Effectiveness of Practices for Improving the Diagnostic Accuracy of Non-ST Elevation Myocardial Infarction in the Emergency Department: A Laboratory Medicine Best Practices Systematic Review

Christopher Layfield¹, John Rose¹, Aaron Alford¹, Susan R. Snyder², Fred S. Apple³,⁴, Farah M. Chowdhury⁵, Michael C. Kontos⁶, L. Kristin Newby⁷, Alan B. Storrow⁸, Milenko Tanasijevic⁹,¹⁰, Elizabeth Leibach⁵, Edward B. Liebow¹¹, and Robert H. Christenson¹²

¹Battelle Memorial Institute
²Geisinger Center for Health Research
³Hennepin County Medical Center
⁴University of Minnesota Medical School
⁵Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA
⁶Virginia Commonwealth University Medical Center
⁷Duke University Medical Center
⁸Vanderbilt University Medical Center
⁹Brigham and Women’s Hospital
¹⁰Harvard Medical School
¹¹American Anthropological Association
¹²University of Maryland School of Medicine

Abstract

Objectives—This article presents evidence from a systematic review of the effectiveness of four practices (assay selection, decision point cardiac troponin (cTn) threshold selection, serial testing, and point of care testing) for improving the diagnostic accuracy for Non-ST-Segment Elevation Myocardial Infarction (NSTEMI) in the Emergency Department.

Design and Methods—The CDC-funded Laboratory Medicine Best Practices (LMBP™) Initiative systematic review A6 Method for Laboratory Best Practices was used.

Results—The current guidelines (e.g., ACC/AHA) recommend using cardiac troponin assays with a 99th percentile upper reference limit (URL) diagnostic threshold to diagnose NSTEMI. The...
evidence in this systematic review indicates that contemporary sensitive cTn assays meet the assay profile requirements (sensitivity, specificity, PPV, and NPV) to more accurately diagnose NSTEMI than alternate tests. Additional biomarkers did not increase diagnostic effectiveness of cTn assays. Sensitivity, specificity, and negative predictive value (NPV) were consistently high and low positive predictive value (PPV) improved with serial sampling. Evidence for use of cTn point of care testing (POCT) was insufficient to make recommendations, though some evidence suggests cTn POCT may result in reduction to patient length of stay and costs.

Conclusions—Two best practice recommendations emerged from the systematic review and meta-analysis of literature conducted using the LMBP™ A6 Method criteria: Testing with cardiac troponin assays, using the 99th percentile URL as the clinical diagnostic threshold for the diagnosis of NSTEMI and without additional biomarkers, is recommended. Also recommended is serial cardiac troponin sampling with one sample at presentation and at least one additional sample taken a minimum of 6 hours later to identify a rise or fall in the troponin level. Testing with high-sensitivity cardiac troponin assays, at presentation and again within 6 hours, is the recommended evidence-based best practice testing algorithm for optimized NSTEMI diagnosis.

Keywords
Acute Coronary Syndrome; Cardiac Troponin; Non-ST-Segment Elevation; Myocardial Infarction

1.0 Introduction
Type 1 myocardial infarction (MI) is one component of the Acute Coronary Syndromes (ACS) [1], a continuum of disease spanning from unstable angina to non-ST-segment elevation myocardial infarction (NSTEMI) and cell death, to ST-segment elevation myocardial infarction (STEMI). The primary cause of most type 1 MIs is the unstable coronary plaque. Type 2 MI is caused by a mismatch of oxygen supply and demand from a variety of causes including coronary spasm, coronary embolism, arrhythmia, anemia, and hypotension [1][2]. Here we consider type 1 and type 2 MI as the single entity of acute myocardial infarction (AMI). The American Heart Association (AHA) conservatively estimated 1,190,000 unique hospitalizations for new or recurrent MI (approximately 70% NSTEMI) that occurred in the United States in 2009 [3]. Classifying NSTEMI separately from STEMI is important because biomarkers play a central role in the diagnosis and management of NSTEMI, whereas STEMI classification is based on the electrocardiogram and biomarkers serve only a confirmatory role in diagnosis.

Biomarkers have evolved to be the cornerstone for the diagnosis of MI [1]; the preferred screening biomarker is cardiac troponin (cTn) [1][4][5]. However, cTn reporting is complicated by the multiple generations of cTn assays that have evolved with varying analytic characteristics and differences among cTn assays which have led, in part, to different diagnostic algorithms that can vary widely. Additionally cTn assays may use subtype I or T as the diagnostic component. Though sensitive assays exist for each subtype, we did not assess these differences, but did include the subtype information where necessary to inform assay selections. The relative diagnostic value of these subtype-specific algorithms in clinical decision making has not yet been assessed.
Our aim was to conduct an evidence-based review of four areas for cTn use for diagnosis of NSTEMI. These areas were selected by an expert panel following a preliminary literature review using the CDC Laboratory Medicine Best Practices A-6 Method [6]. These areas represent potential opportunities for quality improvement in NSTEMI diagnostic practices and include: (1) cTn assay selection, (2) cTn assay diagnostic threshold, (3) use of serial cTn samples to confirm NSTEMI diagnosis, and (4) POCT testing and timely receipt of cTn results.

1.1 Quality Gap 1: Selection of an appropriate biomarker to diagnose NSTEMI

Cardiac troponin assays have evolved with improvements in analytical performance since their introduction in the early 1990s. Early cTn assays were intended to replace for measurements of the MB isoenzyme of creatine kinase (CK-MB, muscle and brain subunits produced by heat muscle). However, by the year 2000, cTn was recognized in the first task force on redefinition of MI document as contributing additional information to CK-MB measurement [7], and is currently the preferred cardiac biomarker [1][4][5]. Based on this report, a consensus-based quality specification of a 10% total coefficient of variation (CV) at the decision point, defined as the 99th percentile of a reference control population became a target performance specification [7]. The International Federation for Clinical Chemistry and Laboratory Medicine (IFCC) has established a table of current, commercially available and research assays, available at www.ifcc.org [8].

