

1) Table 2: please add a heading to the top left cell.

398 **Table 2. The diagnostic performance of HHUS, ABVS and molybdenum-target**
399 **mammography for breast cancer.**

	HHUS	ABVS	Molybdenum-target mammography
SEN (95%CI)	0.86 (0.84, 0.87)	0.90 (0.89, 0.91)	0.81 (0.78, 0.84)

Response: Thanks for your suggestion, we have added heading: Diagnostic performance indicators here.

2) Tables 1 and 3: please add unit to Age.

Response: Thanks to your suggestion, we added the unit "years" after the ages of Tables 1 and 3.

3) Table 1: the number of malignant lesions and benign lesions does not sum to the total number of lesions. Please check it.

24	Huang	2018	Prospective	50	76	18	558	
17	Zhu	2017	Prospective	50.6	130	75	58	√

Response: Thank you for your careful review. We have updated the relevant data against the original article.

17	Zhu	2017	Perspecti ve	50.6	130	72	58
24	Huang	2018	Perspecti ve	50	76	18	58

4) Table 3: Please indicate how data are presented. For example, data are presented as mean±standard deviation or median (interquartile range) or number (frequency) or No. (%), etc. Such a description should be based on specific cases. Please do not directly copy the example sentence.

Response: Thank you for your careful review. The data form in Table3 is DOR(95%CI). In this update, we explain the data presentation in the title.

Table 3. The sub-group analysis for DOR value (95%CI) of HHUS, ABVS and molybdenum-target mammography for breast cancer.

Sub-group	HHUS	ABVS	Molybdenum-target mammography
Year of publication			
Before 2016	24.57 (17.08, 35.34)	86.74 (52.93, 142.14)	30.63 (7.44, 126.20)
After 2016	21.18 (12.74, 35.20)	59.21 (26.95, 130.08)	32.86 (15.76, 68.49)

5) Any abbreviations used in figures and tables, as well as their captions, should be defined in a footnote beneath each corresponding table/figure. Even if they were explained in the main text, full terms must be defined again for clarity, so that figures and tables can be read on their own.

Response: Thank you for your suggestion. In this update, we checked the abbreviations that appeared in all the tables and supplemented the full terms in the footnote.

3. When reporting P values, authors should follow our guidelines as listed below. P values reported on main text should be consistent as those on tables and figures.

Reporting of P values:

- The description of the P value should be in the uppercase format, i.e., "P".
- If P value <0.001, report "P<0.001" to avoid reporting unnecessarily excessive precision (except hypothesis tests that include correlations or studies with exponentially small P values, such as genetic association studies, which can be reported exponentially, e.g., P=1×10-5).
- If 0.001≤ P value <0.01, report the specific P value to 3 decimal places, e.g., "P=0.001" or "P=0.009".
- **If P value≥0.01, report the specific P value to 2 decimal places, e.g., "P=0.01" "P=0.06" "P=0.10" "P=0.90". When the P value is near 0.05, report the specific P value to 3 decimal places, e.g., "P=0.046" or "P=0.052".**
- If the P value is >0.99, report "P>0.99".
- Do not report "not significant" simply because the data is greater than an arbitrary value, and do not report only vague bounds such as P<0.05, as described above, but report the exact P value.

Response: Thank you for your detailed explanation of the publication requirements, and we have checked the full text of the p-value reporting form to guarantee compliance with the journal's standards.