In: Advances in Medicine and Biology ISBN: 978-1-53614-421-5
Editor: Leon V. Berhardt

Chapter 1

PROCALCITONIN AND DIAGNOSIS OF NEONATAL SEPSIS: MOVING ON

Claudio Chiesa[1,*], MD, Lucia Pacifico[2], MD, John F. Osborn[3], PhD, Michela Lavorato[2], MD and Bernhard Resch[4], MD

[1]Institute of Translational Pharmacology,
National Research Council, Rome, Italy
[2]Department of Pediatrics, Sapienza University of Rome, Rome, Italy
[3]Department of Public Health and Infectious Diseases,
Sapienza University of Rome, Rome, Italy
[4]Division of Neonatology, Department of Pediatrics and Adolescent Medicine, Medical University of Graz, Austria

ABSTRACT

To interpret correctly the results of procalcitonin (PCT) accuracy studies for the diagnosis of neonatal sepsis, it is time to begin to debate the methods we use to measure PCT performance, rather than just how PCT performs. Therefore we should move toward the better use of the information available including criteria to: (1) substantiate (or rule-out)

* Corresponding Author Email: claudio.chiesa@ift.cnr.it.

the diagnosis of neonatal sepsis; (2) define the normal(and abnormal) upper limits of PCT; (3) identify which prenatal, perinatal and neonatal variables are in play in modulating host characteristics and PCT response in the healthy, uninfected and infected neonate, respectively; and (4) evaluate the quality of reporting of PCT accuracy studies for diagnosing neonatal sepsis in terms of internal (risk of bias) and external (generalizability) validity of the results.

1. INTRODUCTION

1.1. Neonatal Sepsis and Challenges in Diagnosis

The diagnosis of sepsis is a problem in medicine in general, and in neonatology in particular. Problems in the diagnosis of neonatal sepsis are present at multiple levels: definition of the problem (just what is sepsis), definition of what is not the problem (what constitutes a normal result for any given test), nomenclature (one person's "early onset" sepsis could be well another's "late onset"), and, last but not least, problems with test results and test accuracy.

2. DEFINITION OF NEONATAL SEPSIS

2.1. Heterogeneity Among Studies

There is a considerable heterogeneity among studies regarding the case definition of neonatal sepsis. The gold standard for establishing a diagnosis of neonatal sepsis is through culture. However, culture can be falsely sterile, as suggested by postmortem cultures [1], or because of the low yield caused by insufficient sample volume, intermittent or low-density bacteremia, or suppression of bacterial growth by earlier (i.e., intrapartum) antibiotic administration [2]. Faced with these problems, part of the neonatology literature has abandoned positive cultures as a critical point of the "gold standard". If culture results are negative, but the infants have signs of infection, they are considered to have clinical sepsis. In neonates,

the scenario of a child with a culture-negative clinical sepsis is far more common than blood culture-positive sepsis [3]. Identification of neonatal clinical sepsis is challenging as the clinical signs of neonatal septicemia are not specific and can be very similar to those of other life-threatening diseases such as necrotising enterocolitis, hyaline membrane disease, and perinatal asphyxia [4]. Yet, many early signs of infection in neonates may be simply associated with prematurity or the transition to extrauterine life. On the other hand, asymptomatic status cannot rule out infection [5].

2.1.1. Pediatric Consensus Definition for Neonatal Sepsis

It is striking to see how the neonatal and adult patients have diverged on the issue of sepsis. As emphasized by Wynn J.L. [6], while present consensus definition for sepsis in children and adults are threshold-based and therefore static, a static definition of sepsis in the newborn transitioning to extrauterine life will be associated with continued limitations in diagnostic accuracy because neonatal sepsis is a dynamic and complex condition including a rapid and continuous change in clinical status over time. Thus time of clinical presentation/biomarker measurement, and duration of clinical and laboratory signs can be highly informative in diagnostic studies. As such, there is no universal agreement on the definition of neonatal clinical septicemia in practice and in research [2, 6]. In recent years, some authors have suggested that the presence of just one clinical sign (compatible with infection) along with a C-reactive protein (CRP) value >10 mg/L, is sufficient to make the diagnosis of early- and late-onset neonatal clinical septicemia [7, 8]. Others have assigned infants with only maternal risk factor(s) for infection to the sepsis group [9]. In other studies, the presence of two or three categories (by organ system) of clinical signs of infection in the infant has been taken into consideration to strongly support a diagnosis of septicemia [10, 11]. Absent from neonatology publications, however, are data on just how common a given clinical sign is in all infants ever evaluated (rather than in infants with positive cultures). Thus, with the exception of the clinical scenario of a newborn with clear-cut signs of infection, the possibility that

infants with only clinical evidence of infection may have been assigned an incorrect diagnosis is intrinsic to all studies of this nature [2, 12].