First generation assays allowed for cTn detection within 4 to 12 hours of onset of myocardial injury with peak values between 12 and 48 h post-onset, as they were designed to be similar to CK-MB using ROC curve decision cutoffs. Current, more sensitive cTn assays have receiver operator characteristics curve (ROC) areas exceeding 90% sensitivity for AMI diagnosis within 2 h of symptom onset [9]. Cardiac troponin levels in early presenters still may not exceed the diagnostic threshold, though many assays are sufficiently sensitive to detect very small amounts accurately, i.e. ~10 ng/L or ~0.01 μg/L, of cTn within hours of MI symptom onset [9]. The imprecision and lack of diagnostic sensitivity of early, first generation cTn measurements led to combining this marker with other biomarkers such as myoglobin and CK-MB in an effort to improve medical decision making. The relative, marginal increase in diagnostic value in clinical decision support of adding these additional biomarkers to algorithms using currently available high sensitivity cTn tests has not been thoroughly assessed [10].

1.2 Quality Gap 2: Appropriate cTn assay threshold to diagnose NSTEMI

Although the initial Joint European Society of Cardiology (ESC)/ACC redefinition report indicated by consensus that a 10% total CV was necessary at the 99th percentile cutoff [7], a 2003 report [11] found that none of the then-current cTn assays could meet the 10% CV specification. Since that time cTn assays have improved, and the most recent global task force redefinition of MI report reiterated the recommended, consensus-based decision limit for myocardial injury as the concentration corresponding to the 99th percentile of the reference distribution in healthy people and that assays with ≤20% CV at the 99th percentile URL are clinically acceptable [1]. However, assay values between the 99th percentile and a cTn concentration with a 10% total CV can result in unnecessary, missed, or delayed
treatment pending subsequent evaluation [12]. Kaplan-Meier curves and Cox proportional hazard analyses indicated significant risk for death or AMI recurrence as peak concentration of cTnI increased from low (<0.04 μg/L) to intermediate (0.04–0.10 μg/L) to high (>0.10 μg/L) in a study using the AccuTnI assay (Beckman Coulter Inc.) [13]. In contrast, changing the theoretical CV (up to 20% CV) at the 99th percentile resulted in minimal to no overall effect on diagnosis or outcomes in epidemiological studies using simulated evidence [14][15]. Limited evidence using population studies suggests no difference in diagnostic accuracy using the 99th percentile regardless of CV, although some STEMI patients were included [9]. Comparisons of the World Health Organization (WHO) Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA) and ESC/ACC definitions of MI found significant differences in the prevalence of AMI based on the biomarker used (CK-MB versus cTn) and change in biomarker concentration, but found no difference between the 99th percentile and 10% CV cutoff for cTn [16]. A 2012 study in European hospitals indicated that diagnosis was based on the 99th percentile (38%), the 10% CV (41%), or another cut-off (62%) (lower limit of detection (LLOD), manufacturer’s recommendation, or literature citation) [17].

1.3 Quality Gap 3: Timing of serial samples to confirm ACS diagnosis

Any cTn result above a defined AMI cutoff has been shown to be a significant predictor of 30-day mortality [18]. Serial blood samples may improve timely and accurate diagnosis while reducing adverse outcomes among patients who initially have negative cTn results [18][19][20]. The 2012 task force consensus recommendations specify a criterion for diagnosis of AMI as a rise and/or fall of cTn with at least one value above the 99th percentile URL. However this guidance [10] does not specify the exact temporal sequence of sample collection or magnitude of the rise and/or fall of cTn indicating AMI. Current ACC/AHA consensus-based guidelines recommend initial samples collected at presentation with an additional sample between 6 and 8h post-presentation to identify a rise/fall in cTn level [4]. Other organizations [21][22][23][24][42] also recommend serial sampling but at different time points. Several studies of accelerated diagnostic protocols using contemporary sensitive cTn assays suggest that early serial sampling times (e.g., 0 and 1, 2, or 3 h) combined with use of either absolute or relative cTn concentration change for differential diagnosis may still allow safe clinical decision making [25][26][27][28][29]. Studies using high sensitivity cTn assays have shown that absolute concentrations are preferable in improving clinical specificity compared with percentage changes that were used with contemporary, current generation assays.

1.4 Quality Gap 4: Reduced turnaround time within 60 minutes

The recommended turnaround time (TAT) for cTn results is ≤1 h, with 30 minutes as the ideal [30]. Point of care testing (POCT) may substantially reduce TAT [31][32][33], to achieve the ACC/AHA recommendation of 30 minutes (87.3% of cases) [31]. One study was able to achieve 100% delivery of cTnI results to a nurse or clinician within 20 to 33 minutes after blood draw with a median time of 35 minutes from registration [37]. Reduced TAT has been associated with shorter length of stay (potentially between 8% [34] to 36% [35] reduction) in the emergency department (ED), concomitant cost savings [34], and more
efficient coronary care use [35]. However it is unclear whether use of POCT improves clinical outcomes.

