METHODS

Systematic Review of the Evidence

The evidence review was conducted by the AHRQ's Southern California Evidence-based Practice Center–RAND Corporation. Additional methodological details may be found in the Appendix (available at www.annals.org) as well as in the accompanying article (8) and full report (9). Reviewers searched several databases for studies published from database inception through February 2016 and included prospective and cross-sectional studies. The study population included all adults aged 18 years or older with suspected gout (such as an acute episode of joint inflammation).

The systematic review evaluated diagnostic tests for gout. Evaluated outcomes included accuracy of the test results (sensitivity, specificity, and positive and negative predictive value); intermediate outcomes (results of laboratory and radiographic tests, such as serum urate and synovial fluid crystal analysis and radiographic or ultrasonography changes); clinical decision making (additional testing and pharmacologic or dietary management); short-term clinical (patient-centered) outcomes, such as pain and joint swelling and tenderness; and adverse effects of the tests.

Grading the Evidence and Developing Recommendations

This guideline was developed by the ACP Clinical Guidelines Committee (CGC) according to the ACP guideline development process, details of which may be found in the methods paper (10). The CGC used the evidence tables in the accompanying systematic review (8) and full report (9) when reporting the evidence and graded the recommendations by using the ACP system, which is based on the GRADE (Grading of Recommendations Assessment, Development and Evaluation) method (Table).

Peer Review

The AHRQ evidence review was sent to invited peer reviewers and posted on the AHRQ Web site for public comments. The guideline was peer reviewed through the journal and posted online for comments from ACP Regents and Governors, who represent ACP members at the regional level.

ACCURACY OF TESTS, ALGORITHMS, AND CLASSIFICATION SYSTEMS

Clinical Algorithms Incorporating Clinical Signs and Symptoms

Eleven studies assessed the sensitivity and specificity of clinical algorithms versus evaluation of synovial fluid MSU crystals for diagnosing gout (11–21) and reported widely varying results. Evaluated algorithms included the Rome criteria, New York criteria, American Rheumatology Association criteria, Janssens diagnostic rule, Clinical Gout Diagnosis criteria, monoarthritis of the first metatarsophalangeal joint, SUGAR (Study for Updated Gout Classification Criteria), and 2015 American College of Rheumatology and European League Against Rheumatism gout classification criteria.

Table. The American College of Physicians' Guideline Grading System*

Quality of Evidence	Strength of Recommendation	
	Benefits Clearly Outweigh Risks and Burden or Risks and Burden Clearly Outweigh Benefits	**Benefits Finely Balanced With Risks and Burden**
High	Strong	Weak
Moderate	Strong	Weak
Low	Strong	Weak
Insufficient evidence to determine net benefits or risks		

* Adopted from the classification developed by the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) workgroup.

Moderate-quality evidence showed that several clinical algorithms based on clinical signs and symptoms have good specificity and sensitivity (>80%) for diagnosing gout compared with assessment of synovial fluid MSU crystals. Many algorithms had a higher sensitivity in patients with a longer history of flares (>2 years) than in those with more recent-onset symptoms (≤2 years). The components of the clinical algorithms varied; however, most included the following clinical characteristics: more than 1 attack of acute arthritis, maximum inflammation developing within 1 day, redness observed over joints, painful or swollen first metatarsophalangeal joint, proven or suspected tophi, and the comorbid risk factor of hyperuricemia. Details on each algorithm are provided in the accompanying evidence review (8).

DECT

Low-quality evidence from 3 observational studies (20–22) showed that DECT had a sensitivity of 85% to 100% and specificity of 83% to 92% for predicting gout compared with assessment of synovial fluid MSU crystals or clinical algorithms.

Ultrasonography

Low-quality evidence (20, 23–27) showed that ultrasonography had a wide range in sensitivity and specificity for diagnosing acute gout compared with assessment of synovial fluid MSU crystals or clinical algorithms. Pooled sensitivity from 4 trials (12, 17, 19, 28) showed 74% sensitivity (95% CI, 52 to 88) and 88% specificity (95% CI, 68 to 96).

DIAGNOSTIC ACCURACY BASED ON JOINT SITE AND NUMBER OF JOINTS

Insufficient evidence exists regarding whether the joint site or number of joints involved affected the accuracy of diagnostic tests based on clinical signs and symptoms because no studies were identified that directly tested this association.

DIAGNOSTIC ACCURACY BASED ON SYMPTOM DURATION

Insufficient evidence exists regarding the effect of symptom duration (that is, time since beginning of flare) on the accuracy of gout diagnosis based on clinical signs and symptoms.