Another Perspective of the Miller Forensic Assessment of Symptoms Test – Part II: A Quantitative Review

D. Detullio

ABSTRACT

Reference [1] presented pooled data for the specificity of the M-FAST cut-off, but ignored or excluded data based on poor justifications and used questionable analytic methods. The analyses here corrected the problems associated with [1].

No moderator substantively influenced sensitivity values. Therefore, sensitivity values were pooled across all studies (k = 25) to provide an overall estimate. Overall, the average sensitivity of the M-FAST cut-off was estimated to be 0.87, 95% CI [0.80, 0.91], and 80% of true sensitivity values were estimated to range from 0.63 to 0.96. Thus, there could be methodological scenarios when the M-FAST cut-off may not operate efficiently.

Average specificity values for the M-FAST cut-off were moderated by one variable: the comparison group. On average, specificity values for clinical comparison (k = 15) groups (i.e., 0.80, 95% CI [0.73, 0.85]) were lower than specificity values for non-clinical comparison (k = 11) groups (i.e., 0.96, 95% CI [0.89, 0.99]). Unlike the CIs, the estimated distributions of true specificity values for the two subgroups overlapped, which suggests there could be scenarios when these subgroups share the same true specificity value.

The M-FAST was designed to be a screener to detect potential feigning of psychiatric symptoms. An examinee is never to be designating as feigning or malingering psychiatric symptoms based on only a positive M-FAST result. As a screening instrument, the results here show that the M-FAST cut-off is operating adequately overall and negate the conclusions of [1].

Keywords: Feigning, Malingering, Meta-Analysis, Miller Forensic Assessment of Symptoms Test (M-FAST).

I. INTRODUCTION

Reference [1] concluded that the Miller Forensic Assessment of Symptoms Test (M-FAST; [2]) should no longer be used in practice. Reference [1]’s conclusion was based on methods that were questionable at best and seemingly biased. That is, a number of relevant studies were not included, analyses were based on a small number of studies (k), and typical methods for pooling values were not employed (e.g., random-effects models; [3] - [5]). In the companion piece to this article [46], research findings for the M-FAST cut-off were discussed at length, including the strengths and weaknesses of the associated study designs. However, the results from these studies were not analyzed. Due to the subjectivity of narratively reviewing findings and some limitations of [6] as well as weaknesses in [1], it was decided to re-analyze sensitivity and specificity values for the M-FAST data to overcome the problems of both meta-analyses and provide a fuller picture of the operating characteristics for the recommended M-FAST cut-off ([2], p. 12).

II. AN UPDATED, BALANCED META-ANALYSIS OF THE M-FAST TOTAL SCORE

All the studies included in analyses here were those discussed in the companion article, and the coding of the studies remained identical to the companion article [46]. The coded study-level characteristics and sensitivity/specificity values are presented in Table I in the Appendix.

Separate random-effects models for the sensitivity and specificity values of the M-FAST cut-off ([2], p. 12) were applied. Similar to [6], the sensitivity and specificity values were transformed to logits for analyses and transformed back to proportions for interpretation [7]. Depending on the analyses performed, some studies were excluded because they were not relevant to the features being examined, and this will be noted below.

Heterogeneity statistics were calculated to examine the extent to which the differences among observed sensitivity and specificity values were due to differences in true values. That is, the F statistic indicates how much observed variation is due to variation in true values [8]. Reference [9] provided some rough guidelines for F values of 25%, 50%, and 75% as representing low, moderate, and high amounts of heterogeneity. Reference [8] critiqued F and associated
interpretations because it does not provide estimates for the range of absolute values for true statistics when there is heterogeneity. Therefore, the standard deviation of true statistics needs to be estimated from the observed values to provide an idea of the absolute variability in the distribution of true statistics. The estimated standard deviation of true statistics can be designated as T. Reference [8] showed how T² and the M variance can be used to calculate prediction intervals (PI). Reference [5] advocated for using only the estimated standard deviation of true statistics for identifying the potential range of true statistics because PsIs focus on predicting “the true effect in a new study” ([9], p. 148). In contrast, by using only the estimated standard deviation of true statistics “the focus is on the state of nature in the population and not on the possible results of the next empirical study” ([5], Chapter 5 – subsection: Credibility Intervals, Confidence Intervals, and Prediction Intervals in Meta-Analysis, para. 2). When intervals for true statistics are calculated using only the estimated standard deviation, [5] referred to them as credibility intervals (CrI). For the analyses below, intervals around the estimate for the mean of the distribution of true statistics (i.e., sensitivity and specificity) were calculated using only T. So, the intervals were referred to as CrI, but technically, these were not CrIs as [5] intended. That is, [5] advocated for correcting methodological artifacts in analyses to provide a more precise estimate of true values. It appears such corrections and statistical programs have only been developed in the context of effect sizes, such as r, which prevented the use of the corrections in the analyses of proportions below. Thus, the CrIs below can be viewed as rough approximations for the range of true values. Similar to [5], 80% CrIs were calculated. The intervals were calculated with the logits for T and the logits of proportions. The values were back-transformed to proportions for interpretation.

When sensitivity and specificity values were not reported in studies, the method from [1] was used to estimate sensitivity and specificity values. Then, such estimates were used to approximate the number of true-positives and true-negatives, and these values were used to calculate sensitivity and specificity values in analyses. This extra step resulted in slightly different sensitivity and specificity estimates when compared to the sensitivity and specificity estimates from using the [1] estimate directly, but it was required to know the number of true-positives and true-negatives to run analyses. Nonetheless, when the estimates differed, it was only by ± 0.01. By using the method from [1], it increased the k sensitivity/specificity values, which is an improvement over [6]. K. Ahmadi of [10] and A. Azizian of [11] provided data requested for the analyses. Reference [10] provided sensitivity/specificity values in their study, but data was requested because the number of true-positives and true-negatives were unclear. Reference [6] obtained data from [12] previously for analyses. For a number of the sensitivity/specificity values estimated here, [6] previously attempted to contact authors (e.g., [13] - [18]) for the data to compute such values, but the data could not be obtained.