2.0 Methods

This evidence review followed the CDC’s LMBPTM A6 Method for conducting systematic reviews for evaluating laboratory quality improvement practices [6]. This approach is tailored to the evaluation of laboratory medicine practice effectiveness studies to support evidence-based best practice recommendations and has been reported in detail elsewhere [6]. A review team, including co-coordinators and staff specifically trained in the A6 Method application, conducted the systematic review. An expert panel, contributing medical and diagnostic perspectives provided additional guidance on the conduct of the systematic review and drafted recommendations based on the review results. The expert panel consisted of individuals selected for their diverse perspectives and subject matter expertise in the review topic, laboratory management, and evidence review methods. (See Appendix A for a listing of Expert Panel members.) In accordance with the A6 Method, the results of the evidence review and best practice recommendations were reviewed by the LMBPTM Workgroup, an independent, multi-disciplinary group comprised of 15 members with broad expertise in laboratory medicine, clinical practice, health services research or health policy (See Appendix B for LMBPTM Workgroup members).

The question addressed in this evidence review is: What cTn testing practices are effective at increasing timely and accurate AMI diagnosis for patients presenting in the ED with signs and symptoms suggestive of NSTEMI? An analytic framework for this quality issue of ACS diagnosis is displayed in Figure 1. The elements of the analytic framework were defined using a PICO (population, intervention, comparator, and outcomes) approach that defined the pertinent study parameters [36]. For this review specific study setting and population inclusion criteria were: studies published between 1996 and 2013, performed in the ED with adult patients suspected of NSTEMI, respectively. STEMI patients were specifically excluded.

The review of practice effectiveness studies was based on a literature search strategy using terms developed with the assistance of a research librarian. An initial systematic search was completed in July 2011 which was subsequently updated in 2013 using four electronic databases (PubMed, Cochrane, Embase, and CINAHL), and additional sources including public and private-source professional guideline electronic databases (AHRQ, CLSI, ISO, NACB) for English language, human subjects, practice-relevant articles, and AMI guidelines from 1996 to 2013. We conducted hand searches of bibliographies from relevant secondary literature, consultation with and references from experts in the field including expert panelists, and solicitation of unpublished quality improvement studies resulting in direct data submissions to the LMBPTM website (http://wwwn.cdc.gov/futurelabmedicine/). A separate search for each of the interventions was performed and the results for all searches were then combined. Search terms and strings for each practice are available in Appendix C.

To reduce subjectivity and the potential for bias, all screenings, abstractions, and evaluations were conducted independently by at least two reviewers, and all differences were resolved.
through consensus. Each study was assigned one of three quality ratings (Good, Fair, Poor). Study quality ratings were based on the LMBP™ scoring methods described previously [6]. A study was included if it were considered to provide valid and useful information, met the PICO study inclusion criteria above, and evaluated a specific intervention/practice with at least one finding for a relevant outcome measure reported in a form which was useful for statistical analysis. Search result records that did not meet the inclusion criteria (i.e., not considered studies, not including a practice/outcome of interest), were excluded from further review. Studies not meeting the LMBP™ study quality criteria (Fair or Good quality rating) were also excluded. The Expert Panel then assigned one of three effect size ratings (Substantial, Moderate, or Minimal/None) for each study. Full tables and evaluation information from the systematic review are available in Appendix E.

3.0 Evidence review synthesis and results

Complete information related to the search results is provided in Figure 2. Of the total of 1,376 separate bibliographic records screened for assay selection, serial testing, assay threshold, and POCT, a full text review of 68 potentially eligible studies was conducted. A total of 24 published and 2 unpublished studies (Storrow 2015 [42] was unpublished at time of evaluation) were deemed acceptable for inclusion in this evidence review which totalled 13 published studies on assay selection, 8 published studies and 1 unpublished study on assay threshold, 5 published studies and 2 unpublished studies on serial testing, and 11 published studies on POCT. Six published studies and 1 unpublished study contained data evaluating two practices and two contained data evaluating three practices. While the data presented in Figure 2 is inclusive of all studies included for evaluation, several studies were removed for failing to meet a specific evaluation criterion or due to cTn assay issues. These instances are explained later in this manuscript. Analyses and characterizations of included and excluded studies are provided in Appendix D.

“Body of Evidence” summary tables for each practice are provided for each practice below, including abstracted and standardized information, study quality ratings, and bibliographic reference information. Detailed evidence tables comprise Appendix E.

3.1 Assay selection practice effectiveness evidence

Table 1 summarizes information on 13 published studies that comprise the practice effectiveness body of evidence for comparing use of a cTn assay alone versus use of a multi-biomarker approach. Five studies were eliminated because they used early generation cTn assays [43][44][45][46][47] as evaluated by the Expert Panel. Four studies [48][49][50][65] were disqualified for failing to meet quality criteria. The remaining four studies were rated as “Fair” with “Minimal/None” effect [39] or “Good” with “Moderate” effect [26][38][40]. Two studies [26][40] included a marginal number (7% of the total population) of STEMI patients, which were included in the analysis as supporting material.

The RATPAC study [38], a randomized controlled trial, indicates that myoglobin and CK-MB did not add to the diagnostic value provided by a contemporary, sensitive cTnI assay. Eggers et al. [39] showed a similar result, that CK-MB and in particular, myoglobin, did not offer additional diagnostic value when added to a sensitive cTn assay. This evidence
suggests minimal to no benefit from measuring any biomarkers in lieu of or addition to use of cTn to diagnose NSTEMI because the increase in diagnostic sensitivity comes at a similar or greater decrease in specificity. As displayed in Figure 3 (the cumulative assay specificity at 95% sensitivity by sampling time), the sensitivity and specificity from use of cTn far exceeds that of CK-MB or myoglobin at all-time points. Two other studies [26][40] showed a similar result but generalizability is limited by the inclusion of STEMI patients in their population. The findings are in agreement with the consensus recommendation that cTn is the preferred biomarker for MI diagnosis from the 2012 Global Task Force [1].

