89863_Auto_Edited.docx

WORD COUNT 6467
TIME SUBMITTED 10-JAN-2024 05:26PM
PAPER ID 106050324

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 89863

Manuscript Type: META-ANALYSIS

Urea breath test for *Helicobacter pylori* infection in adult dyspeptic patients: a meta-analysis of diagnostic test accuracy

Fabian Fellipe Bueno Lemos, Caroline Tianeze de Castro, Marcel Silva Luz, Gabriel Reis Rocha, Gabriel Lima Correa Santos, Luís Guilherme de Oliveira Silva, Mariana Santos Calmon, Cláudio Lima Souza, Ana Carla Zarpelon-Schutz, Kádima Nayara Teixeira, Dulciene Maria de Magalhães Queiroz, Fabrício Freire de Melo

Abstract

BACKGROUND

Helicobacter pylori infection has been well-established as a significant risk factor for several gastrointestinal disorders. The urea breath test (UBT) has emerged as a leading non-invasive method for detecting *H. pylori*. Despite numerous studies confirming its substantial accuracy, the reliability of UBT results is often compromised by inherent limitations. These findings underscore the need for a rigorous statistical synthesis to clarify and reconcile the diagnostic accuracy of the UBT for the diagnosis of *H. pylori* infection.

AIM

To determine and compare the diagnostic accuracy of ^{13}C-UBT and ^{14}C-UBT for *H. pylori* infection in adult patients with dyspepsia.

METHODS

We conducted an independent search of the PubMed/MEDLINE, Embase, and Cochrane Central databases until April 2022. Our search included diagnostic accuracy studies that evaluated at least one of the index tests (^{13}C-UBT or ^{14}C-UBT) against a reference standard. We used the QUADAS-2 tool to assess the methodological quality of the studies. We utilized the bivariate random-effects model to calculate sensitivity, specificity, positive and negative test likelihood ratios (LR+ and LR-), as well as the diagnostic odds ratio (DOR), and their 95% confidence intervals (95%CI). We conducted subgroup analyses based on urea dosing, time after urea administration, and assessment technique. To investigate a possible threshold effect, we conducted Spearman correlation analysis, and we generated summary receiver operating characteristic (SROC) curves to assess heterogeneity. Finally, we visually inspected a funnel plot and used Egger's test to evaluate publication bias.

RESULTS

The titles and abstracts of 4,621 studies were screened; 79 articles were retrieved and selected for full-text reading. Finally, 60 studies were included in the diagnostic test accuracy meta-analysis. Our analysis demonstrates superior diagnostic accuracy of ^{13}C-UBT over ^{14}C-UBT, indicated by higher sensitivity (96.60% *vs.* 96.15%), specificity (96.93% *vs.* 89.84%), likelihood ratios (LR+ 22.00 *vs.* 10.10; LR- 0.05 vs. 0.06), and AUC (0.979 vs. 0.968). Notably, ^{13}C-UBT's DOR (586.47) significantly outperforms ^{14}C-UBT (DOR 226.50), making it the preferred diagnostic tool for dyspeptic individuals with *H. pylori* infection. Correlation analysis revealed no threshold effect (^{13}C-UBT r = 0.48; ^{14}C-UBT r = -0.01), and SROC curves showed consistent accuracy. Both ^{13}C-UBT and ^{14}C-UBT showed high AUC values (^{13}C-UBT 0.979; ^{14}C-UBT 0.968) near 1.00, reinforcing their excellent accuracy and endorsing both as reliable diagnostic tools in clinical practice.

CONCLUSION

In summary, our study has demonstrated that ^{13}C-UBT has been found to outperform the ^{14}C-UBT, making it the preferred diagnostic approach. Additionally, our results emphasize the significance of carefully considering urea dosage, assessment timing, and measurement techniques for both tests to enhance diagnostic precision. Nevertheless, it is crucial for researchers and clinicians to evaluate the strengths and limitations of our findings before implementing them in practice.

Key Words: Helicobacter pylori; Urea breath test; Diagnosis; Diagnostic test accuracy; Meta-analysis.

Lemos FFB, Castro CT, Silva Luz M, Rocha GR, Correa Santos GL, de Oliveira Silva LG, Calmon MS, Souza CL, Zarpelon-Schutz AC, Teixeira KN, Queiroz DMM, Freire de Melo F. Urea breath test for Helicobacter pylori infection in adult dyspeptic patients: a meta-analysis of diagnostic test accuracy. *World J Gastroenterol* 2023; In press

Core Tip: The urea breath test (UBT) is a pivotal noninvasive method for detecting Helicobacter pylori; however, its reliability is challenging. This meta-analysis aimed to compare the precision of the 13C-UBT and 14C-UBT in diagnosing H. pylori among adults with dyspepsia, providing insights to enhance clinical strategies.

INTRODUCTION

Helicobacter pylori (*H. pylori*) is a spiral-shaped, gram-negative microaerophilic bacterium that infects approximately 43% of the global population[1]. While the majority of infected individuals remain asymptomatic, chronic gastritis inevitably ensues, leading to a significant burden of morbidity and mortality[2,3]. Adults who are infected with *H. pylori* are at increased risk of developing peptic ulcer disease (PUD), gastric cancer (GC), and mucosa-associated lymphoid tissue (MALT) lymphoma[4–6]. To address this, current guidelines advocate for either a test-and-treat or a scope-and-treat

approach in managing uninvestigated dyspepsia, underscoring the importance of timely diagnosis and intervention[7,8].

Diagnostic testing for *H. pylori* infection typically involves two primary categories: invasive (endoscopic) and non-invasive testing, depending on the application of upper endoscopy[9]. For individuals aged 50 years or older or those with alarm features, the recommended standard diagnostic approach involves upper endoscopy, followed by histopathological examination or rapid urease test (RUT), and occasionally, culture[8]. In contrast, in dyspeptic patients under 50 years without specific risk factors or alarm symptoms, non-invasive methods such as urea breath testing (UBT), stool antigen testing (SAT), and serology are preferred[8,10].

