

OPEN

# Early-Onset Neonatal Sepsis: Still Room for Improvement in Procalcitonin Diagnostic Accuracy Studies

Claudio Chiesa, MD, Lucia Pacifico, MD, John F. Osborn, PhD, Enea Bonci, PhD,  
Nora Hofer, MD, and Bernhard Resch, MD

**Abstract:** To perform a systematic review assessing accuracy and completeness of diagnostic studies of procalcitonin (PCT) for early-onset neonatal sepsis (EONS) using the Standards for Reporting of Diagnostic Accuracy (STARD) initiative.

EONS, diagnosed during the first 3 days of life, remains a common and serious problem. Increased PCT is a potentially useful diagnostic marker of EONS, but reports in the literature are contradictory. There are several possible explanations for the divergent results including the quality of studies reporting the clinical usefulness of PCT in ruling in or ruling out EONS.

We systematically reviewed PubMed, Scopus, and the Cochrane Library databases up to October 1, 2014. Studies were eligible for inclusion in our review if they provided measures of PCT accuracy for diagnosing EONS. A data extraction form based on the STARD checklist and adapted for neonates with EONS was used to appraise the quality of the reporting of included studies.

We found 18 articles (1998–2014) fulfilling our eligibility criteria which were included in the final analysis. Overall, the results of our analysis showed that the quality of studies reporting diagnostic accuracy of PCT for EONS was suboptimal leaving ample room for improvement. Information on key elements of design, analysis, and interpretation of test accuracy were frequently missing.

Authors should be aware of the STARD criteria before starting a study in this field. We welcome stricter adherence to this guideline. Well-reported studies with appropriate designs will provide more reliable information to guide decisions on the use and interpretations of PCT test results in the management of neonates with EONS.

(*Medicine* 94(30):e1230)

## INTRODUCTION

Early-onset neonatal sepsis (EONS), diagnosed during the first 3 days of life, is a leading cause of morbidity and mortality among infants.<sup>1,2</sup> Currently, criteria for EONS usually include documentation of infection in a newborn infant with a serious systemic illness in which noninfectious explanations for

the abnormal pathophysiologic state are excluded or unlikely.<sup>3</sup> However, even culture is not free from error because it can be falsely sterile, as suggested by postmortem cultures,<sup>4</sup> or because of the low yield caused by insufficient sample volumes, intermittent, or low-density bacteremia.<sup>5</sup> Antibiotic treatment prior to blood culture may further reduce the diagnostic performance of blood culture.<sup>5</sup> Yet, clinical signs may be minimal and are similar to those of various noninfectious processes. Hence, antibiotics are often started empirically in infants with perinatal risk factors or clinical signs suggestive of bacterial infection.

The availability of a laboratory test to accurately and more rapidly identify septic neonates than is done by the isolation of microorganisms from body fluid specimens would be of considerable value in improving the outcome of this challenging clinical problem, and in minimizing unnecessary treatment of uninfected patients during the immediate postnatal period. Many laboratory tests including various leukocyte indices and acute-phase proteins have been recommended for the evaluation of suspected infection in the neonate.<sup>3</sup> However, the inability of any single laboratory test to provide rapid, reliable, and early identification of infected (and, as importantly, noninfected) neonates has led to a search for other diagnostic markers.<sup>3</sup> Among those evaluated in recent years has been procalcitonin (PCT).<sup>6</sup>

In 1994, Dandona et al<sup>7</sup> showed that after the injection of endotoxins in normal human volunteers, the PCT concentration was undetectable at 0, 1, and 2 hours, but was detectable at 4 hours peaking at 6 hours, and maintaining a plateau through 8 and 24 hours. Because of this rapid response, the body of literature investigating PCT in adults and children with sepsis, as an attractive alternative to C reactive protein (CRP), has grown rapidly over the last 2 decades,<sup>6,8,9</sup> providing insight as well as posing questions regarding the potential use of PCT for the diagnosis of sepsis. However, clinicians have been less familiar with the potential use of PCT measurements for clinical purposes in neonatal patients,<sup>3</sup> in particular those presenting with EONS.<sup>6</sup> Increased PCT is a potentially useful diagnostic marker of EONS, but reports in the literature are contradictory.<sup>6</sup>

In contrast to CRP, reports in the literature have shown that PCT is an early and specific marker of neonatal sepsis, confirming the importance of the latter in excluding infection shortly after birth.<sup>10</sup> However, falsely increased PCT concentrations have been reported for critically ill newborns presenting with “apparently” noninfectious conditions.<sup>11–16</sup> Although these studies argued for the lack of PCT specificity for the diagnosis of sepsis in neonates, their conclusions should be interpreted with caution. First, arbitrary cut-offs were used to differentiate infectious and noninfectious clinical conditions. For instance, in the report by Lapillonne et al,<sup>16</sup> uninfected neonates (mean postnatal age, 2.3 days) were deemed to have high serum PCT concentrations on the basis of a surrogate cut-off level originally established in children admitted to a pediatric intensive care unit. Second, interpretation on the

Editor: Marsha Medows.

Received: November 4, 2014; revised: May 5, 2015; accepted: July 6, 2015. From the Institute of Translational Pharmacology, National Research Council (CC), Department of Pediatrics and Child Neuropsychiatry (LP), Department of Public Health and Infectious Diseases (JFO), Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy (EB); and Research Unit for Neonatal Infectious Diseases and Epidemiology, Division of Neonatology, Department of Pediatrics and Adolescent Medicine, Medical University of Graz, Graz, Austria (NH, BR).

Correspondence: Claudio Chiesa, Institute of Translational Pharmacology, National Research Council, Via Fosso del Cavaliere, 100, 00133 Rome, Italy (e-mail: [claudio.chiesa@ift.cnr.it](mailto:claudio.chiesa@ift.cnr.it)).