Moderator analyses were performed to examine if sensitivity and specificity values varied meaningfully based on a number of variables. Specifically, it was decided to compare sensitivity and specificity values from studies based on the following variables: feigning methodology; use of the Structured Interview of Reported Symptoms (SIRS; [45]) to identify groups; language of the M-FAST (English vs non-English). Both sensitivity and specificity values were examined for these variables because both of these values were hypothesized to be impacted by the aforementioned moderators. Only specificity values were compared among studies which had clinical and non-clinical comparison groups because sensitivity values are not calculated with the comparison group, as opposed to effect sizes. Therefore, sensitivity values were not expected to be impacted by the comparison group. A meta-regression was performed for sensitivity values from simulation designs based on ratings for the quality of the simulation designs to examine if the strength of the simulation design moderates sensitivity values. Sensitivity values were also compared among studies that did not include any cautions, studies that only included cautions, and studies that provided coaching. These analyses were conducted to overcome limitations of [6] and [1] as well as to examine other potential design moderators. For all moderator analyses, subgroups were not assumed to have equivalent estimates of T [3]. To identify potential moderators, CIs and CrIs were examined as opposed to significance tests to prevent the potential for type I errors. Prior to the moderator analyses, [1] was replicated with the exact sensitivity and specificity values from those studies which were estimated in [1]. All analyses were performed using the meta package in R [19].

A. Reference [1] Update

The majority of the specificity estimates from [1] were underestimates. Therefore, it was decided to re-analyze the specificity values from [1] with specificity values calculated directly with the use of true-positives and true-negatives. Among the five specificity values for clinical samples from [1], an average specificity of 0.68 was observed, 95% CI [0.52, 0.80], $\hat{F} = 80.4\%$, 80% CrI [0.47, 0.83]. The average specificity observed here is a 0.07 increase from that reported in [1] (i.e., 0.61). It is not entirely surprising that there is a high amount of heterogeneity considering that these studies included clinical participants from studies with a variety of methodological differences. Given this heterogeneity, these studies are providing an estimate of the average specificity among all true specificity values for the recommended M-FAST cut-off in studies which do not use the SIRS and had a clinical comparison group. For this study design (i.e., no SIRS and a clinical comparison group), the CrI indicates that 80% of true specificity values are expected to be within the bounds of 0.47 to 0.83. When relying on only five specificity values to provide estimates, true specificity values for this design type could range from poor to what may be considered acceptable for a screening instrument. Therefore, it is possible that there are other methodological characteristics for this type of study that could diminish the specificity of the M-FAST cut-off.

Reference [1] appeared to ignore [20] and excluded [21] based on an erroneous justification. Consequently, it was decided to update [1]’s analysis to include these two studies, which could improve the population estimates given the prior analysis only had five specificity values. When these studies are included (k = 7), the average specificity becomes 0.72, 95% CI [0.60, 0.81], $\hat{F} = 76.3\%$, 80% CrI [0.54, 0.85]. With
these two studies added, the estimated average specificity for the distribution of true specificity values appears to be adequate for a screening instrument. Again, there is still an unsurprisingly high amount of heterogeneity. However, the addition of the two specificity values appeared to improve the estimate of \( T \) for the specificity values given the narrower CrI, which shows that 80% of true specificity values are expected to be between 0.54 and 0.85. The lower limit of this range increased, but it still appears lower than what one may hope for in a screening instrument. However, within this type of study (i.e., No SIRS and clinical comparison), the upper range of estimated true specificity values could be viewed as adequate.

In [22], a comparison group of college students who displayed symptoms of posttraumatic stress disorder (PTSD) were included. Even though this group was composed of college students, the fact that they displayed PTSD symptomatology could warrant their designation as a clinical comparison group. Therefore, the preceding analysis was updated with [22] \((k = 8)\), which brought the average specificity for the M-FAST cut-off to 0.73, 95% CI [0.62, 0.81], \( F = 73.0\% \), 80% CrI [0.56, 0.85]. The addition of [22] only slightly changed the average specificity estimate and CrI from the prior analysis, but this change was still an improvement in that the specificity increased and the estimated width of the CrI decreased.

Lastly, [23] was published after [1], and this study appeared to meet the criteria that [1] used in examining specificity values. The prior analysis was updated a final time to include [23] \((k = 9)\), and the average specificity was found to be 0.75, 95% CI [0.65, 0.82], \( F = 74.4\% \), 80% CrI [0.58, 0.86].

These analyses were performed to correct the problematic analyses and associated interpretations of [1]. The data show that the recommended M-FAST cut-off has sufficient specificity, on average. However, given the CrIs, there are scenarios in which the specificity for the M-FAST cut-off is likely to be poor among studies that do not use the SIRS and have a clinical comparison group. The studies within these analyses that had the lowest specificities were simulation designs that incorporated translated versions of the M-FAST (i.e., [11], [24]). It is worth noting that [10] and [21] found adequate specificities for the M-FAST among clinical comparison groups and used translated versions of the M-FAST (i.e., Persian and Turkish). Perhaps, there is something unique to simulation designs which use M-FAST translations and clinical comparison groups because [11] and [24] were simulation designs whereas [10] and [21] were partial criterion designs (PCDs). Based on this type of design (i.e., no SIRS and clinical comparison group), [1]’s blanket conclusions regarding the M-FAST specificity are erroneous because the average estimates for the distribution of true specificity values as well as the CrIs demonstrate that the M-FAST specificity values span into what may be viewed as acceptable for a screening instrument.