Among non-invasive diagnostic techniques, the UBT has emerged as a prominent method. This approach capitalizes on the urease activity of *H. pylori*, initiating the hydrolysis of ingested urea and consequent release of labeled carbon dioxide[11]. Two commonly utilized isotopic variants, ^{13}C-UBT and ^{14}C-UBT, offer distinctive features. In ^{13}C-UBT, a stable isotope (carbon-13) is employed, and breath samples are collected and analyzed for labeled carbon dioxide using various methods such as mass spectrometry and infrared spectrometry[12]. This method presents important advantages, notably the absence of ionizing radiation, rendering it suitable for repeated application and applicable in vulnerable populations, including pregnant women and children[11,13]. In contrast, ^{14}C-UBT utilizes a radioactive isotope (carbon-14) and primarily relies on scintillation counting for detection[14,15]. Despite its historical use, concerns regarding radiation exposure have diminished its preference in contemporary clinical practice.

In a prior meta-analysis, Ferwana *et al.*[16] assessed the diagnostic accuracy of the UBT, encompassing both ^{13}C-UBT and ^{14}C-UBT, for detecting *H. pylori* infection in adult dyspeptic patients. Despite its high accuracy, the reliability of UBT results was constrained by significant unexplained heterogeneity, persisting even after subgroup analysis[16]. This pattern persisted in subsequent studies, with Zhou *et al.*[17] finding analogous challenges in calculating pooled estimates of diagnostic accuracy for ^{14}C-UBT. Moreover, a subsequent systematic review emphasized that the variability in

thresholds and reference standards across studies limited the data available for pooling accuracy measures at specific UBT thresholds[18].

These findings underscore the need for a rigorous statistical synthesis to clarify and reconcile the diagnostic accuracy of the UBT for the diagnosis of *H. pylori* infection, addressing challenges identified in prior research. To address this gap in the evidence, we conducted a systematic review and meta-analysis to determine the diagnostic accuracy of the UBT for *H. pylori* infection in adult patients with dyspepsia.

MATERIALS AND METHODS

This study adhered to the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA)[19]. These guidelines encompass a 27-item checklist and a 3-phase flowchart, both designed to enhance the transparency of systematic review reporting. Accordingly, our study protocol has been officially registered in the PROSPERO database under the registration number CRD42023449854.

Literature search

This search strategy was designed following the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Version 2.0, 2022)[20]. We performed independent, computer-assisted searches of the (1) PubMed/MEDLINE; (2) Embase; and (3) Cochrane Library databases. Medical Subject Headings (MeSH) and Embase Subject Headings (Emtree) index terms and free-text words were combined. Search terms included "urea breath test," "breath test," "13C-urea breath test," "14C-urea breath test," "13C-UBT," "14C-UBT," "*Helicobacter pylori*," "*H. pylori*," and "dyspepsia." Boolean operators (AND, OR) were also used to narrow or broaden the search as required. No language restriction was applied. To identify additional studies, reference lists were also scanned. Finally, we conducted a 'citing reference' search (by searching articles which cited the included studies) in PubMed/MEDLINE and Embase. Following the search, all identified citations were collated and uploaded into the Rayyan (https://www.rayyan.ai/) tool, and all duplicates were removed.

Selection of studies

Two independent reviewers, Lemos FFB and Calmon MS, screened the references against predefined eligibility criteria. In the case of disagreement, a 3rd researcher, Luz MS, was consulted. Full-text papers were obtained for references considered relevant. If any study was not retrieved, the authors were contacted. Two authors, Lemos FFB and Santos GLC, independently screened the full-text papers against the eligibility criteria. In the case of disagreement, consensus was reached.

We included diagnostic accuracy studies that evaluated at least one of the index tests (^{13}C- or ^{14}C-urea breath tests) against a reference standard (biopsy fragments followed by culture or histopathological examination or rapid urease testing (RUT) and/or not serology/stool antigen-based tests in adult dyspeptic patients. Exclusion criteria were as follows: (1) studies that enrolled children or adolescents under 18 years of age; (2) studies that included only patients with acute upper gastrointestinal bleeding; (3) studies that enrolled subjects who presented for reasons other than dyspeptic symptoms, complicated dyspeptic cases that need surgery, those who received previous therapy for *H. pylori* within the last 3 months, or long-term use of corticosteroids and immunosuppressant drugs; (4) screening studies; (5) studies that did not report true positive, false positive, false negative, and true negative data and the threshold used for the index tests; (6) case-control studies because these are prone to bias[21]; (7) full-text articles not available or articles not available in English, Spanish, or Portuguese.

Data extraction and management

Two review authors, Rocha GR and Santos GLC, independently extracted data from each included study using a pre-piloted data extraction form. In case of discrepancies, a 3rd researcher, Lemos FFB, was consulted. The extracted data included: (1) information about the studies, such as the first author, publication year, and country; (2) details about the study design, including the type of study (prospective and retrospective cohort studies, cross-sectional studies, or randomized clinical trials), the reference standards used, blinding of the index test and reference standard, and the flow and timing (retrospective/prospective); (3) participant information, i.e., the total number of

participants and population characteristics (age, mean ± SD, sex, and disease prevalence); (4) reference standard details, including the time interval between the index test and the reference standard; index test information, including the model (^{13}C- or ^{14}C-UBT), cut-off values, urea dosing, time for measurement after urea administration (min), and measurement technique; (5) and diagnostic accuracy data, including the number of true positives, false positives, false negatives, and true negatives.

Assessment of methodological quality

Two independent reviewers, Luz MS and Silva LGO, conducted critical appraisal using the QUADAS-2 tool. In cases of disagreement, they consulted a 3rd researcher, Lemos FFB. The QUADAS-2 tool is applied in four phases[21]: summarizing the review question, tailoring the tool and producing review-specific guidance, constructing a flow diagram for the primary study, and evaluating bias and applicability. This tool comprises four domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed for the risk of bias, and the first three domains are also evaluated for concerns regarding applicability. It's important to note that "risk of bias" refers to internal validity, i.e., whether there are systematic errors in conducting the study with respect to the specific domain, while "applicability concern" pertains to external validity, i.e., whether there are concerns that the population, index test, or reference standard used in the studies align with the review question. Signaling questions were also included to assist in assessing the risk of bias.