Moving forward, there is increasing focus on the potential of more accurate diagnostic tests to aid in identifying or excluding neonatal sepsis. Many laboratory tests including various leukocyte indices and acute-phase proteins have been recommended for the evaluation of suspected infection in the neonate [2]. The most extensively studied acute-phase reactants have been CRP and procalcitonin (PCT).

3. EXPLORING WHETHER TEST RESULTS ARE "WITHIN NORMAL LIMITS" (OR OUTSIDE)

3.1. Gestational- and Age-Specific Reference Intervals in the Newborn

The newly born baby must undergo extreme physiologic changes to survive the transition from the protective intrauterine existence to the environment of the outside world. It is therefore not surprising that the reference intervals for many laboratory markers (including hormones, proteins, enzymes, immunologic markers, and cytokines) change considerably with time from birth [13]. However, for neonates, assessment of laboratory tests also occurs within a context of prenatal growth and neonatal development [14]. Clearly, determining age-specific reference intervals is crucial for neonatal diagnosis and monitoring, yet organ growth, development, and physiologic function may not be synchronous with age, as in premature infants [14]. Thus analysis of the reliability of laboratory tests for the diagnosis of neonatal sepsis must be done in light of the knowledge that differences in postnatal as well as gestational ages may significantly affect the interpretation of what constitutes the "normal" (and the "abnormal") value of a laboratory test over the neonatal period [15, 16]. In recent years, attempts to develop neonatal reference intervals for various elements of the complete blood count (CBC) as well as CRP and PCT have supported this premise.

3.1.1. Complete Blood Count and Neonatal Reference Intervals

Published reference intervals vary widely for CBC [17-20]. Total and differential white blood cell counts are affected by many factors besides infection, including gestational age [20], infant age in hours [17-20], the method of blood sampling [21, 22], the method of delivery [23, 24], maternal hypertension [18], and the infant's gender [23]. As multiple factors can alter the CBC dynamics, it is not surprising that many different values for the sensitivity and specificity of different components of the CBC as predictors of infection have been published, depending on what values of these tests have been considered "abnormal". For these reasons, this limits the ability to make a definite diagnosis based on these markers alone.

3.1.2. C Reactive Protein and Neonatal Reference Intervals

Data pertaining to CRP reference intervals are limited. In the majority of published reports, CRP upper limits have been obtained from symptomatic uninfected patients [2]. In this respect, the literature offered conflicting cutoff points (1.5–20 mg/L) for the upper limits of CRP. Possible sources of heterogeneity included wide-ranging differences in postnatal age or inaccuracies in reporting it, single vs serial determinations, different sample sizes, and different measurement methods [25]. There have been fewer cross-sectional studies of upper limits for CRP in healthy newborns [26-28]. Again, even among these studies, there was some degree of heterogeneity based on different measurement methods and different sampling times. Thus, in 2001 we first performed a longitudinal study of CRP dynamics during the immediate postnatal period in a large sample of healthy newborns, whose postnatal clinical course from birth to the 4-week follow-up visit was unremarkable [29], implying therefore no need of management (including antimicrobial treatment) throughout this study period. We established the "normal" upper limits for CRP by serial determinations at three fixed neonatal ages: 0, 24, and 48 hours (h) after birth, and found that at birth CRP was significantly lower than at 24 and 48 h of life [29]. Ten years later, we performed a further longitudinal study in which we assessed the influence of both gestational and postnatal ages on

the development of CRP reference intervals during the first 4-5 days of life in the truly healthy preterm as well as term neonate [30]. We constructed the 95% age (hour)-specific reference ranges for CRP in the healthy preterm and term neonate, and found that healthy preterm babies had a lower and shorter CRP response compared with that in healthy term babies, demonstrating the effects of development *per se* on CRP dynamics. Thus maturational changes are the most likely explanation for the increasing likelihood of a pronounced CRP response with increasing gestational age [16]. Variables that induced a significant deviation from these ranges included a number of proinflammatory risk factors such as prolonged rupture of membranes, preterm labor and Group B Streptococcus (GBS) colonization, method of delivery, and intrapartum fetal distress [30]. Hence, the above CRP responses and time-dependent variables argue against the arbitrary selection of a particular CRP as being "normal" (or "abnormal") for ruling out (or diagnosing) sepsis over the entire neonatal period, regardless of many potential prenatal, perinatal and neonatal confounders like gestational age or time point of evaluation.

