Nuclear Imaging for Classic Fever of Unknown Origin: Meta-Analysis

Motoki Takeuchi; Issa Dahabreh; Takashi Nihashi; Mitsunaga Iwata; George Varghese; Teruhiko Terasawa.

Institutions:
1. Department of Emergency and General Internal Medicine, Fujita Health University School of Medicine, Toyoake, Japan
2. Center for Evidence-based Medicine, Brown University, Providence, RI, USA
3. Department of Radiology, Nagoya University Graduate School of Medicine, Nagoya, Japan
4. Department of Infectious Diseases, Christian Medical College, Vellore, India

Running Title:
Nuclear Imaging for Classic FUO

For correspondence:
Motoki Takeuchi (motoki-t@fujita-hu.ac.jp)
Teruhiko Terasawa (terasawa@fujita-hu.ac.jp)
Fujita Health University Hospital, 1-98 Dengakugakubo, Kutsukakecho, Toyoake, Aichi, 470-1192 Japan
Phone/Fax: +81-562-93-2497

Conflict of interests:
All authors have nothing to disclose.

Word counts:
Text: 4,999

Funding:
MEXT, Japan (No. 26460755)
ABSTRACT

Aim: Several studies have assessed nuclear imaging tests for localizing the source of fever in patients with classic fever of unknown origin (FUO); however, the role of these tests in clinical practice remains unclear. We systematically reviewed the test performance, diagnostic yield, and management decision impact of nuclear imaging tests in patients with classic classic FUO.

Methods: We searched PubMed, Scopus, and other databases through 31 October 2015 to identify studies reporting on the diagnostic accuracy or impact on diagnosis and management decisions of fluorodeoxyglucose positron emission tomography alone (FDG-PET) or integrated with computed tomography (FDG-PET/CT), gallium scintigraphy, or leukocyte scintigraphy. Two reviewers extracted data. We quantitatively synthesized test performance and diagnostic yield, and descriptively analyzed evidence about the impact on management decisions.

Results: We included 42 studies with 2058 patients. Studies were heterogeneous and had methodological limitations. Diagnostic yield was higher in studies with higher prevalence of neoplasms and infections. Non-neoplastic causes, such as adult-onset Still's disease and polymyalgia rheumatica, were less successfully localized. Indirect evidence suggested that FDG-PET/CT had the best test performance and diagnostic yield among the four imaging tests; summary sensitivity=0.86 (95%
confidence interval 0.81–0.90); specificity=0.52 (0.36–0.67); diagnostic yield=0.58 (0.51–0.64).

Evidence on direct comparisons of alternative imaging modalities or on the impact of tests on management decisions was limited.

**Conclusions:** Nuclear imaging tests, particularly FDG-PET/CT, can be useful in identifying the source of fever in patients with classic FUO. The contribution of nuclear imaging may be limited in clinical settings where infective and neoplastic causes are less common. Studies using standardized diagnostic algorithms are needed to determine the optimal timing for testing and to assess the impact of tests on management decisions and patient-relevant outcomes.

**Key words:** FUO, FDG-PET, FDG-PET/CT, gallium scintigraphy, leukocyte scintigraphy
INTRODUCTION

Classic fever of unknown origin (FUO) is defined as fever ≥38.3°C (101°F) for three or more weeks in immunocompetent, otherwise healthy patients with no identified cause of fever after undergoing a set of obligatory investigations (1). Common causes of classic FUO include infections, tumors, and non-infectious inflammatory diseases (NIIDs) (2). With recent advances in diagnostic technologies, such as sophisticated imaging tests, improved culture techniques, and molecular diagnostics, the two most common causes (infections and neoplasms) have become less common and undiagnosed cases have become more challenging to investigate (3,4).