Statistical analysis and data synthesis

Eligible studies were subjected to data extraction, and we organized the data into 2x2 tables. In our analysis, we selected only the optimal threshold values for *H. pylori* positivity in cases where multiple thresholds were presented. We added 0.5 to values equal to zero to ensure computational stability and prevent potential issues[22].

To address the anticipated diversity in meta-analyses of diagnostic accuracy studies, we utilized the random-effects model to calculate sensitivity, specificity, positive and negative test likelihood ratios (LR+ and LR-), as well as the diagnostic odds ratio

(DOR)[23]. We also determined the corresponding 95% confidence intervals (CIs). The results of the ^{13}C- and ^{14}C-UBT are presented separately. Subgroup analyses were conducted with a focus on urea dosing, time for measurement after urea administration (in minutes), and the assessment technique employed. To investigate the possibility of a threshold effect in the analysis, we conducted Spearman correlation analysis. A substantial threshold effect was recognized when the correlation coefficient reached or exceeded 0.6[24].

We performed a bivariate random-effects meta-analysis and generated summary receiver operating characteristic (SROC) curves to visually assess heterogeneity. Furthermore, these curves allowed us to predict accuracy by summarizing diagnostic performance as the area under the curve (AUC)[25]. We categorized accuracy levels as follows: fail (0.50 – 0.60), poor (0.61 – 0.70), fair (0.71 – 0.80), good (0.81 – 0.90), and excellent (0.91 – 1.00)[22].

To evaluate publication bias, we conducted a visual inspection of a funnel plot and employed Egger's tests for statistical assessment. The creation of this plot and the assessment of the risk of data due to missing data required a minimum of ten studies.

All analyses were performed using R version 4.2.1, an environment for statistical computing in Vienna, Austria, utilizing the "meta" package (version 6.5-0), "dmetar" package (version 0.1.0), and the "mada" package (version 0.5.11).

RESULTS

Study selection

Database searches initially yielded 10,902 reports, from which duplicates were removed. No additional references were discovered through alternative search methods. Subsequently, the titles and abstracts of 4,621 studies underwent screening, resulting in the retrieval and selection of 79 articles for full-text examination. Ultimately, 60 studies fulfilled the inclusion criteria. The reasons for exclusion were as follows: incorrect population (n = 8), unsuitable reference standard (n = 4), insufficient accuracy measures (n = 2), missing threshold information (n = 2), failure to compare

against the reference standard (n = 1), writing in a foreign language (n = 1), and involvement of the same sample (n = 1). Figure 1 illustrates the flow of information through various phases of the systematic review.

General characteristics of included studies

Among all the studies, 39 (comprising 65%) employed the ^{13}C-UBT as their primary diagnostic test, featuring a median population size of 200 individuals (lower to upper quartile: 84.5-254). Cross-sectional study design was predominant, making up 97.5% of the total, while only one study (2.5%) adopted a randomized controlled trial (RCT) approach. For the ^{13}C-UBT, the median pre-test probability was 51.2% (lower to upper quartile: 47.8-67.6). Various reference standards were used, with the most common being "*H. pylori* culture (*Hp*C) or (histopathological examination (HE) and rapid urease test (RUT))," accounting for 22.5% of cases. Other reference standards included "HE and RUT" (17.5%), "HE" (12.5%), "HE or *Hp*C" (12.5%), "HE or (RUT and serology)" (5%), "RUT" (5%), "HE, *Hp*C, and RUT" (5%), and "*Hp*C" (2.5%). Some studies also combined reference standards, such as "RUT or HE" and "(HE, *Hp*C, RUT) at least two positives," each constituting 2.5% of the sample, as shown in Table 1.

On the other hand, the ^{14}C-UBT accounted for 35% of the total (21 studies) with a median population size of 108.5 (lower to upper quartile: 63.5-125.5). For the ^{14}C-UBT, the median pre-test probability was 64.9% (lower to upper quartile: 43.6-73.1). Various reference standards were employed in these studies, with "HE" being the most prevalent, accounting for 38.1% of cases. Other reference standards included "HE and RUT" (14.3%) and "HpC or HE" (9.5%). Some studies also used combinations of reference standards, such as "HE, RUT, Serology (at least two positive)" and "HpC or [HE and (RUT or Gram staining)]," each comprising 4.8% of the sample, as detailed in Table 2.

Methodological quality assessment

Supplementary Figure 1 and Figure 2 provide a visual representation of the comprehensive methodological quality assessment of the included studies. In the patient selection domain, 22 studies (35.5%) were categorized as having a low risk of

bias, 36 studies (58.1%) were associated with a high risk of bias, and 2 studies (3.2%) were considered to have an unclear risk of bias. In terms of patient selection applicability, 40 studies (64.5%) exhibited low concern, 17 studies (27.4%) showed high concern, and 3 studies (4.8%) had unclear concern.

Within the index test selection domain, 33 studies (53.2%) were rated as having a low risk of bias, 26 studies (41.9%) were identified with a high risk, and 1 study (1.6%) had an unclear risk of bias. Concerning index test applicability, 57 studies (91.9%) displayed low concern, while 3 studies (4.8%) raised high concern.

In the reference standard domain, 47 studies (75.8%) demonstrated a low risk of bias, 12 studies (19.4%) showed a high risk of bias, and 1 study (1.6%) had an unclear risk of bias. Notably, none of the studies raised concerns about the applicability of the reference standard.

Lastly, in the flow and timing domain, 23 studies (37.1%) were associated with a low risk of bias, 23 studies (37.1%) exhibited a high risk of bias, and 7 studies (11.3%) had an unclear risk of bias.