3.1.3. Procalcitonin (PCT) and Neonatal Reference Intervals

Data pertaining to reference intervals for PCT are also limited. In 1997, Monneret et al. first recognized daily variations of PCT concentrations during the first days of life in uninfected ill newborns [31]. One year later, in a cross-sectional study involving healthy newborns who were born after uncomplicated pregnancy and labor and who had normal postnatal courses until day 3 of life (when they were discharged home), we determined the reference intervals for PCT across the range of postnatal hours from birth to 48 h- a time span where the majority of diagnostic problems are encountered [32]. To this end, we constructed 95% age (hour)-specific reference ranges for PCT in the healthy term neonates by following the sequence of six steps described by Royston [13], and found that in these neonates PCT increases gradually from birth to reach peak values at about 24 h of age and then decreases gradually to 48 h of life [32]. In 2011, we first reported a longitudinal study in which we assessed the influence of both gestational and postnatal ages on the development of

PCT reference intervals during the first 4-5 days of life in the truly healthy preterm as well as term neonate [30]. We constructed the 95% age (hour)-specific reference ranges for PCT in the healthy preterm and term neonate, and found that healthy preterm babies had an earlier, higher, and longer PCT response compared with that found in the healthy term baby [30], demonstrating an inverse relationship between stage of development and magnitude of neonatal PCT response. Our data also indicated that PCT reference intervals may be possibly hampered by specific confounders, such as prolonged rupture of membranes and gestational diabetes [30].

In summary, the findings here reported for both PCT and CRP emphasize the need to introduce separate PCT as well as CRP age-specific reference intervals for preterm and term newborns in order to optimize the accuracy of these inflammatory markers for the diagnosis of early-onset neonatal sepsis (EONS). Nonetheless, further prospective studies are clearly needed to refine our gestational- and age-specific PCT as well as CRP reference intervals [30], though they were assessed in 200 "truly" healthy preterm infants (range 30.0–36.0 weeks; 13.5% below 33 weeks of gestational age) and 221 healthy full-term neonates (one observation for each healthy neonate). Prospective studies are also needed to determine the predictability and usefulness of these PCT and CRP reference intervals in the diagnosis of EONS in both preterm and term neonates. Finally, efforts should be made to establish PCT as well as CRP reference intervals in the full-term as well as in the preterm healthy neonate beyond the early neonatal period.

3.1.4. Studies on PCT Reference Intervals in the NICU Setting

Notably, the above description of the physiologic PCT response in the healthy term and preterm neonate during the immediate postnatal period is in contradiction with the report by Turner et al. [33] who found that PCT concentrations, in particular peak values, obtained during the first 4 days of life in a cohort of symptomatic uninfected preterm (range *24-36 weeks* gestation) babies admitted to the neonatal intensive care unit (NICU) were lower than those previously reported in healthy term infants [32]. These authors concluded that PCT concentrations decrease in preterm versus term

babies, but no control group of symptomatic uninfected term babies was provided in their study [33]. Furthermore, PCT reference intervals were calculated in their cohort by repeated measurements on the same infant without checking the magnitude of the bias that this might introduce. Using all the information should decrease sampling error and thus improve precision. However, repeated measurements on the same individual are not equivalent to additional independent observations. If for example, the patients- for whom repeated measurements were made - had particular characteristics (they were more severely ill, they were smaller) then to include the repeated measurements would have introduced a bias into estimates of the reference range including underestimation of the standard deviation. It would have been interesting if Turner et al. [33] had checked how big the bias might have been by repeating the calculation of the reference ranges including only one randomly selected observation from each baby.

Further attempts to assess PCT reference intervals in the preterm neonate included other studies collecting data from NICU settings. Hahn et al. [34] reported PCT reference intervals during 7-60 postnatal days in "non-septic" very low birth weight infants populating the NICU for a variety of clinical conditions (including respiratory distress syndrome, pneumothorax, bronchopulmonary dysplasia, intraventricular hemorrhage) as well as of pulmonary or hemodynamic events, with different degrees of illness severity. Furthermore, in a study involving NICU symptomatic "non-septic" extremely preterm, preterm, and term infants, Fukuzumi et al. [35] reported age-specific percentile-based reference curves of serum PCT concentrations from birth to 5 days of age. This study was based on retrospective data collection (rather than on a prospective sampling method permitting therefore immediate error checking, better data control, additional data for data integrity and consistency, a level of clinical detail appropriate to the problem, and most important, minimizing the problems of missing records). Like Turner et al. [33] and Hahn et al. [34], Fukuzumi et al. included all the repeated measurements to calculate their reference intervals.

