Back in 1974, in an effort to engage residents in research, senior residents were asked to identify the clinical activity they found most frustrating. They responded, “Having to do complete sepsis work-ups on young infants just because they are febrile. Practitioners don’t do that, and we can tell when infants are septic or not.” This prompted a study to determine how well the clinical judgment of senior residents, supplemented by complete blood cell counts and cerebrospinal fluid (CSF) analyses, performed in the evaluation of febrile infants 0 to 8 weeks of age. It was the heyday of group B streptococcal infections, as reflected in the bacteremia rate of 14.8% in the 62 consecutively enrolled infants studied. Infants classified as “ill” or “can’t tell” had rates of 18.5% and 18.8%, respectively, whereas the rate in infants classified as “well” was 5.6%, 1 of 18 infants. The study was small, to be sure, but that was not the major impediment to publication. Journal peer reviewers could not understand why such a study was performed because it was common knowledge that clinical judgment did not apply to very young febrile infants. Other similar studies followed along with the recognition of iatrogenic complications associated with hospitalizing these infants. By the mid-1980s, there was sufficient confirmation that, although well-appearing febrile infants had a lower rate of bacteremia than those who were ill-appearing, the rate was not 0, and infants with bacteremia could not be distinguished with complete certainty from those without bacteremia. Then Keith Powell and his colleagues at Rochester proposed an alternative strategy: instead of focusing on who had bacteremia, focus on those who did not and create criteria for a low-risk category that could permit infants to avoid unnecessary (over)treatment. Thus, the Rochester criteria were born, followed by criteria from Milwaukee, Philadelphia, and Boston, all derived from encounters in hospital emergency departments.

The observation that practitioners did not do what was routine in hospital emergency departments was confirmed in surveys based on self-report. But it was not until 2004 that what practitioners actually did was documented in a study conducted by the American Academy of Pediatrics (AAP) Pediatric Research in Office Settings network. During the subsequent decade, it became clear that the practice of “do everything” was changing even in pediatric emergency departments. But guidance about when and how to “safely do less” was lacking. The AAP clinical practice guideline (CPG) “Evaluation and Management of Well-Appearing Febrile Infants 8 to 60 Days Old” was developed to provide such guidance, based on best available evidence. In this article, we address the process of developing the guideline.

STEP 1: CONVENE A (SUB)COMMITTEE
A subcommittee of the AAP Council on Quality Improvement and Patient Safety was assembled, composed of knowledgeable individuals with broad representation (eg, general pediatrics, infectious diseases, pediatric emergency medicine, general emergency medicine, and hospital medicine) and, as became apparent, differing viewpoints. Individuals with particular expertise in epidemiology and algorithm development were included. The subcommittee identified specific questions to be addressed in a guideline.
STEP 2: ASSEMBLE THE EVIDENCE
Gathering the Evidence

To facilitate the process, the AAP submitted a proposal to the Agency for Health Research and Quality for a systematic review. The agency assigned the review to one of its evidence-based practice centers. Members of the subcommittee identified key articles from their own files that needed to be added, and additional searches were required to address specific questions. To keep up with the rapidly expanding literature, the epidemiologist for the project conducted searches in 2012, 2013, 2016, 2019, and 2020. In addition, subcommittee members alerted him to new studies, including some not yet published.

Compiling the Evidence

As evidence was compiled from various sources, it became clear that the use of process and outcome measures established a priori limited the ability of studies to answer specific questions. Two related examples are the use of serious bacterial infections (SBIs) as an outcome measure and age groupings by month. SBIs include urinary tract infections (UTIs), bacteremia, and meningitis, but the prevalence of UTI overwhelms the prevalence of the other 2 infections, and although it helps researchers by reducing the number of participants needed in a study, it obfuscates what is happening with bacteremia and meningitis. (The CPG authors strongly urge that the use of the term “SBI” be abandoned.) UTI and meningitis can be suspected on the basis of analysis of urine and CSF, respectively, but bacteremia is more problematic: it requires incubation of a blood specimen, which causes a delay in diagnosis. A concerted attempt was made to establish the prevalence of the 3 infections separately. Determining the rate of meningitis in well-appearing infants in various studies required noting how the prevalence was computed, that is, whether the denominator was the number of CSF specimens submitted for culture or the number of febrile infants, including those not deemed ill enough to undergo lumbar puncture. The data in Table 1 from Northern California Kaiser are derived not only from hospitalizations and emergency department visits but also from office visits. If only the CSFs submitted for culture are considered, the rates of meningitis are 1 of 203 (0.5%) in 7- to 28-day-olds and 2 of 186 (1.2%) in 29- to 60-day-olds. If, however, febrile infants who did not undergo lumbar puncture are included, the rates fall to 1 of 330 (0.3%) and 2 of 625 (0.3%) in the 2 age groupings, changing the number needed to test considerably.