Overall accuracy of the ^{13}C-urea breath test

The ^{13}C-UBT test was evaluated for its diagnostic accuracy in 39 studies *via* a comprehensive meta-analysis[26–63]. The results demonstrated a high sensitivity of 96.60% (95%CI: 95.64-97.56; *P*-value < 0.01; I^2 = 65.0%) and an equally impressive specificity of 96.93% (95%CI: 96.04-97.82; *P*-value < 0.01; I^2 = 58.0%) for this test (refer to Figure 3). Additionally, the diagnostic odds ratio (DOR) was calculated at 586.47 (95%CI: 340.03-1011.51), with a positive likelihood ratio (LR+) of 22.00 (95%CI: 15.60-30.10) and a negative likelihood ratio (LR-) of 0.05 (95%CI: 0.04-0.06) as presented in Supplementary Table 1.

Subgroup analysis of the ^{13}C-UBT — Urea dosing

Among the thirty-six studies that documented the urea dosage, a 25 mg urea dose demonstrated notably high sensitivity (98.85%; 95%CI: 95.68-100.00) and specificity (99.13%; 95%CI: 96.73-100.00), as illustrated in Supplementary Figure 2. Increasing the urea dose to 50 mg across four studies resulted in a sensitivity of 95.28% (95%CI: 88.51-

100.00) and a specificity of 94.91% (95%CI: 87.67-100.00). Seventeen studies explored the use of 75 mg of urea in the ^{13}C-UBT, revealing a sensitivity of 96.47% (95%CI: 95.14-97.79) and a specificity of 98.33% (95%CI: 97.59-99.07). In cases where 100 mg of urea was used (in 12 studies), the ^{13}C-UBT demonstrated a sensitivity of 97.31% (95%CI: 95.92-98.70) and a specificity of 96.08% (95%CI: 94.34-97.82). Two studies employing 125 mg of urea showed a sensitivity of 93.76% (95%CI: 86.13-100.00) and a specificity of 88.66% (95%CI: 81.07-96.25). Lastly, in a single study using 250 mg of urea, the ^{13}C-UBT exhibited a sensitivity of 97.06% (95%CI: 91.38-100.00) and a specificity of 98.59% (95%CI: 94.72-100.00).

Subgroup analysis of the ^{13}C-UBT — Time for assessment after urea administration

Among the 36 studies that provided information on the time after urea administration, optimal sensitivity (98.87%; 95%CI: 98.14-99.60) and specificity (98.14%; 95%CI: 96.98-99.30) were achieved when the assessment was conducted 20 minutes after urea administration [in 7 studies (see Supplementary Figure 3)]. Notably, there were variations in sensitivity and specificity for different time intervals following urea administration.

For tests conducted 5 minutes post-urea administration (in one study), sensitivity was 97.83% (95%CI: 93.61-100.0), and specificity was 96.08% (95%CI: 90.75-100.00). Tests performed 10 minutes after urea administration (based on one study) yielded a sensitivity of 97.56% (95%CI: 90.88-100.0) and a specificity of 97.56% (95%CI: 90.88-100.00).

Similarly, in the case of tests carried out at 15 minutes post-urea administration (as reported in five studies), sensitivity averaged at 97.61% (95%CI: 95.68-99.55), with specificity at 95.85% (95%CI: 91.33-100.00). Longer intervals, such as 30 minutes and 60 minutes, as well as tests conducted at multiple time points after urea administration, displayed some variability. For instance, tests performed 30 minutes after urea administration (in 19 studies) had a sensitivity of 95.15% (95%CI: 93.30-96.92) and a specificity of 96.18% (95%CI: 94.48-97.87). A single study conducting tests 60 minutes post-urea administration reported a sensitivity of 96.03% (95%CI: 92.62-99.44) and a

specificity of 97.83% (95%CI: 93.61-100.00). In the case of four studies investigating multiple time points after urea administration, the sensitivity was 96.13% (95%CI: 92.13-100.0), and the specificity was 97.95% (95%CI: 96.08-99.81).

Subgroup analysis of the ^{13}C-UBT — Assessment technique

In our analysis of 38 studies that included data on the 13C-UBT assessment technique, Integrated Cavity Output Spectrometry (ICOS) for measuring CO_2 Isotope Ratios exhibited exceptional performance. ICOS demonstrated a sensitivity of 98.99% (95%CI: 96.20-100.00) and a specificity of 98.55% (95%CI: 94.56-100.00), as visualized in Supplementary Figure 4. In contrast, Infrared spectrometry, assessed in 8 studies, displayed a sensitivity of 94.72% (95%CI: 90.91-98.54) and a specificity of 98.55% (95%CI: 88.17-98.22).

Gas chromatography-mass spectrometry, investigated in a single study, yielded a sensitivity of 91.67% (95%CI: 84.67-98.66) and a specificity of 93.02% (95%CI: 85.41-100.00). Isotope-ratio mass spectrometry, scrutinized in 17 studies, demonstrated a sensitivity of 97.37% (95%CI: 96.45-98.28) and a specificity of 98.38% (95%CI: 84.67-98.66). Molecular correlation spectrometry, examined in a solitary study, exhibited a sensitivity of 97.83% (95%CI: 93.61-100.00) and a specificity of 96.08% (95%CI: 90.75-100.00). Similarly, Laser opto-galvanic effect spectroscopy, reported in one study, recorded a sensitivity of 91.65% (95%CI: 88.31-94.78) and a specificity of 91.92% (95%CI: 88.61-98.21).

Overall accuracy of the ^{14}C-Urea breath test

A total of 21 studies investigated the diagnostic accuracy of the ^{14}C-UBT, revealing a combined sensitivity of 96.15% (95%CI: 94.47-97.82; *P*-value < 0.01; I^2 = 62.0%) and specificity of 89.84% (95%CI: 84.90-94.77; *P*-value < 0.01; I^2 = 78.0%), as depicted in Figure 4[26,64–83]. Within this dataset, a diagnostic odds ratio (DOR) of 226.50 (95%CI: 102.57-500.15), a positive likelihood ratio (LR+) of 10.10 (95%CI: 5.74-16.90), and a negative likelihood ratio (LR-) of 0.06 (95%CI: 0.04-0.08) were observed, as summarized in Supplementary Table 1.