Taken together, the results of these studies attempting to establish reference intervals in the NICU setting have to be interpreted with caution. In the NICU population, there are many postnatal- and gestational-age dependent variables and potential scenarios that may significantly affect the nature and magnitude of the host response, leading to misinterpretation of "normal" (and abnormal) PCT values. Thus it is worrying that, absent from the studies by Turner et al.[33], Hahn et al. [34] and Fukuzumi et al. [35], are data on how interpretation of the so called "physiologic" PCT response obtained from their NICU babies, especially within the narrow group of "very" or "extremely" low birth weight infants, might have been hampered by the severity of underlying illnesses (and their extent of inflammatory reaction) and the intensity of therapeutic interventions and supportive therapies [36]. In view of the wide variation in clinical severity and therapy-based illness severity among NICUs, can these Authors confidently generalize and transfer their findings with equivalent results to other NICUs? Not having developed accurate and reproducible reference intervals implicates the potential for inaccurate diagnosis or patient misclassification, inappropriate treatment, or subsequent unnecessary diagnostic procedures [14].

4. PCT Response to Early-Onset Neonatal Sepsis (EONS): Influence of Illness Severity, Risk Status, Antenatal and Perinatal Complications

4.1. Influence of Illness Severity and Risk Status

Host characteristics and response during the immediate postnatal course are related to illness severity and risk status. As such, values of the laboratory markers measured over the first days of life may be altered by physiologic severity and risk indexes. Individual clinical and birth characteristics, such as gestational age, birth weight and Apgar scores, have been reported to act as proxy measures of morbidity in most PCT diagnostic studies [37]. However, they do not capture the overall morbidity

status. The Score for Neonatal Acute Physiology (SNAP) was designed to measure the physiologic instability of the newborn in the first 24 h of life [38]. As these measurements change over time, the SNAP instrument was also designed to allow for sequential measurement of the continuum of illness severity [38]. It is a validated, objective physiology-based score consisting of 34 items capturing the fundamental construct of newborn admission illness severity [39]. From this, SNAP-perinatal extension (PE) combines SNAP with three additional perinatal factors, providing an overall risk of mortality [40]. Notably, although studies of the diagnostic accuracy of PCT for EONS have yielded variable results [37, 41, 42], absent from most NICU publications are data on how the interpretation of the host response during the immediate postnatal period in the uninfected as well as in the infected neonates might have been influenced by the severity of the underlying illnesses (and their extent of inflammatory reaction).

4.1.1. PCT Reliability for Diagnosis of EONS

PCT has been suggested to be a possible mediator of the systemic inflammatory response syndrome (SIRS) [43]. This prompted us to investigate in a longitudinal study [44] whether some of this variation among published reports might be attributable to differences in population baseline severity and risk status, as well as to specific ante- and perinatal variables, independent of the presence of neonatal infection.

To this end, we used the SNAP and SNAP-PE instruments [39, 40]. A total of 134 ill newborns (19 with EONS and 115 without infection) was available for simultaneous analysis of the association of SNAP, SNAP-PE, and maternal and perinatal variables with PCT concentrations at birth and at 24 and 48 h of life [44]. We found that EONS was associated with significant increases in PCT concentrations at all three time points, independent of severity scores.

4.1.2. PCT Response to EONS and Influence of Maternal and Perinatal Variables

We also found that certain maternal or perinatal variables altered PCT values in the infected as well as in the uninfected neonates [44]. In fact, PCT response was independently associated at birth with clinical chorioamnionitis, at 24 h with birth asphyxia (defined as the need for delivery room intubation), and at both 24 and 48 h of age with preeclampsia. In a previous study [45], we demonstrated that maternal GBS colonization and rupture of membranes ≥18 h may significantly affect the patterns of PCT response in the healthy neonate. Thus, it is not surprising that in the presence of clinical chorioamnionitis, PCT concentrations at birth were found to be significantly increased independently of the presence of infection. As suggested by Gendrel and Bohuon [46], hypoxemia may also be responsible for increased PCT values in neonates. Similarly, we have found that birth asphyxia may affect PCT concentrations. Nonetheless, none of the above variables appeared to interfere with the use of PCT for the detection of EONS. In fact, although certain ante-natal and perinatal variables altered PCT values in the infected as well as in the uninfected neonates, when different cutoff points were used at the three common neonatal ages (0, 24, 48 h), the increases of PCT caused by these variables were relatively small compared with the magnitude of PCT response to infection at each of the three time points.