Table 1: Results of CSF Cultures Obtained (and Not Obtained) From Febrile Infants, Northern California Kaiser

<table>
<thead>
<tr>
<th>Age, d</th>
<th>Meningitis, n</th>
<th>Negative Result, n</th>
<th>Not Tested, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7–28</td>
<td>1</td>
<td>202</td>
<td>127 (38.5)</td>
</tr>
<tr>
<td>29–60</td>
<td>2</td>
<td>164</td>
<td>459 (74.5)</td>
</tr>
<tr>
<td>61–90</td>
<td>1</td>
<td>25</td>
<td>399 (93.9)</td>
</tr>
</tbody>
</table>

Refining the Evidence

Determining the prevalence of bacteremia by age required refining the available evidence. Drawing on the experience of the AAP UTI guideline,16 researchers were contacted and requested to provide data from their studies in a more granular form than presented in the publications. As occurred in the development of the UTI guideline, all of the researchers agreed and shared their data. The refined data permitted recognition that the prevalence of bacteremia decreases from weeks 2 to 3 to week 4 and decreases further in weeks 5 to 8. When the results from 4 large studies involving thousands of febrile infants are combined, the 95% confidence intervals for the 3 groupings do not overlap (Fig 1). This provided the basis for considering 3 subgroups among the 8- to 60-day-olds: 8 to 21 days, 22 to 28 days, and 29 to 60 days. (Gaps for which refining existing evidence was insufficient to answer key questions formed an important section of the CPG entitled “Areas for Research,” a blueprint to inform the next round of studies.)

STEP 3: CONVERT THE EVIDENCE

This step reflects the maturation of the AAP’s CPG development process. Early on, it was acceptable to issue recommendations, including verbs such as “consider.” To make recommendations both actionable and measurable, they are now in the form of key action statements (KASs). The process of converting evidence into KASs was performed as transparently as possible, explicitly considering and reporting the following: benefits; risks, costs, or harms; the ratio of the two; value judgments; the role of patient preferences; exclusions; and intentional vagueness. This would seem like a straightforward task because it is expected that if evidence is of high quality, it should “speak for itself.” As noted by Djulbegovic and Guyton,17 however, evidence “never speaks for itself … it always requires interpretation” and, inevitably, value judgments. Put another way, evidence is numbers, not decisions (see Step 4). To help the subcommittee ensure clarity as well as to assist clinicians in applying the CPG, an algorithm was developed for each age group. Reflecting the difficulty in reducing the complexity of clinical decision-making and different interpretations of the evidence into KASs, 189 revisions of the originally proposed algorithms were performed before the final version was accepted. Algorithms are deceptively black and white, so footnotes are included with each algorithm to provide some nuance but still cannot include all the information regarding the infant and family that clinicians use in decision-making.

STEP 4: GRADE THE KASS

Grading the KASs involves assessing the quality of the evidence and assigning a strength of recommendation to each according to the balance of benefits and harms (Fig 2). The problem comes, as noted above, in the interpretation of the evidence. For example, benefits and risks are rarely expressed in the same units, so how can they be balanced without value judgments? Consider well-appearing

infants and the possibility of bacterial meningitis. The benefit of identifying the 1 infant with occult meningitis is clear; it is also clear that this requires hundreds of lumbar punctures in well-appearing infants. At what point does the benefit no longer justify the number of lumbar punctures? 100? 250? 500? 1000? Subcommittee members struggled with this reflecting different levels of risk tolerance. How numbers are interpreted may also depend on one’s experience and subspecialty. Infectious diseases specialists may read Table 1 from left to right, noting the number of infants with meningitis, whereas primary care pediatricians may read right to left, noting the large percentage of infants who did not undergo a lumbar puncture, even in the first month of life. \(^{18}\) When consensus was not achieved within the subcommittee, additional members were added; discussions continued until consensus was achieved and the words “may” or “need not” were used rather than “should” to allow for variation according to specific details of an infant’s presentation, differences in risk tolerance, and shared decision-making with parents.

**STEP 5: SUBMIT THE DRAFT FOR CRITICAL REVIEW**

When members of the subcommittee were able to agree on the wording of KASs, a draft was circulated to various stakeholders, including 8 AAP committees, an AAP council, 6 AAP sections, 3 other AAP groups, and 3 external groups. Comments and concerns were solicited, and ~1000 were submitted. The subcommittee responded to each and received approval from the AAP that the concerns were addressed satisfactorily.

**STEP 6: REVISE AND SUBMIT FOR ADDITIONAL REVIEW**

With multiple changes prompted by the concerns and comments of the reviewers, a new draft was prepared, submitted to an AAP copyeditor, and then reviewed by an AAP senior vice president, who provided comments. The document was then revised further and submitted to the Board of Directors for approval.

**STEP 7: DISSEMINATE THE CPG**

Once approved by the AAP Board of Directors, the CPG was ready for publication. Plans to disseminate the CPG beyond publication in *Pediatrics* included presentations in various settings, the first of which was at the Pediatric Academic Societies meeting in May 2021.

**CONCLUSIONS**

The new AAP CPG is based on the best current evidence. However, the evidence was not always interpreted the same way by subcommittee members representing different perspectives. Wording changes were negotiated to accommodate the various stakeholders. It is expected (and hoped) that the CPG will be applied thoughtfully. None of the KASs uses the word “must” or the words “must not,” and the CPG begins as follows: “The guidance in this report does not indicate an exclusive course of action or serve as a standard of medical care. Variations, taking into account individual circumstances, may be appropriate.” “Individual circumstances” has many components, notably including the values and preferences of parents. The process of developing the guideline was a lengthy one, but to put the process in perspective, it only took 2 years longer than the initiation, development, and implementation of another key guideline, the Constitution of the United States (Fig 3).


Development of the New AAP Febrile Infant Clinical Practice Guideline
Kenneth B Roberts and Robert H. Pantell
Hospital Pediatrics 2021;11;1028
DOI: 10.1542/hpeds.2021-006201 originally published online August 5, 2021;
Development of the New AAP Febrile Infant Clinical Practice Guideline
Kenneth B Roberts and Robert H. Pantell
Hospital Pediatrics 2021;11;1028
DOI: 10.1542/hpeds.2021-006201 originally published online August 5, 2021;

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://hosppeds.aappublications.org/content/11/9/1028