Gallium-67 scintigraphy was the mainstay staging modality in oncology practice (5) until the introduction of fluorine-18-fluorodeoxyglucose positron emission tomography (FDG-PET) in the 1990s. Because gallium accumulates in both malignant tumors and inflammation, gallium scintigraphy is still used as a component of work-up strategies for patients with classic FUO (2,6). Other scintigraphy methods using autologous white blood cells labeled with indium-111 or Technetium-99m are also used when infectious causes are suspected (2,6). FDG-PET is a functional imaging modality that can be used for localizing malignant tumors, as well as infectious and non-infectious inflammatory lesions, because FDG accumulates in malignant cells and activated leukocytes (6).
Several studies have assessed the diagnostic usefulness of nuclear imaging tests for localizing a source of fever in such patients. However, the interpretation of these studies is not straightforward. First, classic FUO has a broad differential diagnosis and FUO causes vary across clinical settings (3,7). Such variation, coupled with differences in pre-imaging workup algorithms across settings, can affect test performance. Second, variability in the reference standards used across studies (e.g., biopsy for malignancies, cultures for infections, and operational diagnostic criteria for autoimmune disorders) affects the cause-specific performance of tests. Third, nuclear imaging tests are often evaluated using routinely collected clinical data, without a standardized post-imaging diagnostic algorithm; in such cases the imaging results influence the selection of further (confirmatory) tests, and may introduce differential verification (8). Fourth, studies focus on a single nuclear imaging test and rarely report direct comparisons among alternative modalities. Previous systematic reviews (9-12) have also focused on single tests, and have not provided comparative information. Our review attempts to address some of these challenges by synthesizing current evidence on the diagnostic performance and clinical utility of FDG-PET alone or integrated with computed tomography (FDG-PET/CT), gallium scintigraphy, and leukocyte scintigraphy for classic FUO.
MATERIALS AND METHODS

Data Sources and Search Strategy

We searched PubMed and Scopus (from inception until October 31, 2015) with no language restrictions. We used pre-specified search terms for the target condition (e.g., “fever of unknown origin” or “FUO”) and the tests of interest (e.g., “scintigraphy”, “PET”, “PET/CT” or “SPECT”). We perused the reference lists of eligible primary papers and relevant reviews and meta-analyses. We also tracked citations to eligible papers through Scopus, Web of Science, and Google Scholar. Supplemental Table 1 provides the complete search strategy.

Study Selection

Two reviewers independently screened abstracts and examined the full text of potentially eligible papers to identify studies that evaluated FDG-PET or PET/CT using a full-ring scanner, gallium or leukocyte scintigraphy, or single-photon emission computed tomography (SPECT) for at least 10 patients with classic FUO. We only included studies that reported sufficient information to calculate sensitivity and specificity, diagnostic yield (the proportion of patients in whom the imaging results were reported to contribute to the diagnosis of FUO causes), or the proportion of patients in whom the imaging results
were deemed to have contributed to changes in diagnostic or therapeutic strategies planned before imaging. When multiple studies reported results from potentially overlapping patient groups, we only used information from the largest patient group. We excluded studies that exclusively evaluated patients 17 years or younger, or studies that included patients infected with HIV and did not report separate data on non-HIV participants.

**Data Extraction and Assessments of Risk of Bias and Applicability**

One investigator extracted descriptive information, which was confirmed by a second investigator; discrepancies were resolved by consensus. We extracted information on study design, pre-imaging tests, characteristics of enrolled patients, FUO causes, index imaging tests and diagnostic criteria ([13-16](#)), and the reference standard. We categorized studies into three groups (first-, second-, and third-level examinations) based on their pre-imaging diagnostic workup algorithm. We also classified the reported reference standards into three degrees of accuracy (high, moderate, and low accuracy). The Supplemental Material provides detailed descriptions.

Two independent reviewers assessed risk of bias and applicability for each eligible study using items based on the QUADAS-2 tool ([17](#)). We evaluated the risk of differential verification bias as
proposed elsewhere (8). Discrepancies were resolved by consensus.

**Data Synthesis**

For each study, we constructed a 2×2 contingency table consisting of true-positive, false-positive, false-negative, and true-negative results, whereby patients were categorized according to their nuclear imaging test results (positive or negative), and whether the cause of FUO was correctly identified in the imaging-positive site(s) or not (i.e., a cause was identified outside the imaging-positive site(s), or the cause remained unknown).