Subgroup analysis of the ^{14}C-UBT — Urea dosing

Twenty-one studies investigated varying urea dosages in the context of the ^{14}C-UBT. Among these, the use of a 5 μCi marked urea dose, as examined in four studies, demonstrated exceptional sensitivity (99.21%; 95%CI: 98.20-100.00) and specificity (93.43%; 95%CI 86.45-100.00), as depicted in Supplementary Figure 5. Elevating the urea dose to 10 μCi, as explored in a single study, resulted in a sensitivity of 96.72% (95%CI: 92.15-100.00) and a specificity of 80.00% (95%CI: 56.76-100.00). Conversely, when employing 1 μCi of marked urea (in 14 studies), the ^{14}C-UBT exhibited a sensitivity of 96.78% (95%CI: 95.46-98.09) and a specificity of 87.19% (95%CI: 59.76-95.81). Lastly, two studies using 0.75 μCi of urea reported a sensitivity of 88.94% (95%CI: 76.10-100.00) and a specificity of 91.32% (95%CI: 78.18-100.00).

Subgroup analysis of the ^{14}C-UBT — Time for measurement after marked urea administration

When considering the time for measurement after urea administration, an analysis of all included studies consistently revealed the highest sensitivity (98.39%; 95%CI: 96.36-100.00) and specificity (98.71%; 95%CI: 96.58-100.00) when the tests were conducted 15 minutes after urea administration, as illustrated in Supplementary Figure 6.

In studies conducted shortly after urea administration (within 10 minutes, n = 9), the sensitivity was consistently high at 97.83% (95%CI: 96.34-99.33), while specificity was somewhat lower at 79.90% (95%CI: 66.15-93.65). A single study, conducted at 12.5 minutes post-administration, reported a sensitivity of 96.05% (95%CI: 91.67-100.00) and a specificity of 95.12% (95%CI: 88.53-100.00). Studies conducted between 10- and 15-minutes post-urea administration (n = 3) showed a sensitivity of 94.92% (95%CI: 89.31-100.00) and a specificity of 96.00% (95%CI: 88.32-100.00).

However, longer intervals (20, 25, and 30 minutes), as well as tests conducted at various time points after urea administration, exhibited more variability. For instance, studies conducted at 20 minutes post-administration (n = 3) showed a sensitivity of 96.52% (95%CI: 93.50-97.55) and a specificity of 97.23% (95%CI: 94.48-99.97). A single study conducted at 25 minutes post-urea administration reported a sensitivity of 82.29% (95%CI: 76.89-87.69) and a specificity of 84.00% (95%CI: 75.70-92.30). A study conducted

at 30 minutes post-administration yielded a sensitivity of 97.78% (95%CI: 91.69-100.00) and a specificity of 95.12% (95%CI: 88.53-100.00). In the case of two studies that investigated multiple time points after urea administration, the sensitivity was 96.03% (95%CI: 91.79-100.00), and the specificity was 91.02% (95%CI: 76.07-100.00).

Subgroup analysis of the ^{14}C-UBT — Assessment technique

In the assessment of 20 studies with available data on the assessment technique, it was observed that liquid scintillation counting yielded a higher sensitivity of 98.79% (95%CI: 97.90-99.69) while maintaining a specificity of 87.24% (95%CI: 77.69-96.79). Conversely, Solid Scintillation UBT (scintillation counting) demonstrated higher specificity, reaching 97.46% (95%CI: 94.62-100.00), with a sensitivity of 95.40% (95%CI: 91.00-99.80), as illustrated in Supplementary Figure 7.

In contrast, the Heliprobe Analyser, assessed in 7 studies, displayed a sensitivity of 95.41% (95%CI: 93.32-97.50) and a specificity of 88.10% (95%CI: 74.43-100.00). Ultimately, the use of Beta-scintillation counter for the assessment of ^{14}C-UBT resulted in a sensitivity of 98.11% (95%CI: 95.33-100.00) and a specificity of 93.47% (95%CI: 88.11-98.82).

Threshold effect and SROC curve

Spearman's correlation analysis for studies evaluating ^{13}C-UBT revealed a correlation coefficient (r) of 0.48, indicating the absence of a threshold effect. Similarly, ^{14}C-UBT studies exhibited a negligible correlation (r = -0.01), also suggesting the absence of a threshold effect. Visual inspection of the Summary Receiver Operating Characteristic (SROC) curves did not reveal any significant heterogeneity. Both the ^{13}C-UBT (AUC = 0.979, see Figure 5A) and the ^{14}C-UBT (AUC = 0.968, see Figure 5B) displayed excellent diagnostic accuracy.

Publication bias

The funnel plot visualization exposed asymmetry in both the ^{13}C-UBT (Figure 6A) and ^{14}C-UBT (Figure 6B) models. Additionally, Egger's test confirmed the presence of publication bias in both tests. The intercept was 2.54 with a *P*-value < 0.001 for ^{13}C-UBT and 3.04 with a *P*-value < 0.001 for ^{14}C-UBT.

DISCUSSION

Insights from ^{13}C- and ^{14}C-UBT performance analysis

Our analysis has revealed that the ^{13}C-UBT outperforms the ^{14}C-UBT in terms of diagnostic accuracy, as evidenced by the following values: Diagnostic Odds Ratio (DOR), Likelihood Ratios (LR+ and LR-), and Area Under the Curve (AUC) values. Specifically, the ^{13}C-UBT has sensitivity and specificity values of 96.60% (95%CI: 95.64-97.56; *P*-value < 0.01; I^2 = 65.0%) and 96.93% (95%CI: 96.04-97.82; *P*-value < 0.01; I^2 = 58.0%), respectively. In contrast, the ^{14}C-UBT has sensitivity and specificity values of 96.15% (95%CI: 94.47-97.82; *P*-value < 0.01; I^2 = 62.0%) and 89.84% (95%CI: 84.90-94.77; *P*-value < 0.01; I^2 = 78.0%). The LR+ values for the ^{13}C-UBT and ^{14}C-UBT are 22.00 and 10.10, respectively, indicating the likelihood of positive results in individuals with *H. pylori* infection. Conversely, the LR- values, suggesting a reduced likelihood of negative test results for individuals with the infection, are 0.05 for the ^{13}C-UBT and 0.06 for the ^{14}C-UBT.