3.2 Assay threshold practice effectiveness evidence

Information on the eight published studies and one unpublished study of practice effectiveness comparing use of the 99th percentile diagnostic threshold with other possible thresholds (% CV concentrations, LLOD, or ROC area) is summarized in Table 2. A subset of data from an earlier publication [51], from which STEMI patients have been removed, leaving only NSTEMI subjects was provided for analysis.

Of the eight published studies, one study [43] was eliminated because it used an early generation cTn assay. Three studies [37][52][53] were eliminated, after a thorough review, because they did not directly address the diagnostic threshold question. One study [41] was rated as “Fair” quality with “Minimal/None” effect and three published studies [38] [51][54] and one unpublished study [42] were rated “Good” quality with a “Moderate” effect size, for a total of five studies included in the practice effectiveness body of evidence.

Mills et al. [51] provided a re-analysis of previously reported data that shows significantly (p<0.001) improved clinical outcomes from the use of a lower diagnostic threshold of cTn very near the 99th percentile. Here the cTnI assay implemented was reformulated to achieve higher sensitivity, which allowed a lowering of the cutoff at the 10% CV threshold from 0.2 ng/mL (old assay) to 0.05 ng/mL (new assay).

Data from the RATPAC study [38] were re-analyzed to examine various diagnostic cutoffs for the Stratus CS (Siemens Healthcare Diagnostics, Camberley, UK) including 0.07 ng/mL (99th percentile), 0.15 ng/mL (reported ROC cutoff [55]), and use of an arbitrary high cutoff, 0.30 ng/mL. (Re-analysis is not shown.) The 99th percentile cutoff for the cTnI diagnostic performance at presentation for the RATPAC cohort was superior to use of either the ROC decision point (65.2% sensitivity) or the higher cutoff (43.9% sensitivity). Use of the 99th percentile allowed for diagnosis of AMI in 82% (81.8% sensitivity) of RATPAC subjects at presentation, which was significantly more effective than either the ROC cutoff of 0.15 ng/mL (p<0.05) or the 0.30 ng/mL (p<0.0001) cutoff.

The effect of threshold selection was also examined with data from an unpublished multicenter study [42] conducted in urgent care settings. Patients (n=1, 929) presenting with signs and symptoms of acute cardiac ischemia were evaluated using the AccuTnI+3 assay (Beckman Coulter, Inc.) at threshold cut-offs corresponding to the 10% CV (0.04 ng/mL), 15% CV (0.03 ng/mL), 20% CV (0.02 ng/mL) and the 99th percentile URL threshold (0.02 ng/mL). The study findings demonstrated a progressively increasing sensitivity through the range of threshold values with slight loss in specificity, which was recovered through serial
sampling. The PPV decreased from 55.1% to 47.0% at baseline for the 10th CV and 99th percentile concentrations, respectively, while the NPV increased from 97.9% to 98.6%.

Our findings are consistent with recommendations from the Global Task Force that the 99th percentile of a reference control population should be utilized for diagnosis of AMI [1].

3.3 Serial testing practice effectiveness evidence

Five published studies (Storrow [42] was unpublished at time of review) and two unpublished studies, as shown in Table 3, had outcomes of interest and were reviewed. Two of the published studies were rated “Fair” and three were rated “Good”. Four of these studies [26][38][40][42][56] had a “Moderate” effect size and one[39], had a “Minimal/None” effect size rating. The two unpublished studies, [42] and (Christenson 2014), were rated “Good” with a “Moderate” effect size.

Serial sampling increases diagnostic accuracy as evidenced by the studies analyzed. A subset of these assay improvements are illustrated with Figure 4, showing increases in sensitivity and specificity with serial sampling.

Comparing the diagnostic sensitivity at presentation with that of the peak cTnI level in the RATPAC data set demonstrated significant diagnostic improvement from serial sampling (p<0.001) [38].

Kelly et al. [56] used a sensitive cTn assay for measurement at enrollment and in later serial samples. Of 129 patients with a final diagnosis of AMI (13.6% of the population), 99 had cTnI values above the 99th percentile value at enrollment for a diagnostic sensitivity of 76.7% (95% CI 68.5–83.7%). For the 823 patients who ruled-out for AMI, 53 subjects had cTnI values exceeding the 99th percentile at enrollment; specificity 93.6% (95% CI 91.7–95.1%). A subgroup of 109 MI patients from this cohort had serial cTnI measurements. The cTnI sensitivity for these patients at presentation was similar to the overall cohort at 72.5% (95% CI 63.1–80.6). However, in later serial samples, all of these patients were diagnosed with MI at 100% sensitivity, while the specificity of this subgroup remained high at 92.4% (95% CI 89.9–94.4%). These data suggest that a single troponin measurement at ED presentation, cTnI in this trial, has insufficient diagnostic sensitivity for AMI diagnosis, and that serial biomarker measurement is required to achieve appropriate high sensitivity [56].

A pattern of increasing NPV and decreasing PPV was observed in a multi-center study of 1818 patients presenting with possible AMI [40]. Measurements of a sensitive cTnI were obtained at presentation, 3h and 6h time points. The NPV increased from 94%, to 95.3% to 96.0 over the hours of serial sampling, while the PPV decreased from 82.0%, to 79.3%, to 78.7% over this period; neither the changes in NPV nor PPV were significant. Although this cohort included 127 STEMI patients, which is a limitation of this study, the STEMI patients represented a small proportion (7%) of the total cohort. These patterns were supported by a study conducted by Eggers et al. [39] which included 197 patients, 27 (14%) of which were diagnosed with MI. Cumulative NPV increased from admission (97%) through 2 hour (99%) and 3 hour (100%) sampling. Collinson et al. [38] showed increasing sensitivity (84.5% to
94.1%), specificity (97.6% to 98.4%), and NPV (98.7% to 99.9%) with a decreasing PPV (74.7% to 55.2%) with serial sampling taken at 90 minutes after presentation.