The authors have no conflicts of interest to disclose.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ISSN: 0025-7974

DOI: 10.1097/MD.0000000000001230

use of PCT was complicated by diverse study populations.<sup>10</sup> Heterogeneity not only within the study group, but also within categories defined as “sepsis,” “distress,” “infected,” “respiratory distress,” or even “hemodynamic failure,” has been huge.<sup>10</sup> Third, the PCT response was assessed in neonates with wide-ranging differences in postnatal age (hours to weeks),<sup>10</sup> without consideration of the gestational age and birth weight of the baby.<sup>3</sup> However, failure to recognize gestational- and age-specific cut-off values over the first few days of life may confound the interpretation of what constitutes a negative and a positive PCT value for the diagnosis of EONS.<sup>17–21</sup> Fourth, PCT levels obtained from uninfected patients were not compared with PCT reference values for each time point of evaluation. This limitation makes it difficult to determine which neonatal factors may cause a significant deviation. Fifth, it is uncertain how the infectious state in the “uninfected” neonates was ruled out. In the report by Monneret et al,<sup>11</sup> high PCT concentrations were reported during the first 4 days of life in apparently “uninfected” newborns who presented with respiratory distress syndrome (RDS) of various etiologies. However, it would be of interest to know how many patients with severe RDS demonstrated clinical evidence of being “uninfected.” Finally, absent from these studies are data on how the interpretation of the PCT response in the uninfected as well as in the infected neonates might have been hampered by the severity of the underlying illnesses (and their extent of inflammatory reaction).<sup>6,22</sup> It is clear from the above that the evaluation of the clinical usefulness of PCT in ruling in or ruling out neonatal sepsis, in particular EONS, is dependent on study consistency. In this article, we therefore performed a systematic review assessing accuracy and completeness of PCT diagnostic studies for EONS including key elements of design, conduct, and analysis according to the Standards for Reporting of Diagnostic Accuracy (STARD) statement.<sup>23,24</sup>

## METHODS

Our systematic review was conducted, when possible, in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Ethical approval was not necessary for this review study.

### Data Sources

We systematically reviewed PubMed, Scopus, and the Cochrane Library databases up to October 1, 2014. The PubMed combined search term used was: (Procalcitonin OR PCT) AND (neonatal sepsis OR neonatal infections OR neonate). The search terms applied to the Scopus and the Cochrane Library were “Procalcitonin and neonate” and “Procalcitonin,” respectively. The bibliographies of relevant articles were also hand-searched.

### Study Selection Criteria

Studies were eligible for inclusion in our review if they provided measures of PCT accuracy for diagnosing EONS, defined by the National Institute of Child Health and Human Development and Vermont Oxford Networks as sepsis with onset at  $\leq 3$  days of age.<sup>25</sup> We excluded studies that used a PCT semiquantitative assay. Duplicate articles, conference abstracts, or studies written in languages other than English were also excluded.

### Data Extraction

Data extraction was performed independently by 2 authors and included the country of the research, year of publication,

journal, reference standard employed, the type of the study design, the number and specific characteristics of the patients in the septic and nonseptic groups (Table 1), and items related to the quality of the methods and reporting (listed below). Specific data regarding the PCT cut-off level used, the sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve for the diagnosis of EONS were also extracted (Table 1). In cases in which estimates of uncertainties around the observed values of sensitivities and specificities were not calculated using Wilson method,<sup>42</sup> we calculated them.

### Quality Assessment of the Included Studies

In 2003, the STARD statement was published in 13 biomedical journals.<sup>23,24,43</sup> The STARD initiative was developed in response to accumulating evidence of poor methodological quality. Therefore, a data extraction form based on the STARD checklist and adapted for neonates with EONS was used to appraise the overall quality of the included studies which is presented in Table 2. All articles were closely examined to gauge the extent to which they adhered to the STARD checklist by assigning a yes or no response to each item.

Twenty-five items makeup the STARD checklist. When the papers considered in this study were assessed, only those STARD items that have been empirically shown to have a potentially biasing effect on the results of diagnostic accuracy studies<sup>44–47</sup> and those items we deemed they may account for variation between studies in estimates of diagnostic accuracy for neonatal sepsis were evaluated. The following features, with corresponding STARD checklist item number, appear to have a possible biasing effect<sup>44–47</sup> or account for variation between diagnostic studies on neonatal sepsis: a description of the target population (STARD: item #3); a description of where (primary, secondary, or tertiary care setting) patients were recruited and where the test and the reference standard were performed (STARD: item #3); a description of how eligible subjects were recruited (ie, a description whether the study included: neonates who presented only symptoms suggesting sepsis; neonates who presented only maternal risk factors for infection; neonates in whom diagnosis of sepsis had been already established or excluded; a neonatal population selected from existing data files [STARD: item #4]); a description of whether recruitment involved either a consecutive series of patients with and without sepsis or a subselection (STARD: item #5); a description of the data collection process (STARD: item #6); a description of the reference standard for diagnosis (or exclusion) of neonatal sepsis and its rationale (ie, a description of whether: a single or composite reference standard [reflecting 2 or more reference standards] was used to identify all newborns positive to the test for sepsis and verify the index test results in infected babies; a reference standard [along with pertinent details: a single or composite reference standard reflecting 2 or more reference standards or noninfectious clinical conditions—STARD: item #18] was used to exclude EONS and to verify the index test results in uninfected babies; the reference standard was independent of the index test [ie, whether the index test or its comparator formed part of the reference standard in neonates with or without sepsis—STARD: item #7]); a description of whether categories of results of the index test (including cut-offs) and the reference standard were defined before or after obtaining the results (STARD: item #9); information concerning the number and training of the persons executing and evaluating the index test and the reference standard (STARD: item #10); a description of whether or not the readers of the index tests and the reference standard were blind (masked) to

**TABLE 1. Characteristics of PCT Accuracy Studies (1998–2014) for Diagnosing Early ( $\leq 72$  hours)-Onset Neonatal Infection**