Furthermore, the DOR values show a substantial difference between the two tests. The ^{13}C-UBT yields a significantly higher DOR of 586.47 compared to the ^{14}C-UBT's DOR of 226.50. These results indicate that the ^{13}C-UBT is statistically superior at distinguishing dyspeptic individuals with and without *H. pylori* infection, making it the preferred diagnostic tool in this clinical context.

Finally, it is essential to emphasize that our correlation analysis, utilizing both the ^{13}C-UBT (*r* = 0.48) and the ^{14}C-UBT (r = -0.01), yielded no evidence of a threshold effect. Visual examination of the SROC curves revealed no heterogeneity, indicating consistent accuracy assessments across the studies. Additionally, both the ^{13}C-UBT and the ^{14}C-UBT displayed remarkably high AUC values: 0.979 for the ^{13}C-UBT and 0.968 for the ^{14}C-UBT, which approaching 1.00 reinforces the excellent accuracy of these tests in detecting *H. pylori* infection in individuals with dyspepsia. These findings strongly support the reliability of the ^{13}C-UBT and the ^{14}C-UBT as valuable diagnostic tools in clinical practice.

^{13}C-UBT performance: urea dose, assessment timing, and measurement technique selection

Our analysis highlights the critical importance of selecting the appropriate urea dose when conducting the ^{13}C-UBT for diagnosing *H. pylori* infection. While the 25 mg urea dose displays the highest sensitivity (98.85%) and specificity (99.13%), concerns regarding the generalizability of these results arise due to the fact that these findings are primarily based on a single study[33]. In contrast, the use of 75 mg and 100 mg doses is supported by a larger body of evidence, maintaining excellent diagnostic accuracy with sensitivity and specificity exceeding 96%. Conversely, higher doses, such as 125 mg or 250 mg, exhibit a modest reduction in accuracy, particularly in terms of specificity. These findings strongly advocate for the consideration of 75 mg and 100 mg doses when aiming to optimize both sensitivity and specificity.

A crucial factor affecting the performance of the ^{13}C-UBT is the timing of the assessment following urea administration. Our observations reveal that the optimal sensitivity and specificity, both exceeding 98%, are achieved at the 20-minute mark post-urea administration. Tests conducted at shorter intervals, such as 5 minutes and 10 minutes, also demonstrate high sensitivity and specificity, albeit slightly lower than the 20-minute assessment. Conversely, assessments at 15 minutes maintain excellent accuracy, with sensitivity close to 98% and specificity around 95%. However, assessments at longer intervals, such as 30 minutes, 60 minutes, and multiple time points, exhibit some variability, with sensitivity and specificity values slightly lower than the 20-minute assessment. These results highlight the 20-minute assessment as the most reliable time point, offering a balance between high sensitivity and specificity. Nevertheless, the test remains accurate when conducted at shorter intervals.

The choice of assessment technique is also crucial for test accuracy. Integrated Cavity Output Spectrometry (ICOS) is the most accurate technique, with a sensitivity of 98.99% and a specificity of 98.55%. However, it is important to note that ICOS was evaluated in a single study[29], potentially limiting the generalizability of these results. To address this limitation, Isotope-ratio mass spectrometry is a more advisable option. In contrast,

Infrared spectrometry, gas chromatography-mass spectrometry, isotope-ratio mass spectrometry, molecular correlation spectrometry, and Laser opto-galvanic effect spectroscopy yield varying levels of sensitivity and specificity. These findings underscore the significance of selecting the right assessment technique. While ICOS may be preferred when available due to its exceptional accuracy, other factors such as cost, availability, and local expertise should also be considered when making this choice.

^{14}C-UBT performance: urea dose, assessment timing, and measurement technique selection

Our research indicates that the urea dosage utilized in the ^{14}C-UBT can also impact test accuracy. Specifically, a urea dose of 5 μCi was examined in four studies and was found to possess exceptional sensitivity (99.21%) and specificity (93.43%). These findings underscore the potential benefits of employing a 5 μCi dose for the ^{14}C-UBT, as it offers a high level of accuracy in detecting *H. pylori* infection. However, increasing the urea dose to 10 μCi, as investigated in a single study[83], resulted in a slightly lower sensitivity (96.72%) and a specificity of 80.00%. This suggests that while higher urea dosages may still provide reliable results, they may be associated with a decrease in specificity, which could lead to more false-positive results.

On the other hand, the use of 1 μCi of marked urea, which was the most commonly used dosage in 14 studies, resulted in a sensitivity of 96.78% and a specificity of 87.19%. This indicates that a 1 μCi dose remains a viable option for the ^{14}C-UBT, offering a good balance between sensitivity and specificity. Two recent studies using 0.75 μCi of urea reported a sensitivity of 88.94% and a specificity of 91.32%, suggesting that even lower urea doses can provide reasonable diagnostic accuracy[26,64] .

Regarding the time for measurement, tests conducted 15 minutes after urea administration consistently exhibited the highest sensitivity (98.39%) and specificity (98.71%). This indicates that the 15-minute time point is optimal for maximizing test accuracy. Tests conducted within 10 minutes post-administration maintained high sensitivity (97.83%) but had a somewhat lower specificity (79.90%). A single study

conducted at 12.5 minutes post-administration reported favorable sensitivity (96.05%) and specificity (95.12%)[73]. In contrast, longer intervals (20, 25, and 30 minutes) showed more variability, with varying levels of sensitivity and specificity. This suggests that measurements taken beyond 15 minutes may not be as reliable for *H. pylori* detection. Clinicians should carefully consider the timing of the ^{14}C-UBT to ensure accurate results, with a preference for the 15-minute mark when feasible.

Lastly, our analysis of assessment techniques uncovered differences in sensitivity and specificity. Liquid scintillation counting demonstrated the highest sensitivity (98.79%) but had a specificity of 87.24%. In contrast, Solid Scintillation UBT (scintillation counting) showed higher specificity (97.46%) at the expense of sensitivity (95.40%). The Heliprobe Analyser and Beta-scintillation counter also demonstrated moderate sensitivity and specificity. When choosing the assessment technique, the trade-off between sensitivity and specificity should be considered in relation to the clinical context. For instance, if high sensitivity is paramount to avoid missing positive cases, liquid scintillation counting may be the preferred method. Conversely, if high specificity is crucial to minimize false positives, solid scintillation counting could be a better choice.