The contribution of serial sampling was also examined in a multicenter study [42] of 1,929 patients presenting acutely to urgent care settings with signs and symptoms consistent with acute cardiac ischemia. Serial samples were collected at enrollment, 1–3h, 3–6h, 6–9h; cTnI was quantified in these samples using a contemporary sensitive cTnI method. Of the 1,929 subjects, 253 (13%) were diagnosed as having acute MI. Of these, 1,504 (78%) were elevated above the 99th percentile at enrollment and demonstrated an increasing or decreasing temporal cTnI pattern at 3–6h. However 13% of the AMI patients did not show a cTnI elevation in the enrollment sample and the diagnosis may have been missed were it not for serial sampling. On the other hand, for the 1,676 cases for which AMI was ruled out, 20% did not demonstrate a rising or falling cTnI pattern at 3–6h. Of these non-AMI subjects, 11% had values above the cTnI cutoff at enrollment and may have received the incorrect diagnosis were it not for serial sampling. Importantly, because of the large number of non-AMI patients relative to AMI patients, the proportion of patients diagnosed with AMI when enrollment cTnI values were elevated was only 55% (PPV); i.e. 45% of patients with elevated cTnI enrollment values were non-AMI and would possibly have undergone unnecessary work-up or treatment if not for serial sampling.

An unpublished multicenter study (Christenson 2014) included 619 patients presenting to the ED with signs and symptoms of AMI. Serial samples were collected at presentation, 6h (± 3h) and 12h (± 3h). Cardiac troponin I was quantified with a contemporary sensitive assay. Of the 619 patients, 109 (17.7%) were diagnosed as having NSTEMI. Fifty-three (49%) of these AMI patients had cTnI values below the 99th percentile normal value of 40 ng/L on presentation. Some or all of these 53 patients may not have received therapy known to benefit NSTEMI patients had serial sampling not been performed. Seventy-one (65.1%) of the AMI patients showed a rise and/or fall of ≥20 ng/L. Of the remaining 510 patients (82.3%) who ruled-out for AMI, 28 (5.5%) had baseline values above the 99th percentile of normal in the enrollment sample. Only two of the rule-out patients (0.4%) had a change of ≥20 ng/L; the remaining 508 non-AMI patients had a change <20 ng/L. The PPV of cTnI for diagnosing AMI in the presentation sample was 65.4%. Thus, if the cTnI value at enrollment presentation were used to rule-in or rule-out AMI, a large proportion of patients may have been diagnosed as having AMI and could have undergone unnecessary work-up or treatment if not for serial sampling.

Serial sampling allows more accurate diagnostic use of cTn. These findings are supported by the consensus based guidance from the Global Task force that recommends serial sampling for diagnosis of MI [1].

### 3.4 Point of care testing practice effectiveness evidence

Data on the eleven published studies that comprised the practice effectiveness body of evidence comparing use of a cTn POCT with the use of a central laboratory test are summarized in Table 4. Of the eleven studies, three [32][57][58] were excluded for failing (i.e. receiving a 0 score) on at least one of the study quality rating criteria. Of the included studies, three were rated “Good” quality and five were rated “Fair” quality. The effect size of
two studies that measured TAT [41][59] were rated “Substantial”, and one study that measured length of stay (LOS) [59] was rated “Minimal/None”. The results are suggestive of decreased TAT with POCT practice implementation.

For the diagnostic accuracy outcome effect size, one study [61] documented an adverse effect rated “Substantial”, one study [59]) had a favorable effect rated “Substantial”, one study [38] had a favorable effect rated “Moderate”, and three [43][60][62] had effect sizes rated “Minimal/None”. One study [37], rated “Fair” with “Moderate” effect size, found improvement in TAT, but there was not sufficient evidence to evaluate further. The evidence on the effect of the POCT practice on diagnostic accuracy was inconsistent, and no recommendation can be made for or against this practice at this time.

4.0 Additional considerations

4.1 Feasibility of implementation

The data from this evidence review are consistent with the consensus based recommendations for use of cTn alone for the diagnosis of AMI, use of the 99th percentile as the AMI decision point for MI and the use of serial samples for evaluating AMI [1][4]. Use of cardiac markers to diagnose NSTEMI is standard practice [5][10]. Recommendations for the use of cardiac marker assays to diagnose NSTEMI have been part of the NACB [5], ACC/AHA guidelines [4] and ACCF/AHA guidelines [10] for many years; evidence evaluated as part of this review supports those guideline recommendations. Here we examined three questions finding that diagnosis of NSTEMI should be performed using a sensitive cTn assay with a diagnostic threshold at the 99th percentile of normal with serial samples to detect a rise/fall in the cTn level with time. No evidence of diagnostic improvement when using additional biomarkers was discovered, and therefore, the use of any additional biomarkers when a sensitive or contemporary sensitive cTn assay was used is not recommended. The evidence indicates that cTn measurement with a sensitive assay provides important diagnostic information for assessing patient risk of adverse clinical outcomes. Only minimal decrease in diagnostic specificity is incurred. The studies included in this evidence review focused on assessment of NSTEMI and were performed in EDs and cardiac short stay units, all settings in which the diagnosis is determined and where treatment is normally initiated.