Country, Year Reference	Recruitment	Reference Standard in Infected Neonates	Reference Standard in Control Neonates	Sample Studied	PCT Cut-Off, ng/mL	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC
Italy, 1998 <sup>18</sup>	103 NICU preterm and term babies; 28, infected;	(1) A positive culture of $\geq 1$ blood specimen, and clinical signs of infection; or (2) $\geq 3$ categories of clinical signs and/or radiographic evidence of pneumonia, a positive sepsis screen, and historic and clinical risk factors for EONS	Babies with various types of distress who were apparently well within 48–72 hours of age	From birth to 48 hours of age	Upper limits of the reference sample over the first 48 hours of age	92.9 (77.4–98.0)	97.3 (90.8–99.3)	NA
France, 2002 <sup>26</sup>	75, uninfected; 109 preterm and term babies; 21, infected;	(1) Positive culture of blood and/or CSF; or (2) clinical signs, increased or decreased WBC, and positive culture of peripheral samples	Babies with transient distress, or prematurity	At birth	2.5	85.7 (65.4–95.0)	89.8 (81.7–94.5)	NA
Austria, 2003 <sup>27</sup>	88, uninfected; 68 NICU preterm and term babies with suspected sepsis; 41, infected;	(1) Positive blood culture; or (2) $\geq 3$ categories of clinical signs, positive sepsis screen and/or risk factors, and antibiotic therapy $\geq 7$ days	Negative blood culture, negative sepsis screen, and antibiotic therapy $\leq 3$ days	$\leq 12$ hours	$\geq 2$	82.9 (68.7–91.5)	59.3 (40.7–75.5)	NA
Italy, 2003 <sup>22</sup>	27, uninfected; 134 NICU preterm and term babies; 19, infected;	(1) Positive blood culture and clinical signs; or (2) $\geq 3$ clinical signs prompting $\geq 5$ days of antibiotic therapy, and historic and clinical risk factors for EONS	Symptomatic babies who had negative body fluid cultures, and were apparently well within 24–48 hours.	At birth	$\geq 14$ $\geq 1$	63.4 (48.1–76.4) 78.9 (56.7–91.5)	100 (85.7–100) 94.8 (89.1–97.6)	NA NA
Italy, 2004 <sup>28</sup>	115, uninfected; 35 NICU preterm babies; 27, infected; 8, uninfected;	Clinical signs of sepsis, positive hematologic findings, and positive blood cultures	Negative blood cultures, and no clinical/hematologic findings	24 hours	$\geq 100$	94.7 (75.4–99.1)	95.7 (90.2–98.1)	NA
Slovenia, 2007 <sup>29</sup>	25 NICU preterm and term babies with suspected sepsis;	(1) Positive cultures (blood, urine, deep soft-tissue, BAL, or CSF); or (2) clinical evidence of sepsis	Noninfectious clinical conditions (respiratory distress, low cardiac output syndrome, perinatal hypoxia, ileus, anemia)	48 hours Day 3 of life	$\geq 50$ 0.7	84.2 (62.4–94.5) 96.0 (80.5–99.3)	100 (96.8–100) 87.5 (52.9–97.8)	NA NA
Spain, 2007 <sup>30</sup>	8, infected; 17, uninfected; 238 preterm and term babies;	(1) Positive blood culture, and $\geq 3$ clinical signs of infection; or	Negative blood cultures, no clinical signs, and no antibiotic treatment	At birth*	$\geq 0.55$	75.4 (64.0–84.0)	72.2 (65.0–78.4)	0.65 (0.44–0.84)

Country, Year Reference	Recruitment	Reference Standard in Infected Neonates	Reference Standard in Control Neonates	Sample Studied	PCT Cut-Off, ng/mL	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC
Denmark, 2008 <sup>31</sup>	69, infected;  169, uninfected and asymptomatic, being admitted to hospital for prematurity, low birth weight, or $\geq 2$ risk factors for infection	(2) $\geq 3$ clinical signs of infection, $\geq 1$ laboratory finding of infection, and $\geq 2$ risk factors for infection  (1) Positive blood culture; or  (2) clinical signs and CRP $>5$ mg/dL	(1) Clinical signs, CRP $\leq 5$ mg/dL, and antibiotic therapy for 3 days; or  (2) clinical signs, but no antibiotic therapy	12–24 hours*  36–48 hours*	$\geq 4.7$  $\geq 1.7$	73.9 (62.5–82.8)  78.3 (67.2–86.4)	81.1 (74.5–86.3)  79.3 (72.6–84.7)	0.80 (0.74–0.87)  0.82 (0.75–0.89)
Italy, 2008 <sup>32</sup>	123 NICU preterm and term babies with at least one clinical sign suggesting EONS; 29, infected;  94, uninfected 149 NICU preterm and term babies at risk for infection;	(1) Positive culture or pneumonia, $\geq 2$ categories of clinical signs, and positive laboratory screen; or  (2) negative cultures, $\geq 2$ categories of clinical signs, and positive laboratory screen	Negative sepsis screen, or clinical signs for less than 24 hours without use of antibiotics	8 hours <sup>f</sup>  $\leq 12$ hours	$> 25$  $> 2.5$	69.0 (50.8–82.7)  47.4 (27.3–68.3)	86.2 (77.8–91.7)  94.6 (89.3–97.4)	0.83  0.77 (0.68–0.85)
Poland, 2008 <sup>33</sup>	19, infected;  130, uninfected  286 preterm and term babies;	Maternal history of clinical chorioamnionitis, and $\geq 3$ clinical signs, with or without positive blood cultures	NA	12–36 hours  36–60 hours  Cord blood	$> 34$  $> 15$  $> 1.22$	57.9 (36.3–76.9)  73.7 (51.2–88.2)  80.4 (66.8–89.3)	89.2 (82.7–93.5)  86.9 (80.1–91.7)  71.7 (65.7–77.0)	0.77 (0.67–0.85)  0.78 (0.66–0.87)  0.80
Italy, 2009 <sup>34</sup>	46, infected; 240, uninfected 42 NICU preterm babies;	Positive laboratory tests and/or clinical signs of infection	Healthy state	$\leq 48$ hours	NA	91.7 (64.6–98.5)	80.0 (62.7–90.5)	NA
France, 2011 <sup>35</sup>	12, infected; 30, apparently healthy 2151 preterm and term babies with risk factors for EONS;	(1) Positive blood or CSF culture; or  (2) clinical signs and $\geq 2$ abnormal laboratory results	No positive cultures of central samples; no clinical signs; no abnormal laboratory results	Cord blood	0.6	92.3 (75.9–97.9)	97.0 (96.2–97.6)	0.96 (0.95–0.98)
Turkey, 2011 <sup>36</sup>	26, infected;  2125, uninfected 74 preterm babies with PPRM;	Clinical signs of infection and/or radiographic evidence of pneumonia, a positive sepsis screen, maternal and perinatal risk factors for infection, with (39%) or without positive blood cultures	NA	Day 1 of life  Day 3 of life	1.74  1.8	75.0 (57.9–86.7)  87.5 (71.9–95.0)	85.7 (72.2–93.3)  85.7 (72.2–93.3)	NA  NA
Turkey, 2011 <sup>37</sup>	32, infected; 42, uninfected 260 NICU preterm and term babies;	Clinical signs of infection, with (39%) or without positive blood cultures	No evidence of clinical sepsis; no risk factors for infection; negative CRP results as well as blood cultures	At birth	$\geq 0.59$	48.5 (41.2–56.0)	68.5 (58.3–72.2)	NA
	171, infected; 89, uninfected			24 hours	$\geq 5.38$	83.0 (76.7–87.9)	88.8 (80.5–93.8)	NA