### B. Comparison Group

Given the potential for specificity values to be impacted by genuine psychopathology, it was decided to examine whether the comparison group (non-clinical vs clinical) substantively influenced specificity values. This analysis differs from [1] because values were not excluded if they were from studies which made use of the SIRS [45]. Among non-clinical comparison groups \((k = 11)\), the distribution of true specificity values for the M-FAST cut-off was estimated to have an average specificity of 0.96, 95% CI [0.89, 0.99], \( F = 74.2\% \), 80% CrI [0.77, 1.00]. Among clinical comparison groups \((k = 15)\), the distribution of true specificity values for the M-FAST cut-off was estimated to have an average specificity of 0.80, 95% CI [0.73, 0.85], \( F = 77.6\% \), 80% CrI [0.64, 0.90]. The CrIs for the subgroup averages do not overlap. Therefore, specificity values are greater among non-clinical comparison groups on average. High amounts of heterogeneity were present within each group though, suggesting the observed values within the subgroups do not share the same true values. The estimated CrIs show that 80% of true specificity values for the M-FAST cut-off can range from 0.77 to 1.00 among non-clinical comparison groups, and 80% of true specificity values for clinical comparison groups can range from 0.64 to 0.90. Unlike the CIs, the estimated CrIs for subgroups overlap. This could suggest that there are scenarios when studies with clinical and non-clinical comparison groups may share the same true sensitivity value. Therefore, it is possible that an additional variable could interact with the type of comparison group to impact specificity values, especially when considering the amount of heterogeneity within the subgroups.

### C. SIRS vs. No SIRS

Sensitivity and specificity values for the M-FAST cut-off may be greater among those studies which used the SIRS to identify relevant groups in known-groups comparisons (KGCs) and PCDs. Therefore, it was decided to examine whether the average sensitivity and specificity values among studies which used the SIRS were substantively greater than the averages of such values for studies which did not use the SIRS to identify feigning groups. Since these analyses focused on designs which used other psychological measures to establish feigning groups, simulation designs were not included, and [25] was excluded because it was not a simulation design, PCD, or KGC.

For the M-FAST cut-off, the distribution of true sensitivity values from studies which did not use the SIRS \((k = 4)\) was estimated to have an average sensitivity of 0.90, 95% CI [0.56, 0.99], \( F = 81.7\% \), 80% CrI [0.49, 0.99]. The distribution of true sensitivity values for studies that used the SIRS \((k = 7)\) was estimated to have an average sensitivity of 0.88, 95% CI [0.69, 0.96], \( F = 78.8\% \), 80% CrI [0.55, 0.98]. Given the overlap in CIs and estimated distributions of true values (i.e., CrIs), the use of the SIRS does not appear to moderate sensitivity values for the M-FAST cut-off. There were approximately high amounts of heterogeneity within the subgroups, suggesting that the studies within the subgroups do not share the same true sensitivity values. The CrIs for both subgroups were also quite wide: within both distributions of true sensitivity values (i.e., SIRS vs no SIRS), the estimated sensitivity values for the M-FAST cut-off could range from chance to nearly perfect. The overlapping CrIs suggest that designs which do and not use the SIRS could share the same true sensitivity values. In both subgroups, there were single sensitivity values that were noticeably lower from the other values. This may have impacted the
estimated range of true sensitivity values given the small number of studies within both subgroups. In the subgroup where the SIRS was not used, the lowest sensitivity value was observed in [23]. Reference [23] used scales intended to detect feigning somatic and cognitive symptoms. The M-FAST was not designed for identifying possible feigning of somatic or cognitive symptoms, which could explain the low sensitivity value that was observed. The lowest sensitivity value within the SIRS subgroup was drawn from [12], and this sensitivity value was the lowest of all the sensitivity values across studies.

For the M-FAST cut-off, the specificity values from studies which did not use the SIRS ($k = 4$) were found to have an average specificity of $0.84$, $95\%$ CI $[0.78, 0.88], F = 0.0\%$, $80\%$ CrI $[0.84, 0.84]$. The distribution of true specificity values from studies that used the SIRS ($k = 7$) was estimated to have an average specificity of $0.88$, $95\%$ CI $[0.82, 0.92], F = 66.1\%$, $80\%$ CrI $[0.80, 0.93]$. The overlap of the CrIs and CIs suggests that the use of the SIRS does not moderate specificity values. Interestingly, there was no heterogeneity within the subgroup of specificity values when the SIRS was not used. This suggests that all the specificity values from these studies share the same true specificity value and observed differences are due to only sampling error [3], [5]. Consequently, the pooled specificity and associated CI provide estimates for the true specificity value among these studies, but the estimates should be viewed with caution because they are based on a limited number of values. The subgroup of specificity values which used the SIRS fell within the bounds of moderate to high heterogeneity, but the range of estimated true specificity values still generally appear acceptable for a screening instrument.

On average, the sensitivity for the M-FAST total score cut-off appears to be adequate for a screening instrument in study designs that do and do not use the SIRS. However, based on the estimated CrIs, there are likely study designs in which the true sensitivity for the M-FAST cut-off is poor. Again, such CrIs were estimated with small numbers of values, and two of the values (i.e., [12], [23]) could have overly influenced the CrI estimates. Despite the wide CrIs and CIs, the results for the sensitivity values were similar for studies which did and did not use the SIRS. Therefore, concerns that sensitivity values are overly inflated when the SIRS is used to identify feigning groups is not warranted based on the results observed here. The specificity values of the subgroups were adequate as well, and the CrIs did not demonstrate the width that was observed for the sensitivity estimates.

