A Comprehensive Evidence-Based Approach to Fever of Unknown Origin

Ophyr Mourad, MD, FRCPC; Valerie Palda, MD, MSc; Allan S. Detsky, MD, PhD

Background: Fever of unknown origin (FUO) is defined as a temperature higher than 38.3°C on several occasions and lasting longer than 3 weeks, with a diagnosis that remains uncertain after 1 week of investigation.

Methods: A systematic review was performed to develop evidence-based recommendations for the diagnostic workup of FUO. MEDLINE database was searched (January 1966 to December 2000) to identify articles related to FUO. Articles were included if the patient population met the criteria for FUO and they addressed the natural history, prognosis, or spectrum of disease or evaluated a diagnostic test in FUO. The quality of retrieved articles was rated as “good,” “fair,” or “poor,” and sensitivity, specificity, and diagnostic yield of tests were calculated. Recommendations were made in accordance with the strength of evidence.

Results: The prevalence of FUO in hospitalized patients is reported to be 2.9%. Eleven studies indicate that the spectrum of disease includes “no diagnosis” (19%), infections (28%), inflammatory diseases (21%), and malignancies (17%). Deep vein thrombosis (3%) and temporal arteritis in the elderly (16%-17%) were important considerations. Four good natural history studies indicate that most patients with undiagnosed FUO recover spontaneously (51%-100%). One fair-quality study suggested a high specificity (99%) for the diagnosis of endocarditis in FUO by applying the Duke criteria. Two fair-quality studies showed that computed tomographic scanning of the abdomen had a diagnostic yield of 19%. Ten studies of nuclear imaging revealed that technetium was the most promising isotope, showing a high specificity (94%), albeit low sensitivity (40%-75%) (2 fair-quality studies). Two fair-quality studies showed liver biopsy to have a high diagnostic yield (14%-17%), but with risk of harm (0.009%-0.12% death). Empiric bone marrow cultures showed a low diagnostic yield of 0% to 2% (2 fair-quality articles).

Conclusions: Diagnosis of FUO may be assisted by the Duke criteria for endocarditis, computed tomographic scan of the abdomen, nuclear scanning with a technetium-based isotope, and liver biopsy (fair to good evidence). Routine bone marrow cultures are not recommended.

Arch Intern Med. 2003;163:545-551

Fever of unknown origin (FUO) identifies a syndrome of fever that does not resolve spontaneously, in which the cause remains elusive after an extensive diagnostic workup. Petersdorf and Beeson1 first coined the term fever of unknown origin in 1961 and explicitly defined it as a temperature higher than 38.3°C on several occasions and lasting longer than 3 weeks, with a diagnosis that remains uncertain after 1 week of investigations in hospital. Petersdorf and Beeson chose 3 weeks of fever to eliminate self-limited viral illnesses and to allow sufficient time for appropriate initial investigations to be completed. Over the past 40 years, health care has shifted from the inpatient to the ambulatory setting. As a result, it has now become widely accepted that the requirement for a 1-week evaluation in hospital be modified so that evaluations may now be completed in an outpatient setting.2

Fever of unknown origin is frustrating for patients and physicians because the diagnostic workup often involves numerous noninvasive and invasive procedures that sometimes fail to explain the fever. There are over 200 different reported causes of FUO.3,4 To date, there are no published guidelines or evidence-based recommendations for the diagnostic workup of FUO. The body of literature that discusses FUO comprises case series and cohort studies. In FUO, there is no diagnostic gold standard against which other diagnostic tests may be measured. Final diagnoses are determined in a number of ways, including natural history, biopsy, surgery, postmortem examinations as well as other imaging techniques. For these reasons, there is disagreement as to what should constitute a comprehensive diagnostic workup.

To have a structured, sensible, and effective approach, the clinician must have
an understanding of the spectrum of disease and the test characteristics of the various diagnostic modalities available in the evaluation of FUO. A rational approach should also be based on the relative frequencies of the different causes and their importance to the health of the patient. For the purpose of this article, FUO is not intended to encompass those individuals with impaired immunity or unexplained fevers in children. Fever of unknown origin in patients with human immunodeficiency virus infection, patients with known malignancy, and children have a different diagnostic differential and will not be addressed in this article.