Country, Year Reference	Recruitment	Reference Standard in Infected Neonates	Reference Standard in Control Neonates	Sample Studied	PCT Cut-Off, ng/mL	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC
Iran, 2012 <sup>38</sup>	65 NICU preterm and term babies with clinical signs of sepsis or maternal risk factors for EONS; 49, infected; 16, uninfected	(1) Positive blood culture; or (2) ≥3 categories of clinical signs, and positive sepsis screen or maternal risk factors	Negative blood culture, <3 categories of clinical signs, and a negative sepsis screen	≤12 hours	≥1.7	77.6 (64.1–87.0)	81.3 (57.0–93.4)	NA
Switzerland, 2013 <sup>39</sup>	137 NICU preterm and term babies; 33, infected; 104, uninfected	(1) Positive blood or CSF cultures; or (2) ≥3 abnormal findings (including maternal risk factors, clinical signs, and conventional laboratory tests)	≤ 2 abnormal findings; and antibiotic therapy for 24–72 hours	24–36 hours	≥4.7	71.4 (57.6–82.2)	81.3 (57.0–93.4)	NA
France, 2014 <sup>40</sup>	2366 preterm and term babies with suspected EONS; 20, infected; 2346, uninfected	(1) Positive blood or CSF culture; or (2) biologic and clinical signs of sepsis; or (3) a pathogen germ in gastric aspirate	Neither clinical nor biologic signs of sepsis; negative gastric aspirates	Cord blood	0.6	95.0 (76.4–99.1)	87.4 (86.0–88.7)	NA
Austria, 2014 <sup>41</sup>	218 NICU preterm babies with risk factors for EONS; 30, infected; 188, uninfected	(1) Positive blood culture; or (2) ≥3 categories of clinical signs; or (3) ≥1 category of clinical signs, and ≥2 laboratory abnormalities	NA	Cord blood	0.235	80.0 (62.7–90.5)	86.2 (80.5–90.4)	0.83 (0.73–0.94)

AUC = Area under the curve, BAL = Bronchoalveolar lavage, CRP = C reactive protein, CSF = Cerebrospinal Fluid, EONS = Early-onset neonatal sepsis, NA = Not available, NICU = Neonatal Intensive Care Unit, PCT = procalcitonin, PPRM = Preterm Premature Rupture of Membranes, WBC = White blood cells.  
 \* The number of infected and uninfected babies who were sampled for PCT at each of the 3 time points is unknown.  
 † After suspicion of EONS.



**TABLE 2.** Quality of Reporting of PCT Accuracy Studies (1998–2014) for Diagnosing Early ( $\leq 72$  hours)-Onset Neonatal Infection

Category and Item No.	YES	NO
Methods-participants		
Describe the study population:		
1A. the inclusion and exclusion criteria,	17	1
1B. setting, and locations where data were collected.	15	3
Describe participant recruitment:		
2A. Was enrollment of patients based only on clinical signs suggesting infection?	6	12
2B. Were such patients consecutively enrolled?	4	2
2C. Was enrollment of patients based only on maternal risk factors for infection?	3	15
2D. Were such patients consecutively enrolled?	3	0
2E. Were patients identified by searching hospital records?	4	14
2F. Did the study include both patients already diagnosed with sepsis and participants in whom sepsis had been excluded?	5	13
Describe data collection:		
3. Was data collection planned before the index test and reference standard were performed (prospective study)?	9	9
Test methods		
Methods pertaining to the reference standard and the index test:		
4A. Was a composite reference standard used to identify all newborns with sepsis, and verify index test results in infected babies?	17	1
4B. Was a reference standard used to exclude sepsis?	15	3
4C. Was a composite reference standard used to identify all newborns without sepsis, and verify index test results in uninfected babies?	4	11
4D. Did the index test or its comparator form part of the reference standard?	10	8
5. Were categories of results of the index test (including cut-offs) and the reference standard defined after obtaining the results?	16	2
6. Did the study report the number, training and expertise of the persons executing and reading the index tests and the reference standard?	2	16
7. Was there blinding to results of the index test and the reference standard?	2	16
Statistical methods		
8. Describe the statistical methods used to quantify uncertainty (i.e. 95% confidence intervals)?	1	17
9. Describe methods for calculating test reproducibility	1	17
Results-participants and test results		
10A. Describe when the study was done, including beginning and ending dates of recruitment	15	3
10B. Did the study report clinical and demographic (postnatal hours or days, gestational age, birth weight, gender) features in those with and without sepsis?	14	4
10C. Did the study report distribution of illness severity scores in those with and without sepsis?	2	16
11. Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test.	4	14
12. Report a cross-tabulation of the results (including indeterminate and missing results) by the results of the reference standard; for continuous results report the distribution of the test results by the results of the reference standard.	3	15
Results-estimates		
13. Report measures of statistical uncertainty (i.e. 95% confidence intervals)	10	8
14. Report how indeterminate results, missing responses and outliers of index tests were handled	3	15
15. Report estimates of test reproducibility	4	14

the results of the other test (STARD: item #11); a description of the methods for calculating the precision around used measures of diagnostic accuracy (STARD: item #12), and the reporting of the range within which the true values were likely to lie (STARD: item #21); a description of the methods used for calculating test reproducibility (STARD: item #13), and the reporting of the estimates of test reproducibility (STARD: item #24); a description of the study population including septic and nonseptic neonates in which tests were executed (study period, clinical and demographic features, distribution of illness severity scores [STARD: items #14, #15, and #18]); a description of how many participants fulfilling inclusion criteria failed to undergo the index test or reference standard and the reasons of failing to do so (STARD: item #16); a cross-tabulation of the results of the index test by the results of the reference standard, and for continuous results, a description of the distribution of the test results by the results of the reference standard (STARD: item #19); and a description of how indeterminate results, missing data, and outliers of the index test were handled (STARD: item #22).