We estimated summary sensitivity and specificity with their corresponding 95% confidence intervals (CIs) using bivariate random-effects meta-analysis with binomial within-study likelihood when four or more studies were available for the same imaging test (18,19). When a bivariate model failed to converge, we calculated summary sensitivity and specificity separately by univariate random-effects meta-analysis using mixed-effects logistic regression (20,21). We visually assessed between-study heterogeneity by plotting study estimates in the receiver operating characteristic (ROC) space (22). We also constructed hierarchical summary ROC curves and obtained the confidence regions for the summary sensitivity and specificity (23). We performed meta-analysis of diagnostic yield using
univariate random-effects logistic regression (20). We quantified between-study heterogeneity by estimating the between-study variance. Data on diagnostic or therapeutic decision impact were qualitatively synthesized.

To explore heterogeneity, we used subgroup analyses and univariable meta-regressions, when 10 or more studies with pertinent data were available (24). Specifically, we examined study design, geographical area, clinical context, proportion of identified neoplasms and infections, and use of contrast-enhancement or not for FDG-PET/CT.

In stability analyses, we excluded studies that enrolled patients younger than 18 years and did not report separate data on adult participants. To address the heterogeneity in classification and reporting of “benign” (spontaneously regressing) causes, we recalculated sensitivity and specificity for each study considering “benign” cases as disease negative along with cases where no cause was identified (see Supplemental Material and Supplemental Table 2).

We visually assessed the results of studies directly comparing test modalities by plotting all estimates from the same study in the ROC space. We did not perform meta-analyses of comparative studies because few studies were available for each comparison. We also “indirectly” compared test performance by visually assessing summary ROC curves and by estimating relative diagnostic odds
ratios comparing alternative imaging tests (24,25). We also performed study-level univariable
meta-regressions to assess the difference in diagnostic yield between two imaging tests. The detailed
methods used for indirect comparisons are described in the Supplemental Material.

We did not perform tests for funnel plot asymmetry because they do not provide a valid way
for assessing the extent and impact of missing data (26). All analyses were conducted using Stata SE
13.1 (Stata Corp, College Station, TX, USA) and WinBUGS 1.4.3 (MRC Biostatistics Unit, Cambridge,
UK) (27). P-values were two-tailed, and statistical significance was defined as P<0.05.

RESULTS

Literature Flow and Eligible Studies

We screened 6,351 abstracts and evaluated 83 full-text articles (Fig. 1). The Supplemental Material
provides a list of excluded studies along with their reasons for exclusion. A total of 43 unique
publications (4 comparative and 38 non-comparative studies, including 2,058 unique patients) met our
eligibility criteria (Supplemental Table 3). Twenty-two studies (1,137 patients) evaluated FDG-PET/CT,
12 (522 patients) FDG-PET, 6 (397 patients) gallium scintigraphy, and 6 (153 patients)
indium-111-labelled leukocyte scintigraphy.
Study and Patient Characteristics

Studies of gallium scintigraphy and leukocyte scintigraphy had been published since the 1980s, whereas studies of FDG-PET and FDG-PET/CT were more recent, having been published after 2001 and 2008, respectively (Supplemental Table 3). Studies included a median of 48 patients (min–max: 10–162).

Typically, studies retrospectively assessed nuclear imaging tests, as second- or third-level examinations, performed after diagnostic workup that was not standardized in each study. Eight studies (19%) were prospective: one study adopted a standardized work-up algorithm (28) and three studies specified mandatory pre-imaging tests (29-31) as selection criteria, one of which routinely performed thoracoabdominal contrast-enhanced CT (31). The average pre-imaging disease duration ranged from 5 to 41 weeks. Thirteen studies (31%) reported details on the post-imaging diagnostic work-up used for each patient. Positive scan results were often verified with invasive but high accuracy reference standards, such as biopsy (median 44%, 25th–75th percentile: 38%–64%, of study participants with positive scan vs. 13%, 25th–75th percentile: 0%–20%, with negative scan); negative results were typically verified with less invasive but low accuracy reference standards, such as clinical follow-up (median 70%, 25th–75th percentile: 60%–88%). The average post-imaging follow-up ranged from 3 to 30
(median 15) months.