SYSTEMATIC PROCESS USED TO ARRIVE AT FINAL RECOMMENDATIONS

Articles that met the selection criteria were summarized in tabular format. Criteria were developed to assess methodological quality for diagnostic studies and natural history studies based on published methods of the US Preventive Services Task Force. The evidence was systematically reviewed by assigning a quality rating to each article according to a priori criteria. While the importance of research design remains the main basis by which to assess strength of evidence, not all studies within a research design have equal internal validity. To more clearly assess the internal validity of individual studies within research designs, design-specific criteria were used that allow rating of studies into 3 internal validity categories: “good,” “fair,” and “poor.” Thus all individual studies receive 2 codes: 1 for research design and 1 (good, fair, poor) for internal validity within its design. The body of evidence available for each topic was then synthesized, and recommendations were made based on the following considerations: published prevalence of disease, performance characteristics of the test (diagnostic yield, sensitivity, specificity, positive and negative likelihood ratios), harms of the test, strength of the evidence supporting the use of the test (study design and quality rating), and harms of the diagnostic test. For example, elements likely to result in a recommendation to perform the test would be good performance characteristics and no harms, even in the presence of limited evidence, or a test with moderate performance characteristics but multiple fair-quality studies demonstrating some benefit. Tests aimed at detecting common disorders were also more likely to be recommended. Final recommendations used language defined by The Canadian Task Force on Preventive Health Care, which was amended to apply to a diagnostic test. Summary tables of the natural history, prognosis, and diagnostic studies as well as appendices that describe criteria developed for study selection and assessment of the methodological quality of diagnostic tests and natural history studies are available on request from the authors.

PREVALENCE AND SPECTRUM OF DISEASE

Iikuni et al documented that of 5245 patients admitted to the Department of Internal Medicine between 1982 and 1992 at Kitasato University Hospital, Kanagawa, Japan, 153 (2.9%) had FUO. Kazanjian reported that of 6250 infectious disease consults performed at 3 community hospitals in Rhode Island between 1984 and 1990, 86 met the criteria for FUO (ie, 1 FUO in every 73 consults requested). Because FUO encompasses a wide spectrum of both infectious and noninfectious diseases, we believe that a significant proportion of patients will be investigated by general internists, with subspecialist (eg, infectious diseases, oncology, or rheumatology) consulting thereafter.

It is important to understand the spectrum of disease before addressing the utility of the diagnostic tools in the evaluation of FUO. The causes of FUO have traditionally been grouped into 1 of 4 categories: infectious, malignant, inflammatory, and undetermined.

There are 11 series that include over 1000 patients that have reported the diagnostic entities that constitute FUO. Grouping all the patients collected from 1932 until 1994 reveals that the spectrum of disease includes infections in 28% and inflammatory diseases in 21%. Malignancies account for a smaller proportion (17%). A cause is never identified in a significant proportion (19%) of patients.

The spectrum of disease has also changed considerably from the time of the first prospectively collected series of 100 patients. Over the past 40 years, the proportion of cases of FUO caused by infections and neoplasms has decreased. The easy detection of solid tumors and abnormal lymph nodes via ultrasonography and computed tomography (CT) has resulted in the decline of tumors as a common cause of FUO, with the consequence that malignancies are less likely to present with prolonged undiagnosed fever. Patients with undiagnosed FUO used to make up the smallest proportion. At present, the largest proportion of patients who present with FUO never have a cause identified (Figure 1).

The most common infectious causes documented in the literature are tuberculosis and intra-abdominal abscesses. The most common malignancies are Hodgkin disease and non-Hodgkin lymphoma. Temporal arteritis accounts for 16% to 17% of all causes of FUO in the elderly.