Strengths and limitations

This meta-analysis adhered to established guidelines and rigorous methodological principles, enhancing the validity and reliability of our findings. We used a bivariate random-effects model to calculate sensitivity, specificity, likelihood ratios, and the diagnostic odds ratio, alongside generating SROC curves for a comprehensive statistical analysis of the included studies. Subgroup analyses based on urea dosing, measurement timing, and assessment technique were conducted to explore potential sources of variation, while Spearman correlation analysis was used to assess the threshold effect's impact on diagnostic accuracy. Additionally, we assessed publication bias through visual inspections of funnel plots and Egger's tests.

However, it's important to acknowledge inherent limitations in our analysis. These include potential language bias, reliance on available data, and challenges associated

with the inherent heterogeneity in diagnostic accuracy studies. Although we did not impose language restrictions in our search, the inclusion of studies conducted in English, Spanish, or Portuguese may introduce language bias[85]. The exclusion of studies due to unavailability of full-text articles or articles not in these specified languages could potentially lead to the omission of essential data.

Furthermore, the quality of our meta-analysis is closely tied to the quality of the primary studies we included. Biases within these primary studies can affect our analysis outcomes. In particular, we have concerns regarding the inclusion of patients, as there was no reported consecutive patient inclusion in some studies, and the index test was not always performed using a pre-specified threshold. Moreover, the diversity in diagnostic accuracy studies can present challenges when consolidating results, and despite subgroup analyses, residual heterogeneity may impact the broad applicability of our findings. Encouragingly, the visual examination of the SROC curves indicates consistent accuracy assessments across the included studies. Nevertheless, it is imperative to underscore that the reliability of our meta-analysis hinges on the data provided in these included studies. The absence or inconsistency of critical data points can significantly affect the precision of our analysis. Researchers and clinicians should consider these strengths and limitations when applying our findings in their practice.

CONCLUSION

In summary, our study offers crucial insights for selecting optimal diagnostic methods to detect *H. pylori* infection in clinical settings. We found that the ^{13}C-UBT outperforms the ^{14}C-UBT in terms of diagnostic accuracy, making it the preferred diagnostic approach. Furthermore, our findings highlight the significance of precise considerations when choosing urea dosage, assessment timing, and measurement techniques for both the ^{13}C-UBT and ^{14}C-UBT, thus enhancing diagnostic precision.

These insights provide practical guidance to healthcare practitioners when choosing the most suitable diagnostic method for *H. pylori* infection, tailored to their specific clinical context. Factors like diagnostic accuracy, cost, and availability should be carefully

weighed in this decision-making process. Our findings also have the potential to contribute significantly to the standardization of testing procedures, ensuring consistent and reliable results, especially for patients with dyspepsia or suspected *H. pylori* infection. Nevertheless, it's essential for researchers and clinicians to consider the strengths and limitations when applying our findings in their practice.

ARTICLE HIGHLIGHTS

Research background

The urea breath test (UBT) has become a widely accepted non-invasive method for detecting Helicobacter pylori (*H. pylori*). While numerous studies have confirmed its high accuracy, its reliability is often hindered by inherent limitations.

Research motivation

In a previous investigation, the diagnostic accuracy of the UBT, which encompasses both 13C-UBT and 14C-UBT, was evaluated in adult patients with dyspepsia to determine the presence of *H. pylori* infection. Although the test demonstrated a high degree of precision, its reliability was compromised by significant and unexplained heterogeneity, which persisted even after conducting subgroup analyses. This trend continued in subsequent studies, with similar challenges encountered in determining pooled estimates of diagnostic accuracy for 14C-UBT. Furthermore, a subsequent systematic review revealed that the variability in thresholds and reference standards across studies limited the available data for pooling accuracy measures at specific UBT thresholds. These findings underscore the need for a rigorous statistical synthesis to clarify and reconcile the diagnostic accuracy of the UBT for the diagnosis of *H. pylori* infection, addressing challenges identified in prior research.

Research objectives

To evaluate and contrast the diagnostic accuracy of ^{13}C-UBT and ^{14}C-UBT for *H. pylori* infection in adult patients with dyspepsia.

Research methods

We conducted independent searches of PubMed/MEDLINE, Embase, and Cochrane Central databases until April 2022, focusing on diagnostic accuracy studies that evaluated at least one of the index tests (^{13}C-UBT or ^{14}C-UBT) against a reference standard. We utilized the QUADAS-2 tool to assess the methodological quality of the studies, and we calculated sensitivity, specificity, positive and negative test likelihood ratios (LR+ and LR-), as well as the diagnostic odds ratio (DOR) and their 95% confidence intervals (95%CI) using the bivariate random-effects model. We conducted subgroup analyses based on urea dosing, time after urea administration, and assessment technique. To investigate a possible threshold effect, we conducted Spearman correlation analysis, and we generated summary receiver operating characteristic (SROC) curves to assess heterogeneity. Lastly, we visually inspected a funnel plot and used Egger's test to evaluate publication bias.

Research results

A screening of 4,621 studies led to the selection of 60 articles for inclusion in a diagnostic test accuracy meta-analysis after full-text reading. Our analysis highlights the superior diagnostic accuracy of ^{13}C-UBT compared to ^{14}C-UBT, as evidenced by higher sensitivity (96.60% vs. 96.15%), specificity (96.93% vs. 89.84%), likelihood ratios (LR+ 22.00 vs. 10.10; LR- 0.05 vs. 0.06), and AUC values (0.979 vs. 0.968). Particularly noteworthy is the significantly higher DOR of ^{13}C-UBT (586.47) compared to ^{14}C-UBT (DOR 226.50), establishing ^{13}C-UBT as the preferred diagnostic tool for individuals with dyspepsia and *H. pylori* infection. Correlation analysis indicated no threshold effect for both ^{13}C-UBT (r = 0.48) and ^{14}C-UBT (r = -0.01), and the SROC curves consistently demonstrated accurate performance for both tests. The high AUC values (^{13}C-UBT 0.979; ^{14}C-UBT 0.968), nearing 1.00, further affirm the excellent accuracy of both UBT variants, solidifying their reliability as diagnostic tools in clinical practice.