4.1.3. EONS and PCT Specificity

Against the above background, some studies have noted the lack of specificity of PCT for the diagnosis of sepsis in neonates [7, 31, 47-50]. However, their conclusions should be interpreted with caution. First, increased PCT values have been found in critically ill newborns clinically presenting with "apparently" noninfectious conditions including respiratory distress syndrome (RDS), hemodynamic instability, asphyxia, intracranial hemorrhage, and pneumothorax [7, 31, 47-50]. Absent from most neonatal studies, however, is the role of severity of the underlying illnesses (and their extent of inflammatory reaction) in modulating PCT response in the uninfected versus infected neonates [41, 44]. Second,

arbitrary cutoffs have been used to differentiate infectious and noninfectious clinical conditions. For instance, in the report by Lapillonne et al., 16 uninfected neonates (mean postnatal age, 2.3 days) were deemed to have high serum PCT concentrations on the basis of a surrogate cutoff level originally established in children admitted to a "pediatric" intensive care unit [50]. Third, interpretation of PCT has been complicated by diverse study populations [51]. Heterogeneity not only within the study group, but also within categories defined as ''sepsis'', ''distress'', ''infected'', ''respiratory distress'', or even ''hemodynamic failure'' has been remarkable [51]. Fourth, the PCT response has been assessed in neonates with wide-ranging differences in postnatal age (hours to weeks) [51]. However, the use of PCT for EONS and neonatal nosocomial sepsis should be evaluated separately [52]. Fifth, PCT levels obtained from uninfected patients have not been 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. Finally, it is uncertain how the infectious state in the ''uninfected'' neonates has been ruled out. In the report by Monneret et al. [31], for example, high PCT concentrations were reported during the first 4 days of life in apparently ''uninfected'' newborns who presented with RDS of various etiologies. However, it would be of interest to know how many patients with severe RDS demonstrated clinical evidence of being ''uninfected.'' It is clear from the above that the evaluation of the clinical usefulness of PCT in ruling out (as well as in ruling in) neonatal sepsis is dependent on study consistency [52].

5. NOSOCOMIAL NEONATAL SEPSIS, PCT AND SEVERITY OF ILLNESS

5.1. Nosocomial Neonatal Sepsis as Public Health Problem

Recently, nosocomial (or late-onset) neonatal sepsis has become an increasing and persistent public health problem in parallel with the

significant increase in survival of premature infants, associated at the same time with prolonged hospitalization, mechanical ventilation, use of invasive procedures and devices (i.e., intravascular catheters and endotracheal tubes), which are all predisposing factors to nosocomial sepsis [53]. A timely and accurate diagnosis of nosocomial neonatal sepsis is of utmost importance, given the associated mortality rate and long-term adverse outcomes.

5.1.1. Heterogeneity among Studies Regarding PCT Reliability

There is remarkable heterogeneity among studies regarding PCT reliability for diagnosis of late-onset neonatal sepsis [54-63]. Though illness severity is something that changes over time across the entire NICU length of stay [38], no studies have yet investigated whether conflicting results might be attributable to differences in illness severity as well as in intensity of therapeutic intervention (as an indicator of neonatal illness severity and resource utilization [64]) that might modulate the host response. Therefore, a critical step to ensure the correct interpretation of PCT accuracy for the detection of late-onset neonatal infection, would be to measure in newborns with and without nosocomial sepsis sequentially and at later time points, PCT concentrations as well as illness severity (operationalized by the SNAP instrument [39]) and therapy-based illness severity scores (as measured by Neonatal Therapeutic Intervention scoring across system categories [64, 65]), along with assessment of postnatal and gestational ages [66].