Outcomes of patients with FUO is a function of the underlying cause. Overall, 12% to 35% of patients will die from FUO-related causes to 100% of patients with a final diagnosis of malignancy will die within 5 years of the diagnosis. Mortality is much lower if an infection is identified as the cause of FUO (8%-22%). Therefore, the best predictor of survival is disease cat-
category, with malignancy incurring the highest mortality. The prognosis of patients with FUO in whom a cause cannot be identified is excellent. Most of these patients have a spontaneous recovery (51%-100%), and only a small proportion have persistent fever (0%-30%).

**RESULTS**

**LIMITATIONS OF THE LITERATURE ON FUO**

The body of literature that discusses FUO comprises case or cohort studies. There are no randomized controlled trials in the FUO literature. Most of these patients were identified in tertiary care centers; however, a number of studies report their experience from community hospitals.

**LIMITATIONS OF THE LITERATURE ON DIAGNOSTIC TESTS FOR FUO**

In FUO, there is no diagnostic gold standard against which other diagnostic tests may be measured. Final diagnoses are determined in a number of ways, including natural history, biopsy, surgery, and post-mortem examinations as well as other imaging techniques. The diagnostic tests being assessed have been performed at various stages of the investigation. The definitions of true positives, false positives, true negatives, and false negatives vary from study to study. Therefore, calculation and significance of sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratios should be viewed with caution.

**INITIAL INVESTIGATIONS**

Although there is no substitute for a thorough history review and physical examination, the yield from a complete history review and meticulous physical examination is not known, since it has never been studied. Review articles and articles evaluating a diagnostic test in FUO state explicitly that a certain number of investigations must be completed for a case to qualify as FUO. These have varied over the years, and we have compiled the following list of minimum diagnostic evaluations based on reviewing all of the literature (Table 1).

One of the first steps that should be undertaken is to confirm that a true fever exists. Patients should be instructed to record and measure their temperature daily. The fever pattern adds little to the diagnostic workup. Also, all medications should, if possible, be discontinued early in the evaluation to rule out a drug-induced fever. Persistence of fever beyond 72 hours after the suspected drug has been removed allows one to conclude that the drug is not the offending agent in producing the fever.

**RECOMMENDED DIAGNOSTIC TESTS FOR WHICH EVIDENCE EXISTS**

**Abdominal CT**

A CT of the abdomen should be one of the first investigations in FUO, since it has a high diagnostic yield and is likely to identify 2 of the most common causes of FUO: intra-abdominal abscesses and lymphoproliferative disorders. A retrospective case series of an abdominal CT in the workup of FUO reported a diagnostic yield of 19%. Clinical follow-up in 32 of the 47 cases in which the CT scan of the abdomen was normal identified only 1 patient with an intra-abdominal pathologic cause (lymphoma).

**Nuclear Imaging**

Ten studies of fair methodological quality have assessed the test characteristics of nuclear imaging studies in FUO. Technetium (99m-Tc) B.W. 250/(183)–based studies report the highest specificity (93%-94%) but are insensitive (40%-75%). Indium 111 IgG and indium 111–labeled white blood cell scans have poor sensitivity (49%-82%) and a specificity that ranges from 69% to

---

Table 1. Minimal Diagnostic Workup to Qualify as Fever of Unknown Origin

<table>
<thead>
<tr>
<th>Test</th>
<th>1950s</th>
<th>1960s</th>
<th>1970s</th>
<th>1980s</th>
<th>1990s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive history</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete blood cell count + differential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood film reviewed by hematopathologist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Routine blood chemistry (including lactic dehydrogenase, bilirubin, and liver enzymes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urology and microscopy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood (&lt;3) and urine cultures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antinuclear antibodies, rheumatoid factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human immunodeficiency virus antibody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytomegalovirus IgM antibodies; heterophil antibody test (if consistent with mononucleosis-like syndrome)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q-fever serology (if exposure risk factors exist)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest radiography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis serology (if abnormal liver enzyme test result)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Figure 1. The percentage of patients with fever of unknown origin by cause over the past 40 years.
86%.26-30 Ga 76 nuclear scanning is less well studied.23,31,32 The best quality study of gallium scanning reported a sensitivity of 67% and a specificity of 78% but only included 20 patients.23 Fludeoxyglucose F 18 is a promising new alternative tracer that accumulates in both malignant tumors and at sites of inflammation. One recent small fair-quality study reported a sensitivity and specificity of 84% and 86% respectively.23 Fludeoxyglucose F 18–based scans hold promise, but further studies are required to validate its utility.

