**Thoughts on the role of uncertainty in Criteria Assessment.**

Mon Jul 31 10:14:41 2023

Elgin S. Perry, Ph. D.

377 Resolutions Rd., Colonial Beach, Va. 22443

phone: 410-610-1473

email: eperry@chesapeake.net

Peter,

In this document I define 5 approaches to dealing with uncertainty in Criteria Assessment. I have given names to these five approaches, but these names are just for convenient reference within this document, and I doubt that these enjoy universal usage. Of the five approaches, three result in the traditional binary pass/fail endpoints and the remaining two require extending the binary endpoints to three and four endponts.

I have illustrated these approaches on a single line with a water quality measure that I simply call **score**. The score is an observed measure of water quality condition that likely does not agree with the true state of water quality condition. The discrepancy between of the observed score and the true state is quantified by uncertainty. This uncertainty is expressed as an interval along a univariate line (Figure 1). The upper bound represents a point where it is improbable that a score at or above this point would be observed if the true state of the waterbody was below the criterion. The lower bound is defined conversely. These simple figures generalize in an obvious way to using formal measures of uncertainty such as standard errors and confidence intervals and more complex measures of water quality such as confidence envelopes of CFD curves.

When discussing issues associated with the various approaches, I find it convenient to think of criteria assessment as an adversarial process. The adversaries are regulators vs. regulated parties. The role of the regulators is to protect the environment from harm and to require mitigation in instances where harm has occurred. Regulated parties are users of the environment whose goal is to minimize the cost of pursuing their activities in a manner that does not harm the environment. In reality, it is much more complicated than this, but this framework will suffice for this discussion.

Uncertainty in the assessment score gives rise to the possibility of incorrect assessments. If the assessment score indicates failing when in fact the true state of the water body is passing, it will be called a ‘false-fail’. Conversely if the score indicates passing when in fact the waterbody is failing, this is a ‘false- pass’. Both bad decisions have costs. A false pass may result in a failure to require mitigation of environmental damage. A false fail may instigate mitigation when none is required resulting in a loss of resources for real environmental problems.

The **Even Handed Approach** deals with uncertainty by ignoring it. This method uses a score which might be a mean and concludes pass if the mean is above the criterion and fail if it is below. It’s advantages are that it is simple and that it evenly distributes the burden of bad decisions between the adversarial parties. Thus I give the name “Even Handed”. The problem with the even handed approach is the high likelihood of both false-passes and false-fails when the score is in the neighborhood of the criterion. As noted above, this can lead to unacceptable cost on both sides.

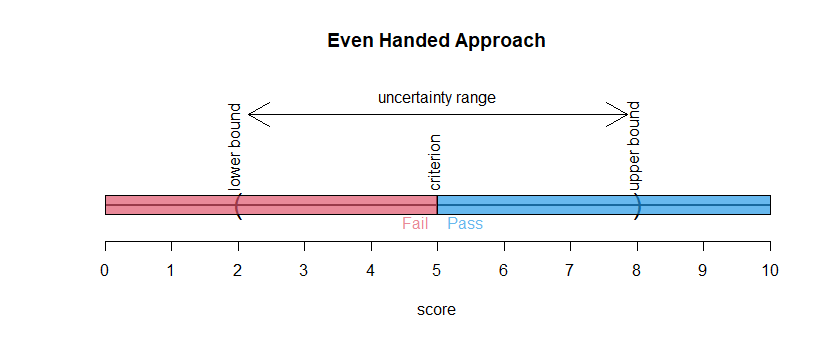


Figure . Even Handed criteria assessment approach.

A long-standing principle in Western judicial systems is the idea of innocence until proven guilty beyond reasonable doubt. One can apply this reasoning to water quality assessment and because this concept has its origins with the Magna Carter (1215), I call the resulting method, the ‘**Magna Carter Approach**’. As noted above, the lower bound sets a limit for which it would be a rare event to obtain a score at or below this bound if the true score were at or above the Criterion (figure 2). Thus, if the lower bound of the uncertainty range sets the dichotomy between pass and fail, the risk of false-fails will be low. However, the risk of false-pass is practically uncontrolled and one would expect that this approach would result in much harm to the environment. Another unfavorable feature of the Magna Carter approach is that it rewards ignorance. Regulated parties benefit from having large ranges of uncertainty and are therefore discouraged from collecting additional data that would improve precision. As a statistician, I cannot endorse an approach that discourages collection of data.

It should be noted that the Magna Carter approach is equivalent to using the lower confidence bound of a confidence interval about a mean as the score to be compared the criterion.