Research conclusions

Our study establishes ^{13}C-UBT as the superior diagnostic approach over ^{14}C-UBT. Furthermore, our findings underscore the critical importance of meticulously considering factors such as urea dosage, assessment timing, and measurement techniques for both tests to optimize diagnostic accuracy. However, it is paramount for researchers and clinicians to thoroughly evaluate the strengths and limitations of our conclusions before integrating them into clinical practice.

Research perspectives

Future research should focus on improving the comprehension, practicality, and dependability of urea breath tests for *H. pylori* infection. This endeavor involves refining techniques, examining sources of variability, exploring threshold effects, conducting longitudinal and comparative investigations, addressing biases, and assessing cost-effectiveness.

89863_Auto_Edited.docx

ORIGINALITY REPORT

13%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	246 words — 3%
2	pubs.rsc.org Internet	130 words — 2%
3	core.ac.uk Internet	86 words — 1%
4	f6publishing.blob.core.windows.net Internet	64 words — 1%
5	www.wjgnet.com Internet	50 words — 1%
6	rsdjournal.org Internet	40 words — 1%
7	Nell, Sjoerd, Jakob W. Kist, Thomas P.A. Debray, Bart de Keizer, Timotheus J. van Oostenbrugge, Inne H.M. Borel Rinkes, Gerlof D. Valk, and Menno R. Vriens. "Qualitative elastography can replace thyroid nodule fine-needle aspiration in patients with soft thyroid nodules. A systematic review and meta-analysis", European Journal of Radiology, 2015. Crossref	33 words — < 1%

8	Jianzhong Mo. "Effects of several oral medications on the results of the 14C-urea breath test", Chinese Journal of Digestive Diseases, 7/2001 Crossref	28 words —	< 1%
9	Haolin Fei, Ziwei Wang, Stefano Tedeschi, Andrew Kennedy. "Boosting visual servoing performance through RGB-based methods", Robotic Intelligence and Automation, 2023 Crossref	27 words —	< 1%
10	www.journalslibrary.nihr.ac.uk Internet	22 words —	< 1%
11	Giulia Scioscia, Donato Lacedonia, Carla Maria Irene Quarato, Pasquale Tondo et al. "Could transthoracic ultrasound be useful to suggest a small airways disease in severe uncontrolled asthma?", Annals of Allergy, Asthma & Immunology, 2022 Crossref	17 words —	< 1%
12	academic.oup.com Internet	16 words —	< 1%
13	Hill, S.A.. "Evidence for the use of B-type natriuretic peptides for screening asymptomatic populations and for diagnosis in primary care", Clinical Biochemistry, 200803 Crossref	15 words —	< 1%
14	www.researchsquare.com Internet	15 words —	< 1%
15	Jeffrey Hom, Samuel H.F. Lam, Kristen M. Delaney, Jessica A. Koos, Sergey Kunkov. "VOMITING, PYLORIC MASS, AND POINT-OF-CARE ULTRASOUND: DIAGNOSTIC TEST ACCURACY FOR	14 words —	< 1%

HYPERTROPHIC PYLORIC STENOSIS—A META-ANALYSIS", The Journal of Emergency Medicine, 2023
Crossref

16 Sanjay Divakaran, Garrick C. Stewart, Neal K. Lakdawala, Robert F. Padera et al. "Diagnostic Accuracy of Advanced Imaging in Cardiac Sarcoidosis", Circulation: Cardiovascular Imaging, 2019 — 14 words — < 1%
Crossref

17 Sylvain Gaudet, Mickaël Begon, Jonathan Tremblay. "Cluster analysis using physical performance and self-report measures to identify shoulder injury in overhead female athletes", Journal of Science and Medicine in Sport, 2019 — 14 words — < 1%
Crossref

18 gastrores.org — 14 words — < 1%
Internet

19 rcastoragev2.blob.core.windows.net — 14 words — < 1%
Internet

20 A Aram, H Hong, C Song, M Bass, JA Platt, S Chutinan. "Physical Properties and Clinical Performance of Short Fiber Reinforced Resin-based Composite in Posterior Dentition: Systematic Review and Meta-analysis", Operative Dentistry, 2023 — 13 words — < 1%
Crossref

21 Arnaud Uguen. "Digital Pathology Slides-based Measurement of Tumor Cells and Lymphocytes Within Cytology Samples Supports the Relevance of the Separation by Size of Nonhematological Tumor and Hematological Nontumor Cells in Liquid Biopsies", Applied Immunohistochemistry & Molecular Morphology, 2021 — 13 words — < 1%
Crossref

22 Salvatore Giovanni Vitale, Gaetano Riemma, Sergio Haimovich, Jose Carugno et al. "Risk of endometrial cancer in asymptomatic postmenopausal women in relation to ultrasonographic endometrial thickness: systematic review and diagnostic test accuracy meta-analysis", American Journal of Obstetrics and Gynecology, 2022
Crossref
13 words — < 1%

23 jamanetwork.com
Internet
13 words — < 1%

24 wprim.whocc.org.cn
Internet
13 words — < 1%

25 Gabriele Campanella, Cristian Navarrete-Dechent, Konstantinos Liopyris, Jilliana Monnier et al. "Deep Learning for Basal Cell Carcinoma Detection for Reflectance Confocal Microscopy.", Journal of Investigative Dermatology, 2021
Crossref
12 words — < 1%

26 Weiwei Gu, Li Gong, Xiaodan Lou, Jiang Zhang. "The Hidden Flow Structure and Metric Space of Network Embedding Algorithms Based on Random Walks", Scientific Reports, 2017
Crossref
12 words — < 1%

27 meridian.allenpress.com
Internet
12 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON
EXCLUDE SOURCES < 12 WORDS
EXCLUDE MATCHES < 12 WORDS