Adequate reporting of 3 key domains, that is, descriptions of participant recruitment, reference standard and index test,

and study population, was considered essential for capturing the global and integrative quality of reporting the accuracy of diagnostic tests for ruling in or ruling out EONS, and therefore they were split into various complementary items.

Two investigators (CC, BR) independently assessed the methodological quality of all eligible studies. The interrater agreement between these 2 reviewers was expressed as overall agreement percentage and more formally tested with the kappa statistic.

## RESULTS

### Study Selection

In Figure 1, we present the flow diagram as recommended by the PRISMA statement showing the process of the selection of the studies included in our review. Specifically, our search identified a total of 273 articles (after removing the duplicates). According to titles and abstracts, 74 of them were selected for further assessment. After a review of full-text, 18 were eligible for inclusion in our review.<sup>18,22,26–41</sup> All articles were identified as studies of diagnostic accuracy and stated this research objective in the introduction. Papers were published between

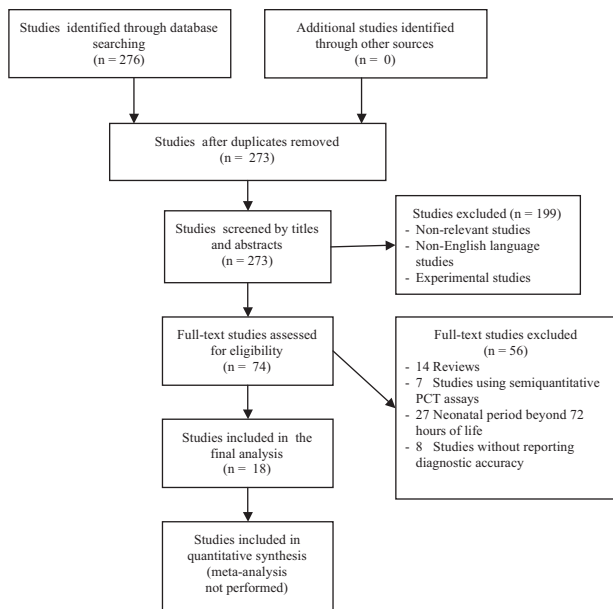


FIGURE 1. Schematic flow chart for the selection of studies.

1998 and 2014. Most studies (15/18) were performed in single perinatal centers.

### Characteristics of the Included Studies

In Table 1, we present the main characteristics of the 18 included studies. Half of the studies had a retrospective design, while the remaining were prospective. Fourteen studies included both preterm and term infants and the remaining 4 involved only preterm newborns. The total number of patients included in the 18 studies was 6547, of whom 680 had EONS and 5867 were uninfected. In all studies, the septic group consisted of neonates with culture proven and/or clinically diagnosed EONS. The nonseptic group consisted of ill neonates with other conditions that were hospitalized in the neonatal unit in all but one of these studies. There was considerable variation in the reference standard used for diagnosing (or excluding) EONS and thus verifying index test results.

### Diagnostic Accuracy of the Included Studies

Table 1 summarizes PCT diagnostic thresholds and accuracy measures such as PCT sensitivities and specificities for identifying (or excluding) EONS. There was over the first 72 hours of life wide variation in PCT sampling times, PCT cut-off values, and results in PCT diagnostic accuracy. The minimum and maximum PCT cut-off used was 0.235 and 100 ng/mL, respectively. The observed values of sensitivity ranged from 47.4% (95% confidence interval [CI] 27.3–68.3) to 100% (95% CI 67.6–100) and specificity from 35.3% (95% CI 17.3–58.7) to 100% (95% CI 96.8–100).

### Quality of Reporting of PCT Diagnostic Studies

We evaluated the 18 included studies for compliance with STARD guidelines. Agreement between the 2 reviewers was very good with an overall agreement percentage of 96%. The kappa statistic had a value of 0.93 (95% CI 0.91–0.95),

indicating a very good agreement. Discrepancies were rectified by consensus.

Table 2 summarizes the extent to which the 18 studies adhered to the STARD items. The reporting of individual items showed wide variation. Overall, the quality of reporting of PCT diagnostic accuracy studies on EONS over the last 2 decades was suboptimal. Information on key elements of design, conduct analysis, and interpretation of test accuracy were frequently missing.

## DISCUSSION

Studies of the PCT accuracy for the diagnosis of EONS have produced discrepant results in the past. Potential sources of the wide variation in PCT sensitivity and specificity include the lack of consistent PCT cut-off values regardless of the time of sampling, and differences in study sample sizes, patient demographic and clinical characteristics, and participant recruitment.

We therefore evaluated the quality of reporting of PCT accuracy studies for diagnosing EONS. The importance of describing how eligible subjects were identified cannot be overemphasized. Selection of a study population is critical in appraising a diagnostic test.<sup>48</sup> It is crucial to describe the population from which patients and patient controls originated, as it allows an assessment of the population “spectrum of disease.” Reported estimates of diagnostic accuracy may have limited clinical applicability in external populations (generalizability) if the spectrum of tested patients is not similar to the patients for whom the test will be used in practice.<sup>49,50</sup> Our analysis showed that while some PCT accuracy studies enrolled only neonates who were suspected of having the disease because of presenting symptoms, other studies recruited neonates who were (initially asymptomatic and) only at risk of developing the disease because of history of maternal risk factors. Yet, other studies included neonates already diagnosed with sepsis and participants in whom sepsis had been already excluded. Finally, there were also designs starting with: 2 separate selection processes to sample patients with sepsis and patients without sepsis (a sampling method producing biased estimates of accuracy known as “the diagnostic case-control” with limited spectrum);<sup>44,51</sup> nonconsecutive sampling of patients (a sampling method leading to the “limited challenge bias”);<sup>52</sup> retrospective data collection; and identification of patients by searching hospital records. These alternative study designs are likely to influence the spectrum of disease in the included patients, and estimates of diagnostic accuracy. In a study of evidence of bias and variation in diagnostic accuracy studies, Rutjes et al<sup>46</sup> found that the largest overestimation of accuracy was found in studies that included severe cases and healthy controls and in those in which 2 or more reference standards were used to verify index test results. The design features associated with overestimation of diagnostic accuracy were also nonconsecutive inclusion of patients, and retrospective data collection.<sup>46</sup> In contrast, the selection of patients on the basis of whether they had been referred for the index test or on the basis of previous test results, rather than on clinical symptoms, was significantly associated with lower estimates of accuracy.<sup>46</sup>

The spectrum of sepsis as well as the range of other conditions that occur in patients suspected of sepsis can vary from setting to setting, depending on what referral mechanisms were in play.<sup>53–55</sup> Thus, the accuracy of tests may vary between primary care and secondary or tertiary care.<sup>24</sup> Most PCT accuracy studies on EONS reported this issue properly by

specifying that the neonatal intensive care setting was the healthcare setting where data were collected.

