The concept of screening, and its underlying complexities, is still poorly understood in many circles. We tell students that there is no free lunch, that if you want to use a screen to identify everyone, or almost everyone, who might have a given abnormality, then you will have to accept that you will falsely identify many individuals who later turn out to be normal. In the jargon of decision analysis theory, high hit rate means high false-alarm rate, and low false-alarm rate means low hit rate. And both numbers are critically dependent on the pass/fail criterion. Nothing illustrates these concepts better than the data provided in the paper “Transient Evoked Otoacoustic Emissions in Adults: A Comparison between Two Test Protocols” in this issue of JAAA. Authors Joseph Kei, Ravi Sockalingam, Clive Holloway, Alan Agyik, Craig Brinin, and Doreen Baine of the University of Queensland, Brisbane, Australia, set out to compare two TEOAE screening protocols in 115 adult ears. The two procedures, offered on a commercially available device, differed only in the time window over which a response is measured and the rate at which clicks are presented. The gold standard was the pure-tone audiometric threshold at each test frequency. The failure criterion was a threshold in excess of 15 dBHL. The authors asked how successfully the TEOAE signal-to-noise ratio (SNR) predicted the presence of such a loss, as a function of the SNR pass/fail criterion. Basic findings were straightforward. The protocol employing a shorter time window and a higher click rate (Quickscreen) yielded slightly better results than the protocol employing a longer time window and slower click rate (Default).

Of particular interest to audiologists engaged in any type of screening, however, are the data that the authors have carefully cataloged in the appendices. Here are found the raw data underlying their clinical decision analysis: hit rate, false-alarm rate, efficiency, predictive value of a positive result, and predictive value of a negative result. They are presented systematically for virtually every possible pass/fail criterion.

As an example, consider the data in Appendix 3. Here are listed all of the results for predicting a loss of more than 15 dB at 4000 Hz. In the upper section (Quickscreen) the first column lists possible TEOAE pass/fail criteria for SNRs ranging from -3 dB to 25 dB. The remaining columns list the various indices of screening outcomes based on each SNR criterion. If we enter the row for SNR = 3 dB (a commonly used pass/fail screening criterion), we see that the hit rate (the percentage of ears correctly identified as having a loss greater than 15 dB) is 55.6 percent. The false-alarm rate (the percentage of ears incorrectly identified as having a loss greater than 15 dB) is 11.3 percent. Not very impressive performance since, at this SNR criterion, almost half of the people we are seeking to identify are missed. But the false-alarm rate is not too bad. Only slightly more than 1 in 10 ears is incorrectly identified as having a loss. We can improve the hit rate by simply changing the TEOAE SNR criterion. Let us try an SNR of 19 dB. Now the hit rate is much better, 88.9 percent. We have only missed about 1 in 10 persons with an actual loss. But at what price? The false-alarm rate is now 85.8 percent. Here we have incorrectly identified almost as high a percentage of ears as we have correctly identified. As you study the HR and FA columns, you will see vividly how hit rate and false-alarm rate move up and down together like reluctant mates. The principle is dramatically illustrated, not only in theory but, here, in actual real-world data.

But there are more columns to explore. Consider efficiency. This refers to the overall accuracy
of the results. It is defined as the percentage of all decisions that are correct, whether passes or fails. The figure of 86.1 percent, for the SNR = 3 dB, looks pretty good, but that is because the ratio of ears without actual losses (true passes) to ears with actual losses (true fails) in this study was large (as it usually is in real-life screening situations). Here we encounter another complexity of screening. Efficiency depends on the ratio of true passes to true fails. In this study there were many more true passes than true fails. So a fairly lax criterion, like SNR = 3 dB, will pass a relatively large number of true passes, and that will make the efficiency look good. But the price paid is that the sensitivity (hit rate) is only 55.6 percent. Using this highly efficient criterion, they have missed almost half of the people they were screening for. Again, changing the pass/fail criterion to SNR = 19 dB raises the hit rate to an acceptable level, 88.9 percent, but the efficiency drops to 29.6 percent. This is because the false-alarm rate is now so high in the true pass group.

The predictive value of a positive result, in the next column to the right in Appendix 3, is defined as the ratio of fails in ears with actual loss to total number of fails (here expressed as a percentage). Similarly, the predictive value of a negative result is the ratio of passes in ears without loss to the total number of passes. Again, these percentages depend critically on the ratio of ears with actual losses to the number of ears without actual losses in the sample. The predictive value of a negative result (far right-hand column) is high because, with either the 3 dB or the 19 dB criterion, only those who are truly normal will pass either criterion. But the predictive value of a positive result declines, from 29.4 percent at the 3 dB criterion to only 9.1 percent at the 19 dB criterion. Even though the hit rate has increased from 55.6 percent to 88.9 percent, the predictive value of a fail has actually declined. This is because of the substantial increase in the false-alarm rate.

Clearly, the choice of a pass/fail criterion must depend on what you want the screening protocol to achieve. If you do not want to miss anyone with a potential problem, then you have to expect that there will be a substantial cost associated with a high false-alarm rate. If you do not want to clog the system with a lot of false positives, then you will have to accept that you are going to miss some, or many, of your targets. If, on the other hand, overall efficiency is your goal, then you have quite a few hard choices to make.

The take-home message is that you have to have some way to handle the false-alarm problem. The answer is follow-up diagnostic evaluation, an essential component of any screening protocol. You have to accept that there will be many false alarms, but they will be sorted out in the second-stage comprehensive diagnostic testing.

Anyone who skips the all-important second stage of diagnostic evaluation, and inappropriately uses a screening technique as a diagnostic tool, should pore over these appendices with a magnifying glass. If, for example, you are screening for auditory processing disorder, you might be tempted to set the pass/fail criterion at one standard deviation from the mean rather than two standard deviations in order to improve the hit rate. But the better the sensitivity (high hit rate) of the screening instrument, the lower the predictive value of an abnormal result. And the associated high false-alarm rate can have unfortunate consequences for the image of our field among both parents and related healthcare professionals.

James Jerger
Editor-in-Chief