The average age ranged from 42 to 62 (median 54) and the proportion of patients with specific FUO etiologies varied across studies (Supplemental Table 4). The median proportion of patients with infectious or neoplastic etiologies as a final diagnosis was 48% (min–max: 23%–80%) for studies of FDG-PET/CT, 31% (min–max: 10%–80%) for FDG-PET, 42% (min–max: 26%–58%) for gallium scintigraphy, and 37% (min–max: 12%–56%) for leukocyte scintigraphy. The proportion of cases deemed to be of infectious or neoplastic etiology was not associated with pre-imaging workup algorithms (Spearman's rho=0.065; P=0.78) or publication year (Spearman's rho=0.22; P=0.15). The proportion of undiagnosed cases also varied substantially (min–max: 4%–56%).

**Test Characteristics**

The Supplemental Material describes how imaging was performed and interpreted. Studies generally adopted standard imaging protocols (13-16), and multiple nuclear medicine physicians visually interpreted the results (Supplemental Tables 5–6). Few studies reported diagnostic thresholds for quantitative assessment to support the visual assessment.
Assessment of Study Risk of Bias and Applicability

We had concerns about high risk of bias and limited generalizability in the majority of included studies for all nuclear imaging tests (Supplemental Fig. 1). Specifically, risk of bias due to differential verification (i.e., selection of reference standard tests on the basis of index test results) was a concern in all studies. Further, the nuclear imaging results were suggested as the basis for the final diagnosis of at least one case of NIIDs (e.g., vasculitis) in seven studies (32-38); such incorporation of index test results into the reference standard can lead to overestimation of test performance.

Test Performance, Diagnostic Yield, and Impact on Management

Studies of FDG-PET/CT produced heterogeneous estimates of sensitivity and specificity (Supplemental Fig. 2A). The summary sensitivity and specificity were 0.86 (0.81–0.90) and 0.52 (0.36–0.67), respectively (Fig. 2). Summary estimates were stable in subgroup analyses, and meta-regression analysis identified no covariates that significantly related with both sensitivity and specificity (data not shown).

The diagnostic yield of FDG-PET/CT also varied across studies (Supplemental Fig. 3A). Studies reporting a higher proportion of neoplasms and infections as the cause of FUO also reported a higher diagnostic yield (Spearman's rho=0.44; P=0.038) (Fig. 3A). FDG-PET/CT successfully localized
a cause of FUO in approximately 60% of patients (summary diagnostic yield=0.58; 0.51-0.64) (Fig. 4).

The summary estimates were similar across subgroups and in stability analyses; meta-regression indicated that the proportion of neoplasms and infections was positively associated with diagnostic yield (P=0.003). Non-neoplastic causes were less successfully localized; adult-onset Still's disease, tuberculosis, and polymyalgia rheumatica were the three causes that FDG-PET/CT most often showed no pathological uptakes leading to diagnosis (Supplemental Tables 7–8).

Studies of FDG-PET reported variable estimates of sensitivity and specificity (Supplemental Fig. 2B). Summary sensitivity and specificity were 0.76 (0.66–0.83) and 0.50 (0.30–0.70), respectively (Fig. 2). In meta-regression, no covariates were significantly associated with both sensitivity and specificity (data not shown). The diagnostic yield also varied across studies (Supplemental Fig. 3B), with a summary estimate of 0.44 (0.31–0.58) (Fig. 4). Again, studies with a higher diagnostic yield also reported a higher proportion of combined neoplasms and infections (Spearman's rho=0.66; P=0.020) (Fig. 3B). A positive association was also suggested by meta-regression between the prevalence of neoplasms and infections and diagnostic yield (P=0.010). Similarly to FDG-PET/CT, standalone FDG-PET frequently failed to localize non-neoplastic causes (Supplemental Tables 7–8).

Test performance of gallium scintigraphy was heterogeneous (Fig. 2C). The summary
sensitivity and specificity were 0.60 (0.45–0.73) and 0.63 (0.37–0.84), respectively (Fig. 2). Diagnostic yield ranged from 0.21 to 0.54 (Supplemental Fig. 3C), and on average, the location of a source of fever was correctly localized in approximately a third of patients (summary diagnostic yield=0.35; 0.25–0.46)(Fig. 4). Data on gallium scintigraphy were too sparse for reliable subgroup analysis.