4.2 Future research needs

Cost of care is also likely reduced but this was not evaluated in this review. While POCT cTn has the potential to reduce TAT [59][41], and LOS [60], the body of evidence regarding this practice was insufficient and inconsistent for conclusions. However, there is sufficient published evidence to determine the equivalency of POCT assays to laboratory assays and we recommend additional studies be conducted to measure the effect of POCT cTn implementation. Future studies should also establish an evidence base for the equivalency of cTI and cTn assays.
4.3 Study limitations

There were several limitations to the approach taken for evaluating the evidence within this review. The studies used in this review did not report all outcome measures consistently and utilized an array of assays which made quantitative data pooling in meta-analytic techniques impractical. Though the specific assays and outcomes were different, the assays used were appropriate at the time of study and comparisons were made which demonstrated the superiority of sensitive assays that are capable of good performance at the 99\textsuperscript{th} percentile of a reference control population. We did not have sufficient evidence to differentiate between sensitive and high-sensitivity cTn assays, which could potentially improve risk stratification. The outcome measures varied according to the specific interest of study, but in all cases the use of sensitive assays resulted in improved outcomes.

The LMBP\textsuperscript{TM} A6 Method is consistent with practice standards for systematic reviews, but all such methods are imperfect and include subjective assessments at multiple points that may produce bias. Rating study quality depends on consensus assessments that may be affected by such things as rater experience and the criteria used. As is the case with most systematic reviews, publication bias must be considered. The restriction to English language studies, while satisfying the requirement for multiple reviewers for each study, may also introduce bias.

5.0 Conclusions and best practices recommendations

The overall evidence included in this review indicates that a stand-alone, contemporary sensitive cTn assay performs significantly better than alternative biomarkers [26][38][39][40] and that the addition of those biomarkers to the cTn assay does not improve diagnosis [38][39]. In addition, recent economic evaluation studies found the new high sensitivity cTn assay without additional biomarkers is likely to be cost effective relative to the earlier conventional cTn assay or a combination of high sensitivity cTn assays with additional biomarker(s) [63][64].

Serial cTn sampling provides a means to reduce risk of misdiagnosis due to limitations in assays, variations in time of presentation after onset of symptoms and other external factors that can effect optimal assay performance. The evidence included in this review indicates that using a strategy for repeat testing results increases the diagnostic sensitivity [26][56] and NPV with a modest decrease in the PPV [39][56]. Our evaluation of the published [38][42] and unpublished (Christenson 2014) evidence also shows that this approach would more effectively diagnose NSTEMI patients for transition to appropriate care and would increase the efficiency of rule-out testing.

The overall strength of the body of evidence included in this review regarding POCT cTn was inconclusive. On one outcome (diagnostic accuracy effect size), one study [61] with “Substantial” effect rating, documented a negative effect direction while another study [59]), also “Substantial effect rating, documented a favorable effect. An additional three studies [38][43][62] documented effect sizes rated “Minimal/None”. Only one study of the eight included [37] found improvement in TAT. Evidence was insufficient and inconsistent for recommendation either for or against the use of point of care cTn testing.
Testing with high-sensitivity cardiac troponin assays, using the 99th percentile URL as the clinical diagnostic threshold for the diagnosis of NSTEMI and without additional biomarkers, is recommended. Also recommended is serial cardiac troponin sampling with one sample at presentation and at least one additional sample collected a maximum of 6 hours later to identify a rise or fall in the troponin level. Testing with high-sensitivity cardiac troponin assays, without other biomarkers, at presentation and again within 6 hours is the evidence-based best practice testing algorithm for most accurate and timely NSTEMI diagnosis.

**Acknowledgments**

Melissa Gustafson, Jim Derzon, Diana Mass, Paul Epner, Judy Berkowitz, Barbara Zehnbauer, Colleen Shaw, the LMBP™ Workgroup

Funding Source: CDC funding for the Laboratory Medicine Best Practices Initiative to Battelle Centers for Public Health Research and Evaluation under contract No. SP0700-00-D-3180, Delivery Order 0723.

**Appendix A: LMBP™ Cardiac Markers Systematic Review Expert Panel Members**

Fred S. Apple, PhD (Hennepin County Medical Center and University of Minnesota Medical School)

Farah M. Chowdhury, MD, MPH (Centers for Disease Control and Prevention, Division for Heart Disease and Stroke Prevention)

Robert Christenson, PhD, DABCC, FACB* (University of Maryland School of Medicine)

Michael C. Kontos, MD (Virginia Commonwealth University Medical Center)

Quaiser Mukhtar, PhD (Centers for Disease Control and Prevention, Division for Heart Disease and Stroke Prevention)

L. Kristin Newby, MD, MHS (Duke University Medical Center)

Alan B. Storrow, MD (Vanderbilt University Medical Center)

Milenko Tanasijevic, MD, MBA* (Brigham and Women’s Hospital & Harvard Medical School)

*Also Laboratory Medicine Best Practices Workgroup member

**Appendix B: LMBP™ Cardiac Markers Systematic Review Workgroup Members**

Robert H. Christenson, PhD, DABCC, FACB, Professor of Medical and Research Technology Director, Rapid Response Laboratories, University of Maryland School of Medicine, Baltimore, MD
Appendix C: LMBP™ Cardiac Markers Systematic Review Structured Search Databases and Terms

Serial Testing Search String
Serial Testing Search String


Assay Selection Search String


Assay Threshold Search String

Appendix D: LMBP™ Cardiac Markers Systematic Review Study
Characterizations and Analyses