6. PCT ACCURACY STUDIES FOR DIAGNOSIS OF NEONATAL SEPSIS

6.1. Moving Beyond the Question "How PCT Performs"

The reliability of PCT for diagnosis of neonatal sepsis in the NICU has yielded markedly heterogeneous results originating from differences in the design and conduct of the studies [37, 41]. Thus to interpret correctly the

results of PCT accuracy studies for the diagnosis of neonatal sepsis, it is time to move away from post hoc analyses debating which arbitrary (single) cutoffs are optimal for diagnosing this disease, and disputing how PCT performs in such clinical settings on the basis of dichotomous results [42, 67-72]. Artificially low or high sensitivity and specificity reported by studies with less than optimal scientific rigor may bias the clinician's findings, potentially resulting in an inaccurate diagnosis of neonatal sepsis [73]. Most worrisome is that there is an abundant, ever-growing body of literature in this field that is still a matter of debate. It is time to begin to debate the methods we use to measure PCT performance, rather than just how PCT performs [74]. Therefore, we should move toward the better use of the information available "before" the performance of a test such as PCT [74], and interpret the results of clinical studies using PCT for neonatal sepsis correctly, by judging internal (risk of bias) and external (generalizability) validity of the results [75, 76]. If the results are biased and are synthesized just by graphing results without any consideration of quality (in terms of potential for bias, lack of applicability, and the quality of reporting), then it is clear that the results and conclusions of the review(s) will also be biased [77]. Though these concerns have been apparently ignored by other systematic reviews, meta-analyses, and reappraisals [42, 67-69], we have followed these precepts in a recent systematic review [37].

6.1.1. EONS and Quality of Reporting of PCT Accuracy Studies

In that systematic review [37], studies were eligible for inclusion if they provided measures of PCT accuracy for diagnosing EONS (defined as sepsis with onset at ≤ 3 days of age). Data extraction included the reference standard employed, type of the study design, the number and specific characteristics of the patients in the septic and non-septic groups, and items related to the quality of the methods and reporting according to the Standards for Reporting of Diagnostic Accuracy (STARD) criteria which were adapted for neonates with EONS [78]. We decided not to use Quality assessment of Studies of Diagnostic Accuracy (QUADAS) because one problem by using the QUADAS is the distinction between the reporting of

a study and its quality [77]. While QUADAS was designed to be used in systematic reviews to evaluate the quality of primary diagnostic accuracy studies [77, 79], STARD was designed to improve the quality of reporting of diagnostic accuracy studies (defined as the number of reported STARD items) over several domains [78], as discussed below.

6.1.2. Participant Recruitment

In evaluating the quality of reporting of PCT accuracy studies for diagnosing EONS, an important objective is to describe how eligible subjects were identified. Selection of a study population is relevant to all studies of diagnostic accuracy. Description of the populations from which patients and patient controls originated enables the assessment of the ‘‘spectrum of disease’’, which is likely to influence the diagnostic performance of the test [80]. Reported estimates of diagnostic accuracy may have limited clinical applicability (generalizability) when the spectrum of tested patients is not similar to the patients for whom the test will be used in practice [75, 77]. Notably, our analysis showed that some studies of PCT accuracy enrolled only neonates who were suspected of having the disease because of presenting symptoms, while other studies recruited neonates who were asymptomatic and only at risk of developing the disease because of antenatal and perinatal history. Other studies included patients already known to have the disease. Finally, the methods used to sample patients for the study were different and included: the “diagnostic case-control” or “two-gate design” [81, 82] often referred to as study in which patients with obvious disease are selected as cases and patients without obvious disease are selected as controls (a sampling method whose estimates are generally not representative of a test’s accuracy in clinical practice, though they provide an indication of the maximum accuracy of a test and are therefore valuable in the technical validation of a test [81, 83]); non-consecutive sampling of patients, with the exclusion of more difficult cases and hence increased estimates of specificity and sensitivity (a sampling method leading to the “limited challenge bias” [84]); retrospective data collection (a sampling method making it difficult to use unambiguous inclusion criteria and to identify

patients who received the index test but whose test results were not subsequently verified [85]); and identification of patients by searching hospital records. Concerns regarding the applicability may also arise when patient selection includes a combination of different populations [86]. In particular, in a different population, such as neonates with early-onset sepsis versus those with late-onset sepsis, the sensitivity and specificity are likely to change. In these circumstances, different postnatal (and gestational) ages and related risk factors may affect the interpretation of the reliability of any laboratory marker including PCT. These alternative study designs may therefore have limited generalizability. Spectrum bias may also result from differences in severity of disease/symptoms between and within neonatal populations. The sensitivity of a test will often vary according to the severity of disease. It is therefore of concern that in our systematