

# Hierarchy of Evidence Within the Medical Literature

Sowdhamini S. Wallace, DO, MS,<sup>a,b</sup> Gal Barak, MD, MEd,<sup>a,b</sup> Grace Truong, MD,<sup>b</sup> Michelle W. Parker, MD<sup>c</sup>

## ABSTRACT

The quality of evidence from medical research is partially deemed by the hierarchy of study designs. On the lowest level, the hierarchy of study designs begins with animal and translational studies and expert opinion, and then ascends to descriptive case reports or case series, followed by analytic observational designs such as cohort studies, then randomized controlled trials, and finally systematic reviews and meta-analyses as the highest quality evidence. This hierarchy of evidence in the medical literature is a foundational concept for pediatric hospitalists, given its relevance to key steps of evidence-based practice, including efficient literature searches and prioritization of the highest-quality designs for critical appraisal, to address clinical questions. Consideration of the hierarchy of evidence can also aid researchers in designing new studies by helping them determine the next level of evidence needed to improve upon the quality of currently available evidence. Although the concept of the hierarchy of evidence should be taken into consideration for clinical and research purposes, it is important to put this into context of individual study limitations through meticulous critical appraisal of individual articles.



<sup>a</sup>Division of Pediatric Hospital Medicine, <sup>b</sup>Department of Pediatrics, Baylor College of Medicine, Houston, Texas; and <sup>c</sup>Department of Pediatrics, Division of Hospital Medicine, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, Ohio

[www.hospitalpediatrics.org](http://www.hospitalpediatrics.org)

DOI:<https://doi.org/10.1542/hpeds.2022-006690>

Copyright © 2022 by the American Academy of Pediatrics

Address correspondence to Sowdhamini S. Wallace, DO, MS, 1102 Bates St, Houston, TX 77030. E-mail: [sswallac@texaschildrens.org](mailto:sswallac@texaschildrens.org)

HOSPITAL PEDIATRICS (ISSN Numbers: Print, 2154-1663; Online, 2154-1671).

**FUNDING:** No external funding.

**CONFLICT OF INTEREST DISCLAIMER:** The authors have indicated they have no conflicts relevant to this article to disclose.

**COMPANION PAPER:** A companion to this article can be found online at [www.hosppeds.org/cgi/doi/10.1542/hpeds.2022-006766](http://www.hosppeds.org/cgi/doi/10.1542/hpeds.2022-006766).

Dr Wallace conceptualized the article, drafted portions of the article including the table and supplemental figure, and compiled the initial and final drafts; Drs Barak and Truong conceptualized and wrote sections of the article, and critically edited the first draft; Dr Parker conceptualized and wrote sections of the article, created the figure, and critically edited the first draft; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

To promote evidence-based practice in the care of hospitalized children, pediatric hospitalists need to understand the hierarchy of evidence in the medical literature and the potential biases and limitations inherent in certain study designs. The steps in evidence-based practice begin with asking a clinical question and acquiring the evidence. When performing a literature search, one can align articles found with their original clinical question to deem whether the article may provide useful evidence. In other words, hospitalists may ponder whether the article addresses the intended population and measures important outcomes. The type of clinical question being asked, whether a question of diagnosis, treatment, or prognosis, in conjunction with knowledge of the evidence hierarchy can inform hospitalists on the ideal study designs to search for in the literature and improve the efficiency of searches through the use of filters for study design. For instance, one may prioritize searching for randomized controlled trials (RCTs) or systematic reviews (SRs) with meta-analyses (MAs) for treatment questions, but if not found, may use a rigorously performed observational study to answer the question.

To appraise the evidence, one must also understand common biases and limitations of certain study designs. For the remainder of this article, we will describe study designs within the medical literature alongside illustrative examples. Strengths and limitations of common study designs can be found in Table 1.

## STUDY DESIGNS

Before delving more deeply into study designs involving human subjects, we would like to note the importance of basic and translational research studies. These studies many times serve as the foundation for clinical research and can help drive the formulation of hypotheses and ascertain the safety of certain drugs and devices before further study in humans. A detailed review of these studies is out of the scope of this article but can

be found in a previous issue of *Hospital Pediatrics*.<sup>1</sup>

As an overview for clinical research studies, we have included an organizational chart for clinical research study designs involving human subjects (Supplemental Fig 1). In the subsequent sections of this article, we will describe different study designs in the order of how they ascend the evidence hierarchy (Fig 1).

## OBSERVATIONAL STUDIES

In observational studies, current practices are evaluated, and the researcher neither intervenes in nor changes the care being provided. Case series and case reports have no comparison group, whereas analytic studies, such as cross-sectional, case-control, and cohort studies, include a comparison group.

## CASE REPORTS AND CASE SERIES

Case reports and case series describe a clinical event such as a diagnosis or treatment. Typically, case reports describe 4 or fewer patients, whereas case series describe >4 patients.<sup>2</sup> In the midst of the coronavirus disease 2019 pandemic, Webb and Osburn published a case series on hospitalized patients with severe acute respiratory syndrome coronavirus 2 infection.<sup>3</sup> These types of studies allow clinicians to share their experiences in caring for patients with rare diagnoses or unusual presentations. However, drawing conclusions from these study types to answer clinical questions is not advised because of potential biases resulting from the small number of patients described and the lack of comparison groups. These studies can help generate hypotheses for subsequent studies.

## CROSS-SECTIONAL STUDIES

Cross-sectional studies are those in which population data are collected at a specific point in time, with simultaneous measurement of the exposure and outcome. This study design is often used for survey-based studies such as the one by Rassbach and Fiore in 2021 in which fellows were surveyed on research and career outcomes after graduation.<sup>4</sup> This

study found that 77% of fellowship graduates continued to conduct research after graduation, and that those who graduated with a master's degree were significantly more likely to continue research after fellowship (91% vs 64%,  $P = .0001$ ).

## CASE-CONTROL STUDIES

In case-control studies, 2 existing groups with or without a disease are compared to identify disease risk factors. Moss et al used this method to identify risk factors that contribute to hospital-acquired venous thromboembolism in adults who are admitted to pediatric hospitals.<sup>5</sup> This type of study is typically quick and economical because there is no need to wait for a disease to develop; this also makes it useful for studying rare diseases as in this example of venous thromboembolism.

## COHORT STUDIES

Retrospective and prospective cohort studies identify a specific patient population in which a subset of individuals has experienced a particular exposure, and compare the rates of disease development in exposed to unexposed individuals over time. In 2019, Desai et al retrospectively evaluated a cohort of infants with bacteremic urinary tract infections to determine the association between duration of intravenous antibiotics and rates of urinary tract infection recurrence resulting in presentation to the emergency department or hospitalization within 30 days of hospital discharge.<sup>6</sup>

## RANDOMIZED CONTROLLED TRIALS

RCTs are prospective studies that evaluate the effectiveness of interventions through comparison with placebo or the existing standard of care. Eligible participants are enrolled and randomly allocated to a treatment arm or comparison arm. Multicenter studies may use cluster randomization of entire groups of participants from hospitals or study sites to 1 arm. In some RCTs, participants and/or study personnel may be blinded to the assigned group. One recent investigator-

**TABLE 1** Strengths and Limitations of Common Study Designs

Study Design	Strengths	Limitations
Cross-sectional (retrospective or prospective)	<ul style="list-style-type: none"> <li>• Low cost and time</li> <li>• Suitable design for diagnostic accuracy studies because investigational test and reference standard obtained at same point in time</li> <li>• Can report prevalence (%) of diseases or outcomes</li> </ul>	<ul style="list-style-type: none"> <li>• Cannot ascertain causality since exposure and outcome occur at 1 point in time</li> <li>• Survey-based research can be prone to missing data or nonresponse bias</li> </ul>
Case-control (retrospective)	<ul style="list-style-type: none"> <li>• Efficient and takes less time to perform</li> <li>• Feasible design for rare outcomes or diseases</li> <li>• Matching of cases and controls by certain research participant characteristics is often used to address confounding factors</li> </ul>	<ul style="list-style-type: none"> <li>• Selection bias can occur if control group is selected from a different population than cases</li> <li>• Recall bias may occur if cases are more likely to remember an exposure than controls or vice versa</li> <li>• Results are limited to odds ratios. Risk difference cannot be measured.</li> </ul>
Cohort (retrospective or prospective)	<ul style="list-style-type: none"> <li>• Feasible design for rare exposures because groups defined by exposure status</li> <li>• Good design when the outcome is common</li> <li>• Multiple outcomes can be measured</li> <li>• Able to ascertain time course of exposure and outcome to build evidence toward causality</li> <li>• Can report incidence and relative risk, absolute risk reduction, and number needed to treat</li> </ul>	<ul style="list-style-type: none"> <li>• Prone to confounding bias given observational design. Measured confounders can be addressed statistically but unmeasured confounders may exist.</li> <li>• Prospective cohorts may be time-consuming and expensive</li> <li>• Loss to follow-up could lead to attrition bias</li> <li>• Surveillance bias can occur because diseases are more likely to be identified than in the general population because of increased monitoring.</li> </ul>
RCTs (prospective only)	<ul style="list-style-type: none"> <li>• Randomization can lead to equal distribution of confounding factors in each group</li> <li>• Allocation concealment can help prevent the research team from knowing the next potential participant's assignment and reduce selection bias</li> <li>• Blinding can help address observation bias</li> </ul>	<ul style="list-style-type: none"> <li>• Resource-intensive and time-consuming to conduct</li> <li>• Limits to generalizability if low consent rate or numerous exclusion criteria</li> <li>• Attrition bias may occur if loss to follow-up or drop-out rates are unbalanced between arms</li> </ul>
SRs and MAs (retrospective of existing evidence)	<ul style="list-style-type: none"> <li>• Rigorous literature search of published and unpublished literature</li> <li>• Appraisal of each included article for bias</li> <li>• Quality assessment of the level of evidence for each outcome can help put findings in context of the quality of evidence it is based on</li> <li>• Makes readers aware of current state of evidence and areas for needed studies</li> </ul>	<ul style="list-style-type: none"> <li>• Studies with biased results may lead to inaccurate qualitative or quantitatively synthesis of results</li> <li>• Quantitative synthesis of results from heterogeneous studies may lead to biased results</li> </ul>

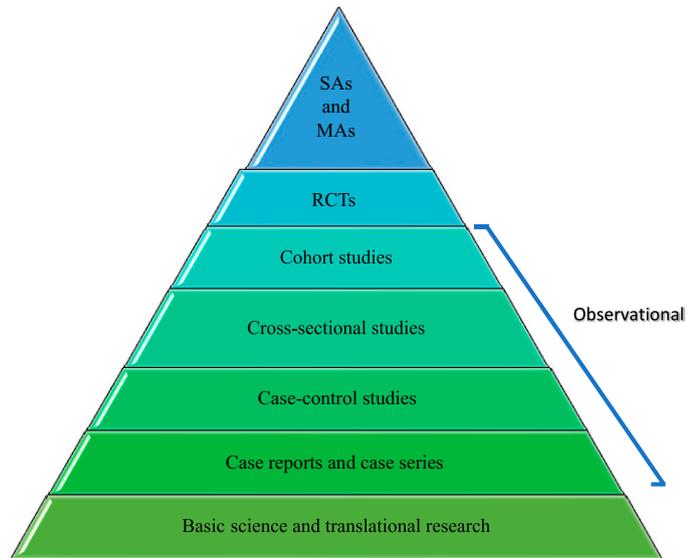
blinded RCT evaluated the effect of interactive versus didactic asthma education on emergency department visits and hospitalizations.<sup>7</sup> An accompanying commentary highlighted the need for more RCTs in the pediatric hospital medicine literature given their relative paucity.<sup>8</sup>

The RCT is considered the best study design to evaluate comparative interventions, and the only study type able to establish causation as opposed to association. Although observational studies are limited by the investigators' abilities to adjust for unmeasured participant characteristics, a well-conducted RCT provides a way to reduce selection and

confounding biases by leveraging the random allocation of participants to balance both observed and unobserved participant characteristics. Blinding is an additional strategy that has long been believed to improve the validity of a trial and 1 that is only possible in a prospective study. Additionally, in RCTs, the research team has greater control over the intervention under evaluation with regard to variables such as exposure amount and timing. The summative result of these study design strengths is a study with higher internal validity than can often be achieved with observational study methods.

## SYSTEMATIC REVIEWS AND META-ANALYSES

SRs and MAs are considered to be the highest-quality study designs when conducted in a rigorous manner. The goal of SRs is to produce an answer to a clinical question on the basis of the synthesis of available evidence evaluated from multiple distinct studies. SRs can focus on any type of research question, whether related to diagnosis, treatment, prognosis, education, or quality improvement. It is important to note that SRs differ from narrative reviews, which seek to describe previously published works on a specific topic but usually



**FIGURE 1** Strength of evidence pyramid.

without systematic methods to determine study eligibility, perform a comprehensive search strategy, and appraise studies for bias. This can potentially make narrative reviews susceptible to selection bias through omission of key studies or bias from emphasis of results produced by flawed methods.

MAs are 1 component of the synthesis phase of SRs. This quantitative statistical process combines numerical outcome data from multiple studies as if they were the result of a single larger study. The methods of SRs/MAs involve formulation of a clinical question, determination of study inclusion and exclusion criteria, a comprehensive literature search of published and unpublished studies, appraisal of bias in studies, determination of the quality of evidence for each outcome studied, and qualitative and/or quantitative synthesis of data. The quality of evidence for each outcome across studies is often determined using the Grading of Recommendations Assessment, Development and Evaluation criteria, which evaluates the initial quality according to the study design, and downgrades or upgrades the evidence level on the basis of several factors. Studies can be downgraded for risk of bias, inconsistency, indirectness, imprecision, or publication

bias, and upgraded for large-effect or dose response.<sup>9</sup>

SRs/MAs potentially offer the highest-quality evidence because of the rigorous methods and broad utilization of study results, regardless of study size or publication status. Unfortunately, when the quality of included studies is low, further studies are often needed to address the clinical question. Furthermore, heterogeneity, either clinical or statistical, may plague SRs/MAs if not identified and addressed by researchers. Outcomes for studies with clinical heterogeneity, meaning differing patient populations, interventions, or outcomes, should not be quantitatively synthesized. Statistical heterogeneity, symbolized by  $I^2$ , measures whether the variability between study results is because of chance. Excessive heterogeneity should alert researchers to potential differences in design, population, or attributes of intervention between studies that is not because of chance. For example, a reanalysis of 2 previous MAs regarding the association between hypertonic saline and hospital length of stay in bronchiolitis patients was able to discern significant heterogeneity that, when accounting for those factors causing heterogeneity, came to a different conclusion than the original

studies.<sup>10</sup> These limitations and sources of error are important to keep in mind when creating or reviewing any SRs/MAs. Further, although out of the scope of this article, hospitalists may come across newer methods of evidence synthesis in the literature, such as individual patient data MAs and network MAs, and one may refer to these referenced resources to understand their unique advantages and limitations.<sup>11,12</sup>

### Diagnostic Testing

There are important nuances of diagnostic testing articles that require consideration. Diagnostic testing research can take the form of various study designs depending on the stage of the research. Initial studies may be descriptive, where diagnostic test results are compared in patients with and without the disease to show that there is a difference in the result. Subsequent studies of diagnostic accuracy are commonly cross-sectional, where the investigational test and a reference standard are employed at 1 point in time to determine sensitivity, specificity, predictive values, and, ultimately, how well the test can distinguish between those with and without disease. Once tests are determined to be accurate, RCTs can

be conducted to determine patient outcomes in those receiving the test versus a different test or no test at all. SRs/MAs can also provide a high level of evidence to answer diagnostic questions on the basis of a synthesis of available literature.