Spectrum bias also results from differences in the severity of sepsis/EONS between populations. It may be that some of the variation in PCT sensitivity in the diagnosis of EONS among published reports might have reflected differences in baseline severity and risk status. Although individual clinical and demographic attributes of the septic and nonseptic neonatal populations were reported in most PCT diagnostic accuracy studies (14/18), proxy measures of morbidity such as gestational age and birth weight do not capture the overall morbidity status. Therefore, it is of greater concern that reporting of the distribution of scores of major measures of illness severity was poor in neonates with and without sepsis.

Differential verification bias is also a key issue in any diagnostic accuracy study. It occurs when some of the index test results are verified by one type of reference standard and other results by a different standard.<sup>49</sup> The effect of differential verification depends on the quality of the different reference tests used. Differential verification poses a problem if the reference standards differ in accuracy and if the choice of reference standard relates to the results of the index test.<sup>49</sup> This usually occurs when patients testing positive on the index test receive a more accurate reference standard than those testing negative. Previous studies that relied on 2 or more reference standards to verify the results of the index test reported estimates of diagnostic accuracy on average 60% higher than those encountered in studies that used a single reference standard.<sup>46</sup> The origin of this difference probably resides in differences between the reference standards in how they define sepsis, or in their quality.<sup>50</sup> Reference standards are not interchangeable. They may not have the same degree of error and may not identify the same segment of the disease spectrum. Thus, worrisome is the fact that in most of the included PCT diagnostic accuracy studies different reference standards were used to diagnose (or exclude) EONS and verify index test results.

A further step in the critical appraisal of PCT accuracy studies on EONS is whether the reference standard was independent of the index test and the comparator of the index test.<sup>48,56,57</sup> When the result of the index test or its comparator is used in establishing the reference standard, incorporation bias may occur. Unfortunately, in 10 (55.5%) of the 18 included studies, the comparator of the index test such as CRP test was also a component of the reference standard. In such situations, it is likely that the person interpreting the results of the comparator will have knowledge of the results of the other test (index test and reference standard).<sup>50</sup>

In order to make a valid comparison between the index test and the standard test, it is essential that the criteria (cut-off values, etc.) are defined before the start of the study. If the cut-offs are decided after the results are obtained, the likelihood that another study will replicate the findings is reduced. Apparently most studies (16/18) included cut-offs post hoc.

We also determined whether the interpretation of the index test or reference standard was influenced by knowledge of the results of the other test. Interpretation of the results of the index test may be influenced by knowledge of the results of the reference standard, and vice versa.<sup>48,50</sup> This may lead to inflated measures of diagnostic accuracy. Information about masking was reported in only 2 of the 18 evaluated reports.

Methods for calculating PCT test reproducibility and measures of test reproducibility were reported poorly, by only 1 and 4 of the 18 included articles, respectively. There may be a lack of understanding about the implications of poor

reproducibility on the overall utility of a test. Furthermore, the value of sensitivity and specificity estimates are unclear in the absence of precise information about test reproducibility.

Information regarding the number, training, and expertise of the persons involved in the execution, and reading of the index test and the reference standard was also among the least commonly reported items of the STARD guidelines. There may be a lack of understanding of the effects of the low level of expertise on the final outcome of the diagnostic accuracy of a test. On the other hand, knowledge of the numbers, training, and experience of operators would help to estimate the repeatability of the test results in different settings.

Measures of diagnostic accuracy will be biased if the result of the index test influences the decision to order the reference standard (an effect known as “partial verification bias”).<sup>48,58–64</sup> According to the STARD statement,<sup>24</sup> partial verification bias occurs in up to 26% of diagnostic studies.<sup>61</sup> It is therefore important to describe how many participants satisfying inclusion criteria failed to undergo index or reference tests and the reasons of failing to do so. This item was reported in a minority of the included publications (4/18). A flow diagram is highly recommended to illustrate the design of the study and provide the exact number of participants at each stage of the study;<sup>23,24</sup> only 2 of the 18 PCT diagnostic accuracy studies had a flow diagram.

Since the reported values of sensitivity, specificity, and area under the ROC curve are only estimates, measures of uncertainty (such as CIs) give readers a range within which true values may lie and indicate the precision of the diagnostic test.<sup>65</sup> CIs were reported in more than half (11/18) of the included studies. However, only one of the 18 studies adequately reported and cited the statistical methods used to quantify these estimates of diagnostic accuracy.

Diagnostic tests may report uninterpretable results for some patients, or call results uncertain, indeterminate, or intermediate. The frequency of such results can vary widely between tests.<sup>57</sup> These problems are sometimes not reported in diagnostic accuracy studies or ignored in analyses.<sup>66</sup> The frequency of these results, by itself, is an important clue of the overall utility of a test.<sup>67,68</sup> This item was reported in only 3 of the 18 studies.

Since the technology for existing tests is rapidly improving, it is important to report the actual dates when the study was performed. This will allow the reader to consider any technologic advances since the study was done. Fortunately, this information was provided in most publications.

In general, the results of our analysis of PCT diagnostic accuracy studies for EONS provide further evidence of the importance of design features in studies of diagnostic accuracy. Studies of the same test can produce different estimates of diagnostic accuracy depending on choices in design. We feel that our results should be taken into account by researchers when designing new PCT accuracy studies for diagnosis of EONS as well as by reviewers and readers who appraise these studies. In that vein, initiatives such as STARD should be endorsed to improve the awareness of design features, the quality of reporting and, ultimately, the quality of study designs before starting a PCT study.