<table>
<thead>
<tr>
<th>Full Text Review Studies</th>
<th>Included Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assay Selection</td>
<td>Five studies were disqualified that reported data from early generation cTn assays [43][44][45][46][47] as evaluated by the Expert Panel. Four studies [48][49][50][65] were disqualified for failing to meet quality criteria. Published studies included: 13 total − 5 − 4 = 4</td>
</tr>
<tr>
<td>13 published studies</td>
<td></td>
</tr>
<tr>
<td>Assay Threshold</td>
<td>One study [43] was disqualified that reported data from an early generation cTn assay. Three studies [37][52][53] were disqualified for failing to directly address the diagnostic threshold question. Published studies included: 8 total − 1 − 3 = 4</td>
</tr>
<tr>
<td>8 published studies</td>
<td></td>
</tr>
<tr>
<td>1 unpublished study</td>
<td>One study [42] was rated “Good” quality with a “Moderate” effect size. (Study [42] also reported a serial sampling analysis.)</td>
</tr>
<tr>
<td>Serial Sampling</td>
<td>Two studies were rated “Fair” [39][56] and three were rated “Good” [26][38][40][42]. Four studies [26][38][40][42][56] had a “Moderate” effect size and one [39], had a “Minimal/None” effect size rating. Published studies included: 5 total − 0 = 5</td>
</tr>
<tr>
<td>5 published studies</td>
<td></td>
</tr>
<tr>
<td>2 unpublished studies</td>
<td>Two studies [42][Christenson, 2014] were rated “Good” with a “Moderate” effect size. (One study [42] also reported an assay threshold analysis.)</td>
</tr>
<tr>
<td>Point of Care Testing</td>
<td>Three studies [32][57][58] were excluded for failing to meet quality criteria. Other studies are reported as follows: One study = Good/Moderate [38]</td>
</tr>
<tr>
<td>11 published studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 studies = Good/Minimal [43][60]</td>
</tr>
</tbody>
</table>
Full Text Review Studies\(^1,2\) & Included Studies\(^2,3\) \\
\(\Delta\) Studies = 2, Fair/Substantial Pos [41][59]; 1, Fair/Substantial Neg [61] \\
1 Study = Fair/Moderate [37] \\
1 Study = Fair/Minimal [62] \\
Graphically represented as follows:

| Substantial (3) | \(\bigcirc\bigcirc(-\bigcirc)\) |
| Moderate (2) | \(\bigcirc\) \(\bigcirc\) |
| Minimal/None (3) | \(\bigcirc\) \(\bigcirc\) \(\bigcirc\) |

Eight studies reported inconsistent effects and direction. No recommendation was made.

Published studies included:
11 total − 3 = \(8\)

Total Studies Included = \(15\)^1

\(^1\) In this particular instance, authors use “studies” to denote “reports” or “articles.”

\(^2\) While Figure 2 is inclusive of all studies included for evaluation, several studies were removed for failing to meet a specific evaluation criterion or due to cTn assay issues.

\(^3\) Later, authors use “studies” to denote practice “analyses”: For example, “A total of 24 published and 2 unpublished studies (Storrow 2014 [42] was unpublished at time of evaluation) were deemed acceptable for inclusion.” See section “3.0 Evidence review synthesis and results” (1) Six published studies contained data evaluating two practices.
Math: \(15 + 6\) extra analyses = 21 analyses
(2) One unpublished study contained data evaluating two practices.
Math: \(21 + 1\) extra analysis = 22 analyses
(3) Two published studies contained data evaluating three practices.
Math: \(22 + 4\) extra analyses = 26 Total Analyses

Published Studies Included = \(4 + 4 + 5 = 13\)
Unpublished Studies Included = \(3 - 1\) (2 analyses from same reference study) = \(2\)
Total Studies Included = \(13 + 2 = 15\)^1

Appendix E: LMBP™ Cardiac Markers Systematic Review Evidence Tables

Laboratory Medicine Best Practices Cardiac Marker Body of Evidence Tables

<table>
<thead>
<tr>
<th>Practice: Assay Selection</th>
<th>Study</th>
<th>Practice</th>
<th>Measures</th>
<th>Results</th>
<th>Total</th>
<th>Rating</th>
<th>Effect Size Rating</th>
<th>Overall Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published</td>
<td>Chang 2010 *</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>Fair</td>
<td>Minimal/None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collinson 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>Good</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eggers 2004</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>Fair</td>
<td>Minimal/None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Engel 2009 *</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>Fair</td>
<td>Minimal/None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hsu 2000</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>Poor</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jernberg 2000*</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>Fair</td>
<td>Minimal/None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jarlander 2000</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>Poor</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keller 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Good</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keller 2009</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Good</td>
<td>Moderate</td>
<td></td>
</tr>
</tbody>
</table>

Clin Biochem. Author manuscript; available in PMC 2016 September 22.
<table>
<thead>
<tr>
<th>Practice: Assay Selection</th>
<th>Study</th>
<th>Practice</th>
<th>Measures</th>
<th>Results</th>
<th>Total</th>
<th>Effect Size Rating</th>
<th>Overall Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCord 2001 *</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>Fair</td>
<td>Minimal/None</td>
</tr>
<tr>
<td>Meune 2011</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Quilici 2004</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>Poor</td>
<td>Poor</td>
</tr>
<tr>
<td>Straface 2008 *</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>Fair</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

* Early generation cTn assay; excluded from analysis

**Abbreviations**

- ACC: American College of Cardiology
- ACS: Acute coronary syndrome
- AHA: American Heart Association
- AMI: Acute myocardial infarction
- AUC: Area under the receiver-operating-characteristic curve
- CDC: U.S. Centers for Disease Control and Prevention
- CI: Confidence Interval
- CK-MB: Creatine kinase – myoglobin
- cTn: Cardiac troponin
- cTnI: Cardiac troponin subtype I
- cTnT: Cardiac troponin subtype T
- CV: Coefficient of variation
- ECG: Electrocardiogram
- ED: Emergency department
- ESC: European Society of Cardiology
- h: Hour
- ICD: International Classification of Diseases
- LMBP: Laboratory Medicine Best Practices Initiative
- MI: Myocardial Infarction
- NSTEMI: Non-ST-segment elevation myocardial infarction
- PICO: Population, Intervention, Comparison, Outcome
POCT  Point of care test
QI    Quality improvement
ROC  Receiver operator curve
STEMI Myocardial infarction with ST elevation
ST Segment In an ECG, the isoelectric line after the QRS complex (Q, R, and S waves representing ventricular polarization) which represents phase 2 cardiac action potential
URL  Upper reference limit