## CONCLUSIONS

The evidence hierarchy should be considered while performing literature searches and to prioritize articles for critical appraisal during a busy clinical service. The ultimate quality of the evidence, however, ultimately depends on evaluation of the strengths and limitations of a particular study during the critical appraisal process. Various free tools exist on the Internet to aid clinicians in critical appraisal sorted by study design as offered through Critical Appraisal Skills Program and Centre for Evidence-Based Medicine.<sup>13,14</sup> Knowledge of the existing evidence base, the evidence hierarchy, and strengths and limitations of study designs should also be considered by researchers seeking to design studies to build upon current evidence for particular questions.

## Take-Home Points

- The evidence hierarchy of study designs can be used to guide efficient literature searches and article appraisals for clinical questions.
- The availability of only lower-level observational study designs for clinical questions may prompt researchers to consider designing higher-level experimental designs if feasible and ethical to do so.

- Each study design inherently has strengths and weaknesses that should be kept in mind during critical appraisal of research articles and design of novel studies.

## REFERENCES:

1. Forster CS, Tang Girdwood S, Morrison JM, Ambroggio L. Changing the paradigm of research in pediatric hospital medicine. *Hosp Pediatr*. 2019;9(9):732–735
2. Abu-Zidan FM, Abbas AK, Hefny AF. Clinical “case series”: a concept analysis. *Afr Health Sci*. 2012;12(4):557–562
3. Webb NE, Osburn TS. Characteristics of hospitalized children positive for SARS-CoV-2: experience of a large center. *Hosp Pediatr*. 2021;11(8):e133–e141
4. Rassbach CE, Fiore D. Council of Pediatric Hospital Medicine Fellowship Directors. Research and career outcomes for pediatric hospital medicine fellowship graduates. *Hosp Pediatr*. 2021;11(10):1082–1114
5. Moss SR, Jenkins AM, Caldwell AK, et al. Risk factors for the development of hospital-associated venous thromboembolism in adult patients admitted to a children’s hospital. *Hosp Pediatr*. 2020;10(2):166–172
6. Desai S, Aronson PL, Shabanova V, et al. Parenteral antibiotic therapy duration in young infants with bacteremic urinary tract infections. *Pediatrics*. 2019;144(3):e20183844
7. Samady W, Rodriguez VA, Gupta R, Palac H, Pongracic JA, Press VG. Interactive inpatient asthma education: a randomized controlled trial. [Published online ahead of print February 22, 2022] *Hosp Pediatr*. 2022:e2021006259
8. Kaiser SV, Schroeder AR, Coon ER. Pediatric hospital medicine needs more randomized controlled trials. [Published online ahead of print February 22, 2022] *Hosp Pediatr*. 2022:e2021006429
9. Guyatt GH, Oxman AD, Vist GE, et al. GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–926
10. Brooks CG, Harrison WN, Ralston SL. Association between hypertonic saline and hospital length of stay in acute viral bronchiolitis: a reanalysis of 2 meta-analyses. *JAMA Pediatr*. 2016;170(6):577–584
11. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221
12. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. *BMJ*. 2013;346:f2914
13. Critical Appraisal Skills Programme. CASP checklists. Available at: [casp-uk.net/referencing/](http://casp-uk.net/referencing/). Accessed May 3, 2022
14. University of Oxford. Centre for evidence-based medicine critical appraisal tools. Available at: [ceb.m.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools](http://ceb.m.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools). Accessed May 3, 2022