Studies of leukocyte scintigraphy reported rather homogeneous estimates of specificity; however, estimates of sensitivity were variable (Supplemental Fig. 2D). The summary sensitivity and specificity were 0.33 (0.24–0.44) and 0.83 (0.61–0.94), respectively (Fig. 2). Estimates of diagnostic yield ranged from 0.08 to 0.31 (Supplemental Fig. 3D), and overall, the FUO cause was correctly identified based on the scan results in only a fifth of patients (summary diagnostic yield=0.20; 0.14–0.28)(Fig. 4). The summary estimates were stable in subgroup and sensitivity analyses, although data were limited.

Five studies of FDG-PET/CT (31,39-42), three studies of FDG-PET (28,29,37), and three studies of leukocyte scintigraphy (43-45) performed univariable or multivariable analyses to identify predictors of the impact of tests on diagnosis or therapeutic management (Supplemental Table 9). The three most commonly assessed predictors were c-reactive protein, leukocyte counts, and erythrocyte sedimentation rate, although studies were highly heterogeneous regarding how candidate predictors were
measured and how they were incorporated in the models (e.g., dichotomized or transformed). Overall, no predictive model was validated. Three studies (28,34,46) reported how often scans altered diagnosis and two studies (37,46) reported how often scans altered therapeutic decisions (Supplemental Table 10). For example, FDG-PET or PET/CT affected diagnostic and therapeutic management in 44% (46), and 36% of cases (37), respectively.

Comparisons of Test Performance and Diagnostic Yield among Imaging Tests

One study directly compared FDG-PET with FDG-PET/CT (47), another compared FDG-PET with gallium scintigraphy (48), and two compared FDG-PET with leukocyte scintigraphy (49,50). This limited evidence was insufficient to establish any nuclear imaging modality as superior to any other (Supplemental Fig. 4). Regarding indirect comparisons of test performance, visual assessment of the summary ROC curves and meta-regression suggested that FDG-PET/CT outperformed standalone FDG-PET, gallium scintigraphy, and leukocyte scintigraphy (Supplemental Figs. 2 and 5). Similarly, visual and quantitative indirect comparisons of diagnostic yield suggested that FDG-PET/CT was more likely to correctly identify the cause of FUO than alternative tests (Supplemental Figs. 3 and 6). Detailed descriptions are reported in the Supplemental Material.
DISCUSSION

We examined 42 studies involving nearly 2,000 patients with classic FUO to evaluate the performance and clinical validity of nuclear imaging modalities for localizing the source of fever. We found that FDG-PET/CT had the best test performance among nuclear imaging tests, and correctly pinpoints the anatomical location of FUO pathologies in approximately 60% of patients for whom basic laboratory tests and anatomical imaging have failed. Indirect comparisons of test modalities suggested that FDG-PET/CT had better performance than alternative nuclear imaging modalities. However, few studies reported direct comparison among tests, and the diagnostic yield of FDG-PET and PET/CT was lower in clinical contexts where the prevalence of neoplastic and infectious causes of FUO was low. Furthermore, the data on gallium scintigraphy and leukocyte scintigraphy were derived from older studies and thus may not be applicable to contemporary clinical practice.

Our findings about the test performance of FDG-PET and FDG-PET/CT are in general agreement with previous systematic reviews (9-12). However, we extended previous work by examining diagnostic yield and demonstrating a positive association between yield and the prevalence of infections and neoplasms. This suggests that FDG-PET and FDG-PET/CT may be less useful in clinical settings
where these causes are less common. In addition, we assessed evidence on comparative test performance and other outcomes more thoroughly than previous reviews. We found that FDG-PET/CT is the best choice and can localize most malignant or infectious lesions. A negative scan, however, does not exclude other causes of fever, such as NIIDs. Gallium scintigraphy can be used to localize malignancies when it is the only readily available modality. Leukocyte scintigraphy has a limited role in most clinical contexts, in view of its low diagnostic yield. Finally, we qualitatively summarized data on the impact of FDG-PET and FDG-PET/CT on diagnostic and therapeutic management. Lack of data, however, precluded reliable assessment on how FDG-PET or FDG-PET/CT affects estimates of the probability of particular FUO cases, or how it alters diagnostic or therapeutic management decisions. Arguably, these outcomes are more relevant to clinical practice than test performance, which suggests that more research, ideally comparative studies of alternative testing strategies incorporating nuclear imaging tests, is needed to assess clinical outcomes beyond test performance (51).