## CONCLUSIONS

With respect to the clinical usefulness of PCT in diagnosing or ruling out EONS, it is time to debate the methods used to measure PCT performance, rather than just how a given test performs. To interpret correctly the results of PCT accuracy



studies for EONS, readers must understand the design, conduct, and data analysis and must be able to judge the internal validity and generalizability of the results. This goal can only be achieved through complete transparency of the reporting of the articles. The results of our analysis involving PCT accuracy in diagnosing EONS show that the quality of reporting for many of the STARD items, that have been shown to have a potentially biasing effect on the results of diagnostic accuracy studies or appear to account for variation between studies, is substandard. For this reason, this article does not include any summary results or meta-analysis. Authors and peer reviewers are encouraged to adhere to and enforce the STARD guidelines, because there is clearly room for improvement in the reporting of PCT diagnostic accuracy studies for EONS.

## REFERENCES

- Weston EJ, Pondo T, Lewis MM, et al. The burden of invasive early-onset neonatal sepsis in the United States, 2005–2008. *Pediatr Infect Dis J*. 2011;30:937–941.
- Stoll BJ, Hansen NI, Sánchez PJ, et al. Early onset neonatal sepsis: the burden of group B Streptococcal and *E. coli* disease continues. *Pediatrics*. 2011;127:817–826.
- Chiesa C, Panero A, Osborn JF, et al. Diagnosis of neonatal sepsis: a clinical and laboratory challenge. *Clin Chem*. 2004;50:279–287.
- Squire E, Favara B, Todd J. Diagnosis of neonatal bacterial infection: hematologic and pathologic findings in fatal and nonfatal cases. *Pediatrics*. 1979;64:60–64.
- Connell TG, Rele M, Cowley D, et al. How reliable is a negative blood culture result? Volume of blood submitted for culture in routine practice in a children's hospital. *Pediatrics*. 2007;119:891–896.
- Pacifico L, Osborn JF, Natale F, et al. Procalcitonin in Pediatrics. *Adv Clin Chem*. 2013;59:203–263.
- Dandona P, Nix D, Wilson MF, et al. Procalcitonin increase after endotoxin injection in normal subjects. *J Clin Endocrinol Metab*. 1994;79:1605–1608.
- Chiesa C, Pacifico L, Mancuso G, et al. Procalcitonin in pediatrics: overview and challenge. *Infection*. 1998;26:236–241.
- Reinhart K, Meisner M. Biomarkers in the critically ill patient: procalcitonin. *Crit Care Clin*. 2011;27:253–263.
- van Rossum AMC, Wulkan RW, Oudesluys-Murphy AM. Procalcitonin as an early marker of infection in neonates and children. *Lancet Infect Dis*. 2004;4:620–630.
- Monneret G, Labaune JM, Isaac C, et al. Procalcitonin and C-reactive protein levels in neonatal infections. *Acta Paediatr*. 1997;86:209–212.
- Blommendahl J, Janas M, Laine S, et al. Comparison of procalcitonin with CRP and differential white blood cell count for diagnosis of culture-proven neonatal sepsis. *Scand J Infect Dis*. 2002;34:620–622.
- Janota J, Stránák Z, Bělohávková S, et al. Postnatal increase of procalcitonin in premature newborns is enhanced by chorioamnionitis and neonatal sepsis. *Eur J Clin Invest*. 2001;31:978–983.
- Bonac B, Derganc M, Wraber B, et al. Interleukin-8 and procalcitonin in early diagnosis of early severe bacterial infection in critically ill neonates. *Pflugers Arch*. 2000;440(Suppl 5):R72–R74.
- Franz AR, Kron M, Pohlandt F, et al. Comparison of procalcitonin with interleukin 8, C-reactive protein and differential white blood cell count for the early diagnosis of bacterial infections in newborn infants. *Pediatr Infect Dis J*. 1999;18:666–671.
- Lapillonne A, Basson E, Monneret G, et al. Lack of specificity of procalcitonin for sepsis diagnosis in premature infants. *Lancet*. 1998;351:1211–1212.
- Van Lente F, Pippenger CE. The pediatric acute care laboratory. *Pediatr Clin North Am*. 1987;34:231–246.
- Chiesa C, Panero A, Rossi N, et al. Reliability of procalcitonin concentrations for the diagnosis of sepsis in critically ill neonates. *Clin Infect Dis*. 1998;26:664–672.
- Coffin CM, Hamilton MS, Pysker TJ, et al. Pediatric laboratory medicine: current challenges and future opportunities. *Am J Clin Pathol*. 2002;117:683–690.
- Chiesa C, Osborn JF, Pacifico L, et al. Gestational-and age-specific CRP reference intervals in the newborn. *Clin Chim Acta*. 2011;412:1889–1890.
- Chiesa C, Natale F, Pascone R, et al. C reactive protein and procalcitonin: reference intervals for preterm and term newborns during the early neonatal period. *Clin Chim Acta*. 2011;412:1053–1059.
- Chiesa C, Pellegrini G, Panero A, et al. C-reactive protein, interleukin-6, and procalcitonin in the immediate postnatal period: influence of illness severity, risk status, antenatal and perinatal complications, and infection. *Clin Chem*. 2003;49:60–68.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Reporting of Diagnostic Accuracy. *Clin Chem*. 2003;49:1–6.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem*. 2003;49:7–18.
- Polin RA. Committee on Fetus and Newborn. Management of neonates with suspected or proven early-onset bacterial sepsis. *Pediatrics*. 2012;129:1006–1015.
- Guibourdenche J, Bedu A, Petzold L, et al. Biochemical markers of neonatal sepsis: value of procalcitonin in the emergency setting. *Ann Clin Biochem*. 2002;39:130–135.
- Resch B, Gusenleitner W, Müller WD. Procalcitonin and interleukin-6 in the diagnosis of early-onset sepsis of the neonate. *Acta Paediatr*. 2003;92:243–245.
- Distefano G, Curreri R, Betta P, et al. Procalcitonin serum levels in perinatal bacterial and fungal infection of preterm infants. *Acta Paediatr*. 2004;93:216–219.
- Pavcnik-Arnol M, Hojker S, Derganc M. Lipopolysaccharide-binding protein, lipopolysaccharide, and soluble CD14 in sepsis of critically ill neonates and children. *Intensive Care Med*. 2007;33:1025–1032.
- López Sastre JB, Solís DP, Serradilla VR, et al., Grupo de Hospitales Castrillo. Evaluation of procalcitonin for diagnosis of neonatal sepsis of vertical transmission. *BMC Pediatr*. 2007;7:1–9.
- Bender L, Thaarup J, Varming K, et al. Early and late markers for the detection of early-onset neonatal sepsis. *Dan Med Bull*. 2008;55:219–223.
- Santuz P, Soffiati M, Dorizzi RM, et al. Procalcitonin for the diagnosis of early-onset neonatal sepsis: a multilevel probabilistic approach. *Clin Biochem*. 2008;41:1150–1155.
- Kordek A, Hałasa M, Podraza W. Early detection of an early onset infection in the neonate based on measurements of procalcitonin and C-reactive protein concentrations in cord blood. *Clin Chem Lab Med*. 2008;46:1143–1148.
- Spada S, Cuccu A, Mussap M, et al. Reliability of procalcitonin in neonatology. Experience in 59 preterm newborns. *J Matern Fetal Neonatal Med*. 2009;22(Suppl 3):96–101.
- Joram N, Muller JB, Denizot S, et al. Umbilical cord blood procalcitonin level in early neonatal infections: a 4-year university hospital cohort study. *Eur J Clin Microbiol Infect Dis*. 2011;30:1005–1013.