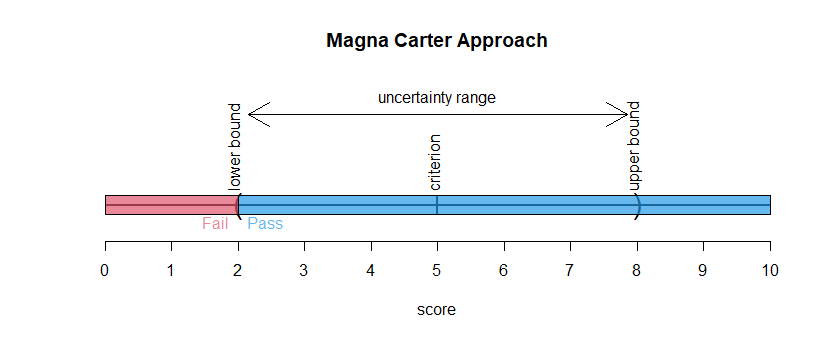


Figure . Magna Carter criteria assessment approach.

Just as the Magna Carter approach is focused on controlling the risk of a false-fail, there is a converse approach the controls the risk of a false-pass. I call this the ‘**IRS approach**’ because it is reminiscent of the IRS requiring that you prove you are entitled to a deduction before you are allowed to claim it. With the IRS approach, the upper bound of the uncertainty range sets the dichotomy between pass and fail and the risk of false-passes will be low (Figure 3). While the Magna Carter approach puts an unfair burden on the environment, the IRS approach puts an unfair burden on regulated parties. This creates a scenario where mitigation resources are unnecessarily squandered and possibly real environmental problems go un-mitigated. It does negate the problem of encouraging ignorance. Regulated parties would be encouraged to collect sufficient data to get the upper bound as close to the criterion as possible.

As with the Magna Carter approach, the IRS can be equivalently defined as using the upper confidence bound of a confidence interval about a mean as the score to be compared the criterion.

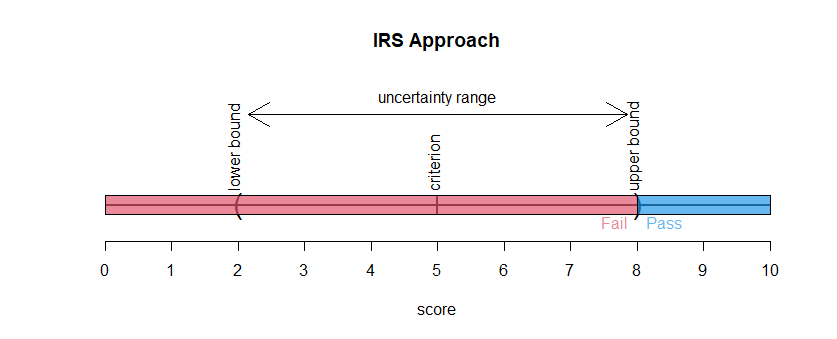


Figure . IRS criteria assessment approach

Among the binary pass/fail approaches, the even-handed approach seems the most fair, but it has the problem of high rates of false-pass and false-fails. By introducing a third outcome category, the problem of high likelihood of false-passes and false-fails can be controlled. I call this the ‘**Uncertainty Approach**’ because it introduces an uncertain category between the lower and upper bounds of the uncertainty range (Figure 4.). Using confidence interval principles of statistics, the risk of false-fails and false passes in the uncertainty approach can be controlled at an acceptable level. The disadvantage of this approach is that when uncertainty is large, many assessments will be unresolved.

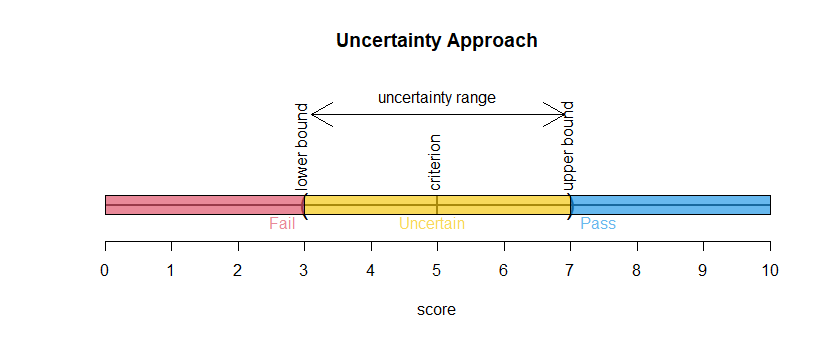


Figure . Uncertainty criteria assessment approach.