Our review has several limitations. First, our summary estimates were based on heterogeneous estimates from studies with varied designs, and we noted substantial between-study heterogeneity. We believe that variability in the prevalence of underlying FUO etiologies across studies and differences in work-up algorithms are the main sources of the observed statistical heterogeneity; it is hard to account
for these factors without access to primary study data. Second, the included studies used multiple and imperfect reference standards and were deemed likely to have produce biased results because of differential verification and incorporation of the index test result in the reference standard. Differential verification and incorporation bias are likely to lead to overestimation of test performance (52) and, thus, our summary estimates should be interpreted with caution. Third, our comparative results depend largely on indirect comparisons, which may be confounded by differences across studies that cannot be addressed analytically (25). We thus view our indirect comparisons as suggestive of possible differences among modalities that merit further research in properly designed and conducted comparative studies.

Fourth, the available literature provides only limited evidence on the impact of nuclear imaging tests on clinical management.

CONCLUSION

FDG-PET/CT, FDG-PET, gallium scintigraphy, and leukocyte scintigraphy are useful imaging modalities for localizing the source of fever in patients with classic FUO for whom a routine diagnostic workup has been unsuccessful in establishing a diagnosis. FDG-PET/CT is the most promising nuclear test with the highest diagnostic yield. However, little is known about the impact of all nuclear imaging
tests on diagnostic and therapeutic decisions. Future studies should use designs with a pre-specified
diagnostic algorithms before nuclear imaging, and standardized protocols for imaging and image
interpretation, in order to clarify which modalities are most useful. Also, given the heterogeneous test
performance across studies and the high cost of nuclear imaging, identifying factors that modify test
performance and impact on management, and using methods that combine nuclear imaging results with
other information, should be a priority.
REFERENCES


12. Sioka C, Assimakopoulos A, Fotopoulos A. The diagnostic role of (18) F


systematic reviews of diagnostic test accuracy. 


27. Thompson J. Bayesian Analysis with Stata. College Station, TX: Stata Press; 2014.


positron emission tomography/computed tomography in fever of unknown origin. 
*Jornal of Peking University (Health Sciences).* 2015;47:175-180.


Figure 1. Study Flow Diagram.

Database searches through Oct 2015
- PubMed (n = 5,519)
- Scopus (n = 2,519)

Records after duplicates removed (n = 6,351)

Records screened (n = 6,351)

Records excluded (n = 6,272)

Full-text publications retrieved for detailed evaluation (n = 79)

Reports excluded on the basis of detailed review (n = 40)
- Editorial or letter: 4
- Case reports: 2
- Review articles: 2
- Studies targeting non-FUO patients: 6
- Studies of patients 18 years: 4
- Studies including patients with HIV infection: 5
- Studies of irrelevant imaging tests: 6
- Studies with pertinent data not extractable: 5
- Not relevant: 6

Studies included in supplemental searches (n = 4)
- Google scholar: 4
- Scopus: 0
- Web of Science: 0
- References: 0

Total publications included in the meta-analysis (n = 43)
Figure 2. Meta-Analysis of Sensitivity and Specificity.

Diamonds (proportional to the number of patients) represent point estimates; extending lines represent the 95% confidence interval (CI) of each estimate.
Figure 3. Diagnostic Yield Plotted against the Prevalence of Infections and Neoplasms for Studies of FDG-PET/CT (A), FDG-PET (B), Gallium Scintigraphy (C), and Leukocyte Scintigraphy (D).

The size of each circle is proportional to the sample size for each study.
**Figure 4.** Meta-analysis of Diagnostic Yield.

Squares (proportional to the numbers of patients) represent point estimates; extending lines represent the 95% confidence interval (CI) of each estimate.