References


12. Bonham M, Miller S. Clinical comparison of 99\textsuperscript{th} percentile and 10% coefficient of variation cutoff values for four commercially available troponin I assays. LABMEDICINE. 2009; 40(8): 470–3. DOI: 10.1309/LMFB0C68TIZGVHOHV


Figure 1.
LMBPTM QI Analytical Framework: Cardiac Biomarkers
Figure 2.
Systematic Review Flow Diagram
Figure 3.
Cumulative Assay Specificity at 95% Sensitivity by Sampling Time
Figure 4.
Serial Testing Evaluation of cTn Assays
### Table 1

<table>
<thead>
<tr>
<th>Practice: Assay Selection</th>
<th>Study Quality Rating</th>
<th>Effect Size Rating</th>
<th>Overall Strength of Body of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chang 2010&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Collinson 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Eggers 2004</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Engel 2007&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Hsu 2000</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Jernberg 2000&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Jurlander 2000</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Keller 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Keller 2009</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>McCord 2001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Meune 2011</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Quilici 2004</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Straface 2008&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

<sup>a</sup> Early generation cTn assay; excluded from analysis

3 Studies = Good/Moderate
1 Study = Fair/Moderate
5 Studies = Fair/Minimal
4 Studies = Poor Excluded
### Table 2

Assay threshold practice effectiveness evidence

<table>
<thead>
<tr>
<th>Practice: Assay Threshold</th>
<th>Study</th>
<th>Practice</th>
<th>Measures</th>
<th>Results</th>
<th>Total</th>
<th>Rating</th>
<th>Effect Size Rating</th>
<th>Overall Strength of Body of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apple 2006</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>Fair</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Body 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>Good</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Collinson 2012</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>Good</td>
<td>Moderate</td>
<td>6 Studies = Good/Moderate</td>
</tr>
<tr>
<td>Kontos 2004</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>Good</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Mills 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>Good</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Morrow 2001</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>Good</td>
<td>Minimal/None</td>
<td>1 Study = Good/Minimal</td>
</tr>
<tr>
<td>Mozina 2010</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>Fair</td>
<td>Minimal/None</td>
<td></td>
</tr>
<tr>
<td>Straface 2008&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>Good</td>
<td>Moderate</td>
<td>1 Study = Fair/Moderate</td>
</tr>
<tr>
<td>Unpublished</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storrow 2014&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>Good</td>
<td>Moderate</td>
<td>1 Study = Fair/Minimal</td>
</tr>
</tbody>
</table>

<sup>a</sup> Early generation cTn assay; excluded from analysis

<sup>b</sup> Storrow study data were analyzed as unpublished prior to publication in 2015 [42].
### Table 3

Serial testing practice effectiveness evidence

<table>
<thead>
<tr>
<th>Practice: Serial Sampling</th>
<th>Study Quality Rating</th>
<th>Effect Size Rating</th>
<th>Overall Strength of Body of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Published</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eggers 2004</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Collinson 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Kelly 2011</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Keller 2011</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Keller 2009</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Unpublished</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christenson 2014</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Storrow 2014</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

5 Studies = Good/Moderate
1 Study = Fair/Moderate
1 Study = Fair/Minimal/None

b Storrow study data were analyzed as unpublished prior to publication in 2015 [42].
# Table 4

Point of care testing practice effectiveness evidence

<table>
<thead>
<tr>
<th>Practice: Point of Care Testing</th>
<th>Study Quality Rating</th>
<th>Effect Size Rating</th>
<th>Overall Strength of Body of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apple 2006</td>
<td>1 2 1 2 6 Fair</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Bock 2008</td>
<td>1 2 1 2 6 Fair</td>
<td>Substantial (Neg)</td>
<td></td>
</tr>
<tr>
<td>Caragher 2002</td>
<td>2 2 2 1 7 Fair</td>
<td>Substantial</td>
<td></td>
</tr>
<tr>
<td>Collinson 2012</td>
<td>2 2 2 3 9 Good</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Di Serio 2009</td>
<td>1 2 2 0 5 Fair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hallani 2005</td>
<td>0 2 1 2 5 Fair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee-Lewandrowski 2011</td>
<td>2 2 2 1 7 Fair</td>
<td>Minimal/None</td>
<td>1 Study = Good/Moderate</td>
</tr>
<tr>
<td>Loten 2010</td>
<td>2 2 1 3 8 Good</td>
<td>Minimal/None</td>
<td>3 Studies = Fair/Substantial</td>
</tr>
<tr>
<td>Mozina 2010</td>
<td>1 2 2 1 6 Fair</td>
<td>Substantial</td>
<td>1 Study = Fair/Moderate</td>
</tr>
<tr>
<td>Singer 2005</td>
<td>0 2 2 1 5 Fair</td>
<td></td>
<td>1 Study = Fair/Minimal</td>
</tr>
<tr>
<td>Straface 2008</td>
<td>2 2 2 2 8 Good</td>
<td>Minimal/None</td>
<td>3 Studies = Excluded</td>
</tr>
</tbody>
</table>