In an attempt to solve the issue of unresolved outcomes with the Uncertainty Approach, I introduce a 2/4 outcome approach that I call the ‘**Conditional Approach**’ (Figure 5.). In this approach, we go back to the binary outcome of pass/fail for scores above and below the criterion, but we add a qualifier to the pass/fail outcome that reflects our level of certainty. Moving from left to right in Figure 5, the four outcomes are:

1. Fail or unconditional fail, where we are reasonably certain the data indicate a true condition below the criterion.
2. Conditional fail, the data indicate a true state below the criterion, but we are less certain of that condition.
3. Conditional pass, the data indicate a true state above the criterion, but we are less certain of that condition.
4. Pass or unconditional pass, where we are reasonably certain the data indicate a true condition above the criterion.

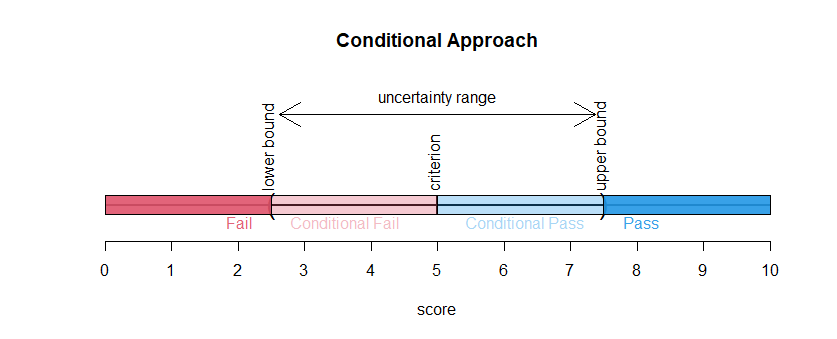


Figure . Conditional criteria assessment approach.

My thinking is that the four categories would allow for different levels of mitigation enforcement. For example:

1. Fail – mitigation required
2. Conditional Fail – mitigation optional, additional data required
3. Conditional Pass – additional data optional
4. Pass – no mitigation action recommended.

To me it seems that the Conditional Approach has a lot of advantages in terms of fairness, lowering concern about rounding, and ensuring that limited resources get used wisely, but it could be a heavy lift to part with tradition..end of file

From: Peter Tango <<ptango@chesapeakebay.net>>

To: eperry <<eperry@chesapeake.net>>

**Subject: Hi Elgin - question about our CFD and the reference curve**

Date sent: Thu, 27 Jul 2023 21:16:13 +0000

Copies to: Rebecca Murphy <<rmurphy@chesapeakebay.net>>, Gary Shenk <<GShenk@chesapeakebay.net>>

Hi Elgin!

I hope all is well for you and your lovely wife down there in these sultry summer days. Overall, not complaining though when more than half the country seems locked in >100 degree conditions, we have had many summer days at home in the 80s, nights sometimes down into the 60s. No complaints at all 😊.

Tickling your brain - We have a discussion bubbling around about the certainty of our assessment using the CFD curve and the comparison with the reference curve. On the one hand, I can recall a simulation study you did with CHLA on a segment in VA to suggest the assessment was like flipping a coin, you gave some sense about the certainty about the assessment.

Newer question that is kind of an old question being revisited now - folks are asking about rounding. Case and point, we have the 10% reference curve, we have 30 day mean 3 year assessment results for summer season. And if we find a segment is 10.00001 or 0.00001 out of attainment, we say it is not meeting the criteria. Any violation (0.1, 0.001, 0.0001, etc) is considered to be out of attainment. Similarly, 9.9999 is in attainment. So - We don't round up, and we don't round down. The final answer gets classified, meets criteria or not, just as the numbers relate to the crisp 10% threshold.

A new proposal asks that given some segments are so fractionally close to the decision threshold, should we have an uncertainty interval around the 10% curve (or maybe the assessment curve?). There is a request for considering something so close to the decision threshold that we need more data to produce a stronger classification, e.g. if something is within + or - 0.05 (fictitious # for show here) it is labeled "uncertain". This has meaning in EPA regulatory reporting world to have this uncertain category of reporting.

My faded memory seems to think you gave some thought to this question of uncertainty around the assessment in the process of us documenting it. Since we never altered this rule (i.e., we do not have any "uncertainty interval" to work with), I am thinking you may not have landed on a happy place solution. If you did, we welcome revisiting it. If you didn't, would you have an option or two for setting some sort of confidence boundary on the 1) space time reference curve, and/or 2) the assessment curve? Or, perhaps, is it advisable to try to go down that path at all given all the uncertainties in the process from sample collection to measurement uncertainty to rounding (we do round during the interpolation process to 2 significant figures with the data, but the final result is taken without rounding to characterize a segment as meeting or not meeting its criteria.

You thoughts and comments welcome as we prepare to discuss this further in a September CAP WG meeting as well as preparing for that meeting.

Take care, thanks so much for any thoughts you might share. Have a good night,

Cheers!

Peter T. 😊