D. Feigning Methodology

Three types of feigning methodologies have been used to study the operating characteristics of the M-FAST total score: PCDs, KGCs, and simulations. Given the differences in these methodologies, it was decided to examine if the operating characteristics of the M-FAST cut-off differed as a function of feigning methodology. As was discussed in the companion article [46], two studies ([2], [26]) were considered to fall somewhere between PCDs and KGCs. Consequently, an additional category of feigning methodology was coded for the purposes of these moderator analyses (i.e., KGC/PCD). Reference [25] was not included in analyses because it was not any of the aforementioned feigning methodologies.

For the M-FAST cut-off, the sensitivity results for the four subgroups of feigning methodology were as follows: simulation ($k = 14$), $0.86$, $95\%$ CI $[0.79, 0.91], F = 74.6\%$, $80\%$ CrI $[0.69, 0.94]$; PCD ($k = 4$), $0.90$, $95\%$ CI $[0.56, 0.99], F = 81.7\%$, $80\%$ CrI $[0.49, 0.99]$; KGC ($k = 5$), $0.87$, $95\%$ CI $[0.54, 0.97], F = 76.7\%$, $80\%$ CrI $[0.42, 0.98]$; KGC/PCD ($k = 2$), $0.90$, $95\%$ CI $[0.79, 0.96], F = 0.0\%$, $80\%$ CrI $[0.90, 0.90]$. All the CIs overlapped with one another as well as the CrIs, suggesting that these feigning methodologies could share the same true sensitivity values. With respect to the PCD and KGC subgroups, the sensitivity CIs and CIs were quite wide, ranging from around chance to nearly perfect. This could be due to the small number of values within these subgroups, which likely led to the CIs and CIs being influenced by individual values (i.e., [12], [23]). Nonetheless, the CIs show that there could be true sensitivity values for the M-FAST cut-off among KGCs and PCDs that are unacceptably low. The KGC/PCD group did not have any heterogeneity, but this subgroup was based on only two studies. Therefore, the lack of heterogeneity should not be viewed as particularly meaningful.

The specificity results for the four subgroups of feigning methodology were as follows: simulation ($k = 14$), $0.92$, $95\%$ CI $[0.79, 0.97], F = 78.6\%$, $80\%$ CrI $[0.50, 0.99]$; PCD ($k = 4$), $0.84$, $95\%$ CI $[0.78, 0.88], F = 0.0\%$, $80\%$ CrI $[0.84, 0.84]$; KGC ($k = 5$), $0.90$, $95\%$ CI $[0.83, 0.94], F = 73.6\%$, $80\%$ CrI $[0.81, 0.95]$; KGC/PCD ($k = 2$), $0.82$, $95\%$ CI $[0.73, 0.89], F = 0.0\%$, $80\%$ CrI $[0.82, 0.82]$. Again, all the CIs for the subgroups overlapped with one another, suggesting that the average specificity values did not meaningfully differ among subgroups. Two subgroups did not show any heterogeneity, but with the small number of values within each subgroup, the lack of heterogeneity may not be reliable. The CrI for the specificity values within simulation designs was particularly wide, which is not surprising considering this subgroup included the lowest specificity values.

These results support that the M-FAST cut-off operates adequately on average across different types of feigning methodology. Therefore, the M-FAST operating characteristics appear to be consistent across designs that focus on internal validity and generalizability [27].

E. Language

The M-FAST has been translated to languages other than English. Therefore, it was decided to examine if sensitivity and specificity values for the M-FAST cut-off differ when the M-FAST is administered in a language other than English. Reference [24] was not included in analyses because the full sample was not administered a translated version of the M-FAST (i.e., 54.2% received a French version). Even though slightly more participants received a translated version, there were concerns about incorporating this study with others in which all participants received a translated version of the M-FAST.

When administered in English ($k = 20$), the M-FAST cut-off was found to have an average sensitivity of $0.83$, $95\%$ CI $[0.76, 0.89], F = 78.6\%$, $80\%$ CrI $[0.61, 0.94]$. When the M-FAST was not administered in English ($k = 4$), the cut-off was found to have an average sensitivity of $0.98$, $95\%$ CI $[0.76, 1.00], F = 0.0\%$, $80\%$ CrI $[0.80, 1.00]$. The overlapping CIs and CIs suggest that non-English translations of the M-
FAST do not moderate sensitivity values. Interestingly, there was no estimated heterogeneity within the non-English M-FAST subgroup, but the CrI was not the same width as the CI, which should occur when there is no heterogeneity. It is likely this occurred due to the data structure for this subgroup, and so the results should be viewed with caution. Furthermore, it was expected to observe some heterogeneity in the non-English M-FAST subgroup because all the studies included different translations. This raises additional concerns regarding the observed results.

With regards to specificity, the cut-off for the English M-FAST (k = 21) was found to have an average specificity of 0.90, 95% CI [0.84, 0.93], F = 70.4%, 80% CrI [0.70, 0.97]. The cut-off for translated versions of the M-FAST (k = 4) was found to have an average specificity of 0.86, 95% CI [0.59, 0.96], F = 75.7%, 80% CrI [0.54, 0.97]. The language of the M-FAST did not substantively impact the specificity of the cut-off because the overlapping CIs and CrIs suggest that these subgroups could share the same true specificity values. However, for the translated subgroup, the CrI was quite wide. This suggests that the specificity for the M-FAST cut-off could be particularly low in some methodologies which employ a non-English M-FAST.