Technetium (99m-Tc BW 250/183)–based scans are therefore most likely to be diagnostically helpful (positive likelihood ratio, 5.7-12.5) because of their high specificity (93%-94%). The other tests have been shown to be either poorly discriminating (gallium 67) or inconclusive because the studies were of poor overall quality.

The only potential toxic effect related to imaging studies such as CT and nuclear studies appears to be radiation exposure. The levels of radiation involved in nuclear medicine studies are usually considerably lower than a patient would receive in a conventional radiographic study or CT scan. Owing to its minimal toxicity and overall good test characteristics, nuclear imaging studies are helpful in localizing a potential infectious or inflammatory focus. Technetium should be the tracer of choice.

The Duke Criteria

Infective endocarditis is an important cause of FUO and accounts for 1% to 5% of all cases.17,8,9,12-14 The Duke criteria have a very high specificity in patients with FUO (99%; 95% confidence interval, 97%-100%), and thus should be used to identify patients with suspected infective endocarditis.31 The design and retrospective nature of the study that assessed the utility of the Duke criteria in identifying those with infective endocarditis may have biased the results toward a higher specificity. Sensitivity data are more difficult to determine from the literature. The same authors determined that in 27 patients without FUO in whom the diagnosis of infective endocarditis was histologically and/or bacteriologically confirmed, the sensitivity of the Duke criteria was 82%.34

Liver Biopsy

The diagnostic yield from liver biopsy is 14% to 17%.35,36 Physical examination findings of hepatomegaly or abnormal liver profile are not helpful in predicting which patients will have an abnormal liver biopsy result. In patients without FUO, complications from liver biopsies are reported in 0.06% to 0.32%.37-40 Death as a direct result of the liver biopsy occurs in 0.009% to 0.12%. We believe that the benefits of a liver biopsy outweigh what we consider are minimal risks.

Temporal Artery Biopsy

There is no single large series composed solely of elderly patients with FUO. Two studies (Esposito and Glickman17 and Knockaert et al39) identified temporal arteritis as the cause of FUO in 16% and 17%, respectively.17,38 A decision analysis in the management of suspected giant cell arteritis concluded that a “biopsy and treat positive cases” is the preferred strategy when the likelihood of disease is intermediate.41 Temporal artery biopsy is a safe surgical procedure42-44 with rare complications including damage of the facial nerve,45 skin necrosis,46 and drooping of the eyebrow.47 Color duplex ultrasonography of the temporal arteries may be a helpful alternative to temporal artery biopsy in the diagnosis of temporal arteritis, with a reported sensitivity and specificity of 93% when a halo, stenosis, or occlusion is identified.48 Temporal arteritis accounts for a large proportion of causes of FUO in the elderly, and thus a temporal artery biopsy should be performed in elderly patients with unresolved FUO.

Leg Doppler Imaging

Venous thrombosis can present with prolonged fever. Three series6,10,49 reported a deep vein thrombosis as the cause of FUO in 2% to 6% of patients. Although deep vein thrombosis accounts for a small percentage of causes of FUO, leg Doppler imaging is safe and may identify a treatable cause.

Diagnostic Test for Which Evidence Exists to Recommend Against: Bone Marrow Cultures

The diagnostic yield of bone marrow cultures in immunocompetent individuals was found to be 0% to 2%.50-51 Owing to the very low diagnostic yield from bone marrow cultures in FUO, bone marrow cultures are not recommended in the diagnostic workup. Physicians must use their discretion in determining whether there are other indications to perform a bone marrow biopsy.

