

Levels of Evidence For Primary Research Question¹

Study Type	Question	Level I	Level II	Level III	Level IV	Level V
Diagnostic —Investigating a diagnostic test	Is this (early detection) test worthwhile? Is this diagnostic or monitoring test accurate?	<ul style="list-style-type: none"> • Randomized controlled trial • Testing of previously developed diagnostic criteria (consecutive patients with consistently applied reference standard and blinding) 	<ul style="list-style-type: none"> • Prospective² cohort³ study • Development of diagnostic criteria (consecutive patients with consistently applied reference standard and blinding) 	<ul style="list-style-type: none"> • Retrospective⁴ cohort³ study • Case-control⁵ study • Nonconsecutive patients • No consistently applied reference standards 	<ul style="list-style-type: none"> • Case series • Poor or non-independent reference standards 	<ul style="list-style-type: none"> • Mechanism-based reasoning
Prognostic —Investigating the effect of a patient characteristic on the outcome of a disease	What is the natural history of the condition?	<ul style="list-style-type: none"> • Inception² cohort study (all patients enrolled at an early uniform point in the course of their disease) 	<ul style="list-style-type: none"> • Prospective² cohort³ study (patients enrolled at different points in their disease) • Control arm of randomized trial 	<ul style="list-style-type: none"> • Retrospective⁴ cohort³ study • Case-control⁵ study 	<ul style="list-style-type: none"> • Case series 	<ul style="list-style-type: none"> • Mechanism-based reasoning
Therapeutic — Investigating the results of a treatment	Does this treatment help? What are the harms ⁶	<ul style="list-style-type: none"> • Randomized controlled trial 	<ul style="list-style-type: none"> • Prospective² cohort³ study • Observational study with dramatic effect 	<ul style="list-style-type: none"> • Retrospective⁴ cohort³ study • Case-control⁵ study 	<ul style="list-style-type: none"> • Case series • Historically controlled study 	<ul style="list-style-type: none"> • Mechanism-based reasoning
Economic	Does the intervention offer good value for dollars spent?	Computer simulation model (Monte Carlo simulation, Markov model) with inputs derived from level I studies, lifetime time duration, outcomes expressed in dollars per quality-adjusted life year's (QALYs) and uncertainty examined using probabilistic sensitivity analyses	Computer simulation model (Monte Carlo simulation, Markov model) with inputs derived from level II studies, lifetime time duration, outcomes expressed in QALYs and uncertainty examined using probabilistic sensitivity analysis	Computer simulation model (Monte Carlo simulation, Markov model) with inputs derived from level II studies, relevant time horizon, less than lifetime, outcomes expressed in QALYs and uncertainty examined using stochastic multilevel sensitivity analysis	Decision tree over the short time horizon with inputs data for original level-II and III studies and uncertainty as examined by univariate sensitivity analysis	Decision tree over the short time horizon with inputs data informed by prior economic evaluation and uncertainty as examined by univariate sensitivity analysis

1. Level-I through IV studies may be graded downward on the basis of study quality, imprecision, indirectness, or inconsistency between studies or because the effect size is very small; these studies may be graded upward if there is a dramatic effect size. For example, a high-quality randomized controlled trial (RCT) should have >80% follow-up, blinding, and proper randomization. The Level of Evidence assigned to systematic reviews reflects the ranking of studies included in the review (i.e., a systematic review of Level-II studies is Level II). A complete assessment of the quality of individual studies requires critical appraisal of all aspects of study design.

2. Investigators formulated the study question before the first patient was enrolled.

3. In these studies, “cohort” refers to a nonrandomized comparative study. For therapeutic studies, patients treated one way (e.g., cemented hip prosthesis) are compared with those treated differently (e.g., cementless hip prosthesis).

4. Investigators formulated the study question after the first patient was enrolled.

5. Patients identified for the study on the basis of their outcome (e.g., failed total hip arthroplasty), called “cases,” are compared with those who did not have the outcome (e.g., successful total hip arthroplasty), called “controls.”

6. Sufficient numbers are required to rule out a common harm (affects >20% of participants). For long-term harms, follow-up duration must be sufficient.