These results provide some initial encouragement for the translation of the M-FAST, but the non-English subgroup was based on a heterogeneous sample of languages and participants with different cultural backgrounds. Consequently, the results here should not be used in support of specific translations of the M-FAST or to support its use with examinees from specific cultures. If [24] administered a translated version of the M-FAST to the entire sample, it could have been included in analyses, and assuming the observed specificity remained the same, it may have impacted the results. Future studies need to be conducted with translated versions of the M-FAST to better understand how it operates in languages other than English and with examinees from various cultural backgrounds. This analysis is too coarse to detect the nuances of how language and culture can impact assessment results.

F. Caution and Coaching

When cautions and/or coaching are incorporated into feigning designs, sensitivity values are expected to be lower when compared to those designs without cautions and/or coaching [27]. Consequently, it was decided to compare studies based on the presence of cautions or coaching. For this moderator analysis, designs that incorporated any form of coaching (i.e., symptom-based or detection-based) were combined into the same subgroup because there was only a single study which relied upon symptom-based coaching (i.e., [22]). Studies that included cautions were split into two separate categories: caution regarding the ability to detect feigning (i.e., detection-caution) or caution to present believably (i.e., believable-caution; see Table I in the Appendix). If both types of cautions were included, the studies were coded as detection-caution. It was decided to split these two cautions separately because a caution which informs participants that feigning can be detected seems similar to detection-based coaching. That is, detection-based coaching informs participants that feigning can be detected and attempts to provide strategies to avoid such detection. The only difference between detection-cautions and detection-based coaching is that detection-cautions do not inform participants how to avoid the feigning detectors. It does not appear that these different types of cautions have been theorized to differ previously [27]. So, this comparison was somewhat exploratory. If studies included both a caution and coaching, they were coded as coaching.

The sensitivity results for the M-FAST cut-off among the four subgroups were as follows: no cautions/coaching (k = 2), 0.76, 95% CI [0.39, 0.94], F = 94.9%, CrI [0.44, 0.93]; detection-caution (k = 7), 0.88, 95% CI [0.78, 0.94], F = 60.1%, CrI [0.72, 0.95]; believable-caution (k = 2), 0.88, 95% CI [0.80, 0.93], F = 0.0%, 80% CrI [0.88, 0.88]; coaching (k = 3), 0.82, 95% CI [0.73, 0.89], F = 0.0%, CrI [0.82, 0.82]. All CIs and CrIs overlap, suggesting that the presence of cautions and coaching did not substantively influence sensitivity values for the M-FAST cut-off. Unfortunately, there were only two studies within two of the subgroups. One would expect the no cautions/coaching subgroup to have the highest average sensitivity, but the feigning design for one of the two studies (i.e., [28]) focused on feigning PTSD symptoms associated with a motor vehicle accident. Given that participants were feigning such specific symptoms, it likely led to a low sensitivity value and impacted the average [2]. It is suggested to focus more attention to the results for the cautions and coaching subgroups. The averages for these subgroups align with what might be expected for cautions and coaching. That is, coaching would lead to lower sensitivity values than simple cautions, but the overlapping CIs and CrIs suggest that these designs could share the same true sensitivity values. The coaching and believable-caution subgroups had no heterogeneity, suggesting the values within these subgroups share the same true value. However, the sensitivity values for the coaching subgroup were drawn from a limited number of studies with varied methods of coaching. Consequently, it is possible that the homogeneity within this subgroup was due to the low k. Similarly, the homogeneity within the believable-caution subgroup was likely due to the low k. With studies that have been conducted thus far, the M-FAST cut-off has sufficient sensitivity in the presence of cautions and coaching.

G. Feigning Procedures

A rating system was developed for the quality of feigning procedures within simulation designs based on selected factors from [27]. These factors are abbreviated in Table I in the Appendix and include the following: specific, real-world feigning scenario; simulators’ familiarity with the feigning scenario; caution to present believably or caution regarding scales to detect feigning; coaching; internal motivation prompts; incentives; relevance of the research; post-manipulation simulation checks. In developing the rating system, it was decided to assign different weights to select factors in accordance with their potential importance for feigning procedures within simulation designs. Because the ratings were intended to be summed across the factors to obtain a total score, the differential weighting would increase the chances of studies with stronger simulation methodology to have higher total scores than studies with weaker methodology. For instance, if two studies had the same number of factors present but one of the studies had
methodological factors present, the differential weighting would show this in the total score.