Areas of Uncertainty

Surgical Exploration of the Abdomen

All of the studies reporting the diagnostic yield of exploratory laparotomy in FUO are of poor methodological quality. Most of the studies were performed in the pre-CT era,52-57 whereas only 1 study examined the role of surgery in the post-CT era.58 In that study, CT of the abdomen was performed in 14 of 25 patients, and 10 had abnormal findings on CT (hepatomegaly, splenomegaly, and/or retroperitoneal nodes). The diagnostic yield in those who had a normal CT and those who did not have a CT was not reported. The mortality rate was 4%, with 12% experiencing postoperative complications.

The diagnostic yield of laparoscopy was evaluated in 1 study in the pre-CT era and determined to be 44% with no mortality and minimal morbidity reported.59 Liver biopsy was performed in 63 of 70 of these patients at laparoscopy, and it is not clear what proportion of final diagnoses were contributed to by the liver biopsy results alone. The role of surgery in the post-CT era remains unclear.

Empiric Therapy

The utility of empiric therapy, such as antibiotics, antituberculosis agents, or corticosteroids has not been studied in FUO. This, however, is not an uncommon practice.
The proposed algorithm needs to be evaluated prospectively before its validity can be ascertained. Information obtained from a thorough history review, repeated physical examinations, and initial laboratory studies may direct the physician to tests that do not conform to the algorithm. The algorithm is meant only as a framework, with necessary adjustments and provisions made according to pretest probability. The framework was derived from considerable evidence; however, one should not neglect the impact of the art of medicine and clinical experience on pretest probabilities, thus allowing for deviations from the proposed algorithm.

CONCLUSIONS

The diagnostic workup of FUO remains complex; however, considerable evidence exists to guide empiric testing. Historically, the spectrum of disease includes "no diagnosis" (19%), infection (28%), inflammatory diseases (21%), and malignancies (17%), with deep vein thrombosis (3%) and temporal arteritis in the elderly (16%-17%) being important considerations. The diagnostic workup should begin with a thorough history review and physical examination. Routine noninvasive investigations (Table 1) are recommended in all patients prior to identifying a patient as having FUO. The Duke criteria have a very high specificity (99%) in patients with FUO and suspected infective endocarditis, and thus should be used to identify endocarditis as the cause of FUO. When the initial investigations are not helpful in identifying a cause, the clinician should then proceed to imaging. These should include a CT of the abdomen and a technetium-based nuclear scan. A CT of the abdomen has a high diagnostic yield (19%) and carries a low risk. Two fair-quality studies show that technetium-based scans have a high specificity but are insensitive. Leg Doppler imaging should be considered the next step in identifying deep vein thrombosis as a potential reversible and easily treatable cause. A temporal artery biopsy should be considered in elderly patients with FUO. There is fair evidence to suggest that
The prognosis of FUO is dependent on the etiological category. Undiagnosed FUO has a very favorable outcome. Patients in whom the above diagnostic investigations fail to identify a cause should be followed clinically with serial history reviews and physical examinations until the fever resolves or new diagnostic clues are found.

Accepted for publication June 19, 2002.

No financial support was requested or rendered for this body of work.

We thank David McNelley, MD, and Hillar Vellend, MD, from the Divisions of Infectious Disease and General Internal Medicine, University Health Network, Toronto, Ontario, for reviewing the manuscript and providing valuable feedback.

Corresponding author: Ophyr Mourad, MD, Division of General Internal Medicine, St Michael’s Hospital, Room 4-140, Cardinal Carter Wing, 30 Bond St, Toronto, Ontario, Canada M5B 1W8 (e-mail: MouradO@smh.toronto.on.ca).