36. Canpolat FE, Yiğit S, Korkmaz A, et al. Procalcitonin versus CRP as an early indicator of fetal infection in preterm premature rupture of membranes. *Turk J Pediatr*. 2011;53:180–186.
37. Altunhan H, Annagür A, Örs R, et al. Procalcitonin measurement at 24 hours of age may be helpful in the prompt diagnosis of early-onset neonatal sepsis. *Int J Infect Dis*. 2011;15:e854–e858.
38. Abdollahi A, Shoar S, Nayyeri F, et al. Diagnostic value of simultaneous measurement of procalcitonin, interleukin-6 and hs-CRP in prediction of early-onset neonatal sepsis. *Mediterr J Hematol Infect Dis*. 2012;4:e2012028.
39. Schlapbach LJ, Graf R, Woerner A, et al. Pancreatic stone protein as a novel marker for neonatal sepsis. *Intensive Care Med*. 2013;39:754–763.
40. Lencot S, Cabaret B, Sauvage G, et al. A new procalcitonin cord-based algorithm in early-onset neonatal infection: for a change of paradigm. *Eur J Clin Microbiol Infect Dis*. 2014;33:1229–1238.
41. Steinberger E, Hofer N, Resch B. Cord blood procalcitonin and interleukin-6 are highly sensitive and specific in the prediction of early-onset sepsis in preterm infants. *Scand J Clin Lab Invest*. 2014;74:432–436.
42. Newcombe RG, Altman DG. Proportions and their differences. In: *Statistics with Confidence*, Altman DG, Machin D, Bryant TN, Gardner MJ, editors. 2nd ed. BMJ Books, Bristol, England; 2000: Chapter 6, 45–56.
43. Korevaar DA, van Enst WA, Spijker R, et al. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med*. 2014;19:47–54.
44. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061–1066.
45. Whiting PA, Rutjes W, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004;140:189–202.
46. Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174:469–476.
47. Wilczynski NL. Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication-before and after study. *Radiology*. 2008;248:817–823.
48. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926–930.
49. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Assessing methodological quality. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0.*, Deeks JJ, Bossuyt PM, Gatsonis C, editors. The Cochrane Collaboration; 2009: Chapter 9, 1–27 [available online at: <http://srdta.cochrane.org/>].
50. Whiting PF, Weswood ME, Rutjes AW, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:1–13.
51. Rutjes AW, Reitsma JB, Vandenbroucke JP, et al. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem*. 2005;51:1335–1341.
52. Philbrick JT, Horwitz RI, Feinstein AR, et al. The limited spectrum of patients studied in exercise test research. Analyzing the tip of the iceberg. *JAMA*. 1982;248:2467–2470.
53. Knottnerus JA, Knipschild PG, Sturmans F. Symptoms and selection bias: the influence of selection towards specialist care on the relationship between symptoms and diagnoses. *Theor Med*. 1989;10:67–81.
54. Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol*. 1992;45:1143–1154.
55. Melbye H, Straume B. The spectrum of patients strongly influences the usefulness of diagnostic tests for pneumonia. *Scand J Prim Health Care*. 1993;11:241–246.
56. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. *CJEM*. 2008;10:174–175.
57. Higgins JPT, Green S. Guide to the contents of a Cochrane protocol and review. In: *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0.*, Higgins JPT, Green S, editors. The Cochrane Collaboration (Cochrane Book Series): John Wiley & Sons Ltd, Chichester, West Sussex, England; 2008: Chapter 4, 51–80.
58. Choi BC. Sensitivity and specificity of a single diagnostic test in the presence of work-up bias. *J Clin Epidemiol*. 1992;45:581–586.
59. Zhou XH. Effect of verification bias on positive and negative predictive values. *Stat Med*. 1994;13:1737–1745.
60. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*. 1983;39:207–215.
61. Greenes RA, Begg CB. Assessment of diagnostic technologies. Methodology for unbiased estimation from samples of selectively verified patients. *Invest Radiol*. 1985;20:751–756.
62. Cecil P, Kosinski AS, Jones MT, et al. The importance of work-up (verification) bias correction in assessing the accuracy of SPECT thallium-201 testing for the diagnosis of coronary artery disease. *J Clin Epidemiol*. 1996;49:735–742.
63. Diamond GA, Rozanski A, Forrester JS, et al. A model for assessing the sensitivity and specificity of tests subject to selection bias. Application to exercise radionuclide ventriculography for diagnosis of coronary artery disease. *J Chronic Dis*. 1986;39:343–355.
64. Diamond GA. Off Bayes: effect of verification bias on posterior probabilities calculated using Bayes' theorem. *Med Decis Making*. 1992;12:22–31.
65. Harper R, Reeves B. Reporting of precision of estimates for diagnostic accuracy: a review. *BMJ*. 1999;318:1322–1323.
66. Simel DL, Feussner JR, DeLong ER, et al. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med Decis Making*. 1987;7:107–114.
67. Mower WR. Evaluating bias and variability in diagnostic test reports. *Ann Emerg Med*. 1999;33:85–91.
68. Poynard T, Chaput JC, Etienne JP. Relations between effectiveness of a diagnostic test, prevalence of the disease, and percentages of uninterpretable results. An example in the diagnosis of jaundice. *Med Decis Making*. 1982;2:285–297.