All the factors outlined in [27] could be seen as important to the feigning condition in simulation designs. However, it appears some factors are likely to have greater influence on the feigning component of these designs. Simulation designs are intended to enhance internal validity, and manipulation checks confirm whether participants instructed to feign symptoms actually followed directions. Therefore, manipulation checks assist with determining whether feigning caused elevations on the M-FAST total score. With this in mind, manipulation checks for the feigning condition could be considered the most important factor outlined in [27], and manipulation checks were weighted with the highest rating (i.e., 4). Incentives in simulation designs are intended to increase participants’ motivation to successfully feign, similar to individuals who feel motivated by incentives to feign in the real-world. Considering the role that incentives may play, designs which included positive incentives (e.g., monetary reward) were rated a 2. Reference [27] noted that negative incentives (e.g., “public disapproval”) should be a core element for all response-style research because of findings that negative incentives improved simulators’ feigning performance (p. 599). If negative incentives were present, a rating of 3 would be assigned, but none of the designs here included negative incentives. Similar to incentives, [27] recommended that simulation instructions should include prompts to increase simulators’ internal motivation to successfully feign. Reference [27] split these prompts into positive (e.g., “can you beat the test?”) and negative (e.g., “Unfortunately, past research shows that most persons like yourself do not have the skill to ‘beat the test’…”) (p. 599). Because such prompts are intended to increase participants’ internal motivation to feign successfully, a study would have been assigned a rating of 2 when such a prompt was present. Unlike incentives, differences between positive and negative internal motivation prompts were not indicated in [27]. This led to the decision to code a 2 if either were present, but none of the studies coded here appeared to include positive or negative internal motivation prompts. Cautions regarding the ability to detect exaggerating or cautions to present believably help to prevent “the risk of producing distorted and possibly bogus results” ([27], p. 596). For the coding here, a rating of 2 was assigned if participants were presented with a detection-based or believable-based caution given the lack of differences observed between these groups in the moderator analysis above. Similar to cautions, incorporating coaching in designs could reflect real-world scenarios when examinees attempt to obtain information to successfully feign. Reference [27] discussed how detection-based coaching is superior to symptom-based coaching. Therefore, a rating of 3 was assigned when detection-based coaching was present, and a rating of 2 was assigned when symptom-based coaching was present. Reference [27] noted that conveying the real-world relevance of the research to simulators could increase their engagement and motivation. In developing this rating system, it was questioned whether explaining the relevance of the design would be as important as internal motivation prompts or incentives to increase motivation. Furthermore, participants may be informed of how the research is relevant to the real-world but still not see it as relevant to themselves. Consequently, if this factor was present, it was only coded a 1. Reference [27] noted that feigning instructions should be “sufficiently explicit that researchers will be able to make a meaningful interpretation of the resulting data” (p. 595). However, it was also discussed that there may be scenarios when general feigning instructions are applicable, and instructions that are too specific may be difficult for participants to follow appropriately. With these conflicting considerations in mind, it was decided to provide a rating of 1 when specific feigning instructions were provided in the design. Lastly, if participants possibly had familiarity with the simulation scenario through direct experience, it was decided to assign a rating of 2 for the study because participants’ experiences may have allowed them to relate to the scenario more easily. Also, if the participants experienced a similar situation, it could increase the generalizability of the results. A rating of 0 was assigned anytime a factor was not present in the study design.

With the rating system that was developed, a total score was created for the feigning condition in simulation designs by summing across the ratings for the factors previously discussed. This total score could range from 0 to 18. Among the studies (k = 14), the total score for the rating system ranged from a low of 1 [29] to a high of 12 [30] with a M = 6.57 (SD = 3.37) and Mdn = 6. Given that the included studies did not include the full span of total ratings, predictions based on the meta-regression may require some extrapolation and should be interpreted with caution.

A meta-regression was performed to examine if the total score for the ratings of simulation procedures predicted sensitivity values (k = 14) for the M-FAST cut-off. The regression coefficient for the total rating was found to be -0.07, 95% CI [-0.22, 0.08], and the intercept coefficient was found to be 2.24, 95% CI [1.17, 3.31]. These coefficients are on the logit scale. The CI for the total rating regression coefficient does not suggest that the quality of simulation procedures substantively influence sensitivity values for the M-FAST cut-off.

III. DISCUSSION AND LIMITATIONS OF META-ANALYSIS

The purpose of the above analyses was to overcome weaknesses of [1] and limitations with [6] as well as to examine other potential moderators of the operating characteristics for the M-FAST cut-off. With respect to sensitivity values, none of the above moderators meaningfully impacted the sensitivity of the recommended M-FAST cut-off in that the CIs and CrIs overlapped, suggesting that the subgroups could share the same true sensitivity values. In other words, based on the data thus far, the use of the SIRS, cautions or coaching, methodology type (e.g., simulation vs KGCs), language, and simulation design quality do not moderate sensitivity values for the M-FAST cut-off. Some of these moderator analyses conflict with that of [1], specifically the moderator analysis focusing on cautions and coaching. Reference [1] observed a significant difference between studies that included cautions compared to those studies without cautions. This conflicting result was likely because [1] performed what could be considered fixed effects analyses, whereas the analyses here were random effects. Fixed effects analyses can lead to type I errors [4],

DOI: http://dx.doi.org/10.24018/ejmed.2021.3.6.1143
Vol 3 | Issue 6 | December 2021
46
and it seems likely that the significant moderator for cautions in [1] was a type I error.

Given that none of the moderators impacted the sensitivity values for the M-FAST cut-off, it was decided to perform a final meta-analysis with all sensitivity values to provide an overall average sensitivity for the M-FAST cut-off. The overall average sensitivity for the M-FAST cut-off ($k = 25$) was found to be $0.87$, $95\%$ CI $[0.80, 0.91]$, $I^2 = 74.7\%$, $80\%$ Crl $[0.63, 0.96]$. The CrI shows that there could be true sensitivity values for the M-FAST cut-off that are less than acceptable depending on methodological characteristics. One may look to individual studies to develop potential hypotheses for scenarios when the true sensitivity of the M-FAST cut-off is not optimal. For instance, studies which focus on feigning symptoms not targeted by the M-FAST (e.g., [23]) may have a low true sensitivity value, which would not be surprising [2]. Overall, as a screener for feigning symptoms of a mental illness, the sensitivity of the M-FAST is adequate based on the data available at this time.

Only one moderator variable had an influence on the specificity values for the M-FAST cut-off: the comparison group (i.e., clinical vs. non-clinical). This finding was not surprising based on recommendations for feigning designs, and it aligns with the results from [1] to some extent. However, the results differ in the degree of specificity for the M-FAST cut-off among clinical comparison groups. As discussed in the companion article [46], [1] included a narrow set of studies based on questionable justifications as well as used values that underestimated the specificity values from studies. The moderator analysis here showed that the estimated average of true specificity values for the M-FAST cut-off among clinical comparison groups was $0.80$, which appears adequate for a screening instrument. Even though the CIs for the subgroups did not overlap, which suggests differences between the averages of the two subgroups, the CrIs demonstrated overlap. This suggests that the distributions of true specificity values for clinical and non-clinical comparison groups could overlap. The CrI for the clinical comparison group showed that there could be designs for which the true specificity of the M-FAST cut-off is $0.64$. Again, one might consider examining individual studies to develop potential hypotheses for when the true specificity value of $0.64$ may occur. Among the studies examined here, the lowest specificities were found among clinical samples which included translated versions of the M-FAST. Thus, the specificity of the M-FAST cut-off may be especially poor among this unique population of studies. Consequently, to claim that the M-FAST cut-off operates poorly among all clinical samples is an overly simplistic conclusion [1].