Table 2. Recommendations for Diagnostic Testing in FUO

<table>
<thead>
<tr>
<th>Maneuver</th>
<th>Effectiveness*</th>
<th>Level of Evidence</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Duke criteria</td>
<td>Specificity = 99%</td>
<td>Fair (1 fair-quality study)</td>
<td>The Duke criteria has a very high specificity in patients with FUO and suspected infective endocarditis and thus should be used to identify endocarditis as the cause of FUO. (Recommend)</td>
</tr>
<tr>
<td>Abdominal CT</td>
<td>Diagnostic yield = 19%</td>
<td>Fair (1 fair-quality study)</td>
<td>A CT of the abdomen has a high diagnostic yield and is likely to contribute to identifying the cause of FUO. (Recommend)</td>
</tr>
<tr>
<td>Tc 99m BW 250/183 nuclear scan</td>
<td>Specificity = 93%-94% Sensitivity = 40%-75% +ve LR = 5.7-12.5</td>
<td>Fair (2 fair-quality studies)</td>
<td>Technetium-based scans have a high specificity and poor sensitivity. Technetium is the tracer of choice in the evaluation of FUO. (Recommend)</td>
</tr>
<tr>
<td>In 111 IgG nuclear scan</td>
<td>Specificity = 69%-79% +ve LR = 2.7-3.2</td>
<td>Fair (1 fair- and 1 poor-quality study)</td>
<td>In 111 IgG-based scans have a poor sensitivity and specificity and are thus not the nuclear tracer of choice. (Recommend against)</td>
</tr>
<tr>
<td>In 111–labeled WBC scan</td>
<td>Specificity = 78%-86% Sensitivity = 45%-60%</td>
<td>Fair (1 fair- and 2 poor-quality studies)</td>
<td>In 111–labeled WBC scan is helpful in identifying suspected infectious processes. (Recommend)</td>
</tr>
<tr>
<td>Gallium 67 scan</td>
<td>Specificity = 70%-78% Sensitivity = 54%-67%</td>
<td>Fair (1 fair- and 1 poor-quality study)</td>
<td>Gallium-67–based scans have a poor sensitivity and specificity and are thus not the nuclear tracer of choice. (Recommend against)</td>
</tr>
<tr>
<td>ESR, C-reactive protein, MRI, echocardiography</td>
<td>. . . . . .</td>
<td>. . . .</td>
<td>There is no evidence to make recommendations for or against the use of ESR, C-reactive protein, MRI, bone scan, and echocardiography in the evaluation of FUO. (Insufficient evidence to recommend)</td>
</tr>
<tr>
<td>Empiric therapy</td>
<td>. . . . . .</td>
<td>. . . .</td>
<td>Empiric therapy with antibiotics, anti-TB agents, or corticosteroids should not be given to patients with FUO because they often obscure or confuse the diagnosis. (Insufficient evidence to recommend)</td>
</tr>
<tr>
<td>Liver biopsy</td>
<td>Diagnostic yield = 14%-17%</td>
<td>Fair (2 fair-quality studies)</td>
<td>Liver biopsy has a high diagnostic yield and minimal toxicity. (Recommend)</td>
</tr>
<tr>
<td>Bone marrow cultures</td>
<td>Diagnostic yield = 0%-2%</td>
<td>Fair (2 fair-quality studies)</td>
<td>Owing to the very low diagnostic yield from bone marrow cultures in FUO, bone marrow cultures are not recommended in the diagnostic workup. (Recommend against)</td>
</tr>
<tr>
<td>Laparotomy/laparoscopy</td>
<td>. . .</td>
<td>Poor (8 poor-quality studies)</td>
<td>Eight poor-quality studies revealed a high diagnostic yield in the pre-CT era. Surgical exploration of the abdomen is associated with significant morbidity and mortality. There are no studies evaluating the utility of surgical exploration in the post-CT era. (Insufficient evidence to recommend)</td>
</tr>
</tbody>
</table>

References


Abbreviations: CT, computed tomography; ESR, erythrocyte sedimentation rate; FUO, fever of unknown origin; In, indium; MRI, magnetic resonance imaging; TB, tuberculosis; Tc, technetium; WBC, white blood cell.

*Diagnostic yield is defined as number of positive tests divided by the number of tests performed; +ve LR is sensitivity/1–specificity.