Even though these analyses represent an improvement over [6] and [1], there are still limitations that need to be discussed. The most obvious limitation concerns the small number of $k$ within most subgroups. This likely impacted estimates for the distributions of true sensitivity and specificity values. Reference [3] demonstrated that Ps begin to stabilize when around 24 studies are included in analyses. Now, Ps as calculated in [8] were not applied here, but it is plausible that the small $k$s in subgroups still had an impact on the estimates for the distributions of true values. Due to the small $k$s, interactions between moderator variables also were not examined [5]. Such interactions would have likely led to even smaller $k$s for subgroups or no $k$s for some combinations. If it was possible to examine theoretically plausible interactions between moderators, it might have uncovered homogenous subgroups. The small $k$s in subgroups could also explain why some subgroups were observed to be homogenous, as opposed to sharing the same true value.

As with most meta-analyses, the coding of studies required some subjectivity and judgement calls [31]. It was attempted to rely on low-inference coding to improve the reliability and representation of the coding, but with the variation in which information was reported in studies, there were difficulties in sufficiently coding all variables for studies. For example, [32] asked participants post-simulation questions, but it was unclear if it was confirmed whether the participants followed the simulation instructions. So, this study was coded as not having a post-simulation manipulation check. Additionally, two subgroups of cautions were created for the caution/coaching moderator analysis, and a subgroup of studies considered KGCs/PCDs were created for the feigning methodology moderator analysis. Such coding is likely unique to this meta-analysis, and other researchers may not make such designations. The most significant limitation in coding was the rating system applied for assessing the methodological quality of the simulation procedures. It was attempted to base the rating system on empirical findings and theories underlying feigning methodologies [27], but it is highly likely that alternative weights and criteria could have been applied. Entire studies are devoted to the development of measures for rating the methodological quality of studies (e.g., [33]). Consequently, it might have been overly ambitious to create a method of rating simulation designs for the purposes of this analysis. Although, even if the rating system here was based on a separate study, scales for rating the quality of study designs have been criticized [34].

Similar to [6], the analyses here relied on separate random-effects models for pooling sensitivity and specificity values. More complex analyses have been recommended (e.g., [35]), but it is possible that the results here are similar to what may be observed using more complex analyses [36], [37]. The analyses here are also an improvement on [1], which was akin to a fixed effects meta-analysis and had lower $k$ values. Reference [1] also relied on estimates of specificity values when such estimates could be directly calculated.

Meta-analyses depend upon the data that is available. The results in [6], [1], or here should not be the final word on the M-FAST. As more studies are completed on the M-FAST, additional meta-analyses should be performed to understand the bounds of the operating characteristics of the M-FAST cut-off. With a larger $k$ of M-FAST studies, moderators and interactions between moderators could be more fully explored and potentially reveal homogenous scenarios in which the M-FAST cut-off operates adequately and poorly. It may also be worthwhile for future meta-analyses to make use of bivariate or Bayesian models when pooling sensitivity and specificity values. Although it is less specific to future M-FAST meta-analyses, the system used here to rate the quality of simulation procedures may also be worthwhile to consider for future development. There are scales for rating the quality of studies, but they are much more general (e.g., [33]). With the frequency that meta-analyses are performed on scales to detect feigning (e.g., [38]–[41]), a scale for rating the quality...
of simulation designs could be beneficial, and such a scale could be used in future moderator analyses of the M-FAST.

IV. CONCLUSIONS

Reference [1] attempted to provide the firm conclusion that the M-FAST should no longer be used in practice. This conclusion is erroneous based on the available data. The operating characteristics of the M-FAST are a bit more nuanced than [1] implied. That is, on average, the sensitivity of the M-FAST cut-score is adequate (i.e., 0.87). However, there can be methodological scenarios when the cut-score operates poorly when considering the CrI ranged from 0.63 to 0.96. None of the moderator analyses substantively revealed under what conditions the sensitivity for the M-FAST cut-off would be poor. However, when examining the studies individually, the sensitivity for the cut-off appears to be poor in the context of feigning somatic or cognitive symptoms (e.g., [23], [42]). The M-FAST was not developed to be a screener for somatic or cognitive feigning. Therefore, it is expected that the cut-off would perform poorly in such a situation, and it should not be used to screen for feigning cognitive or somatic symptoms. The M-FAST cut-off has shown some ability to detect potential feigning of PTSD symptoms (see [10], [22]), but this has not always been the case (see [28]). The M-FAST was developed to be a brief screener, and therefore, items could not be incorporated to detect potential feigning of the full spectrum of psychological disorders. This is even acknowledged in the M-FAST professional manual [2]. Therefore, if examinees are feigning specific symptoms (e.g., PTSD), the M-FAST may not detect such individuals.

Reference [1] emphasized the false-positives associated with the M-FAST. False-positives are less of a concern with the M-FAST because it is not intended to make any conclusions regarding feigning or malingering. Nonetheless, the results above show that the M-FAST cut-off appears to have sufficient specificity among clinical comparison groups on average (i.e., 0.80). As with the sensitivity values, there are probably scenarios when the specificity of the M-FAST cut-off is likely to be poor given the CrI ranged from 0.64 to 0.90. Again, when looking at the individual studies, the M-FAST cut-off had the lowest specificities in designs that incorporated clinical comparison groups and translated versions of the M-FAST (i.e., [11], [24]). Reference [1] ignored that translated versions of the M-FAST were incorporated in these designs. This is important to consider because it suggests that the interaction between genuine pathology and cultural/linguistic factors may lead to greater false positives, as opposed to only psychopathology.

As with any instrument, there are appropriate and inappropriate uses of the M-FAST. When using the M-FAST, a positive M-FAST result should never be used to decide if an examinee is feigning or malingering psychiatric symptoms because it is only a screening instrument. Positive M-FAST results should only be used to decide if comprehensive testing, additional observations, and collateral information need to be collected to determine if an examinee is feigning or malingering.
### APPENDIX

<table>
<thead>
<tr>
<th>Ref</th>
<th>Sens</th>
<th>Spec</th>
<th>Comp</th>
<th>Method</th>
<th>SIRS</th>
<th>Translated M-FAST</th>
<th>Specific Scenario</th>
<th>Familiarity</th>
<th>Caution</th>
<th>Coaching</th>
<th>Internal Motivation</th>
<th>Incentives</th>
<th>Relevance</th>
<th>Manipulation Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>[28]</td>
<td>0.51</td>
<td>0.87</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[29]</td>
<td>0.91</td>
<td>1.00</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[15]</td>
<td>0.67</td>
<td>0.95</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>Yes</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>Positive incentive</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[22]</td>
<td>0.78</td>
<td>0.79</td>
<td>C</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Symptom-based</td>
<td>Positive incentive</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[2]</td>
<td>0.93</td>
<td>1.00</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>No</td>
<td>No</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>Positive incentive</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>[32]</td>
<td>0.90</td>
<td>0.69</td>
<td>C</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>No</td>
<td>BC</td>
<td>No</td>
<td>No</td>
<td>Positive incentive</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[16]</td>
<td>0.85</td>
<td>1.00</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>yes</td>
<td>No</td>
<td>No</td>
<td>DC</td>
<td>Detection-based</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>[30]</td>
<td>0.90</td>
<td>1.00</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>DC</td>
<td>Combination</td>
<td>Positive incentive</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[43]</td>
<td>0.86</td>
<td>0.95</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>BC</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[11]</td>
<td>1.00</td>
<td>0.57</td>
<td>C</td>
<td>sim</td>
<td>NA</td>
<td>yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[17]</td>
<td>0.86</td>
<td>0.68</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[18]</td>
<td>0.82</td>
<td>0.73</td>
<td>NC</td>
<td>sim</td>
<td>NA</td>
<td>no</td>
<td>Yes</td>
<td>Yes</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>[24]</td>
<td>0.86</td>
<td>0.37</td>
<td>C</td>
<td>sim</td>
<td>NA</td>
<td>NA²</td>
<td>Yes</td>
<td>Yes</td>
<td>DC</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>[25]</td>
<td>0.74</td>
<td>0.74</td>
<td>C</td>
<td>NA</td>
<td>NA</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[10]</td>
<td>0.90</td>
<td>0.87</td>
<td>C</td>
<td>PCD</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[42]</td>
<td>0.83</td>
<td>0.91</td>
<td>C</td>
<td>KGC</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[14]</td>
<td>0.92</td>
<td>0.72</td>
<td>C</td>
<td>KGC</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[26]</td>
<td>0.86</td>
<td>0.83</td>
<td>C</td>
<td>KGC/PCD</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[12]</td>
<td>0.25</td>
<td>0.95</td>
<td>NC</td>
<td>KGC</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[15]</td>
<td>0.88</td>
<td>0.93</td>
<td>C</td>
<td>KGC</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[21]</td>
<td>1.00</td>
<td>0.78</td>
<td>C</td>
<td>PCD</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[2]</td>
<td>0.93</td>
<td>0.82*</td>
<td>C</td>
<td>KGC/PCD</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[23]</td>
<td>0.53</td>
<td>0.85</td>
<td>C</td>
<td>PCD</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[20]</td>
<td>0.80</td>
<td>0.85</td>
<td>C</td>
<td>PCD</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: Sensitivity and specificity values in italics were estimated using the method from [1]. Ref = reference; Sens = sensitivity; Spec = specificity; Comp = comparison group; NC = nonclinical; C = clinical; Sim = simulation; KGC = known-groups comparison; PCD = partial criterion design; SIRS = Structured Interview of Reported Symptoms; DC = detection-caution; BC = believable-caution; NA = not applicable. All the factors coded for simulation designs were drawn from [27].

²[2] reported a specificity value of .83 for the M-FAST cut-off, but after recreating ns (e.g., true-positives, true-negatives), a specificity value of .82 was obtained. These calculated ns were used in analyses. *[24] was not coded as “yes” or “no” for language because the entire sample did not receive a translated or English version of the M-FAST.
ACKNOWLEDGEMENT

D. Detullio would like to thank T.D. Kennedy (PhD) for his helpful comments. D. Detullio would like to thank (in alphabetical order by last name) K. Ahmadi (PhD), A. Azizian (PhD), and D. Hill (PsyD) for providing requested data.

FUNDING

D. Detullio did not receive any funding as part of this research.

CONFLICT OF INTEREST

D. Detullio does not declare any conflict of interest. The conclusions and opinions in this article do not represent the views of Western State Hospital.

REFERENCES


[36] Dahabreh IJ, Trikalinos TA, Lau J, Schmid CH. Univariate and bivariate likelihood-based meta-analysis methods performed comparably when marginal sensitivity and specificity were the targets of inference. Journal of Clinical Epidemiology. 2017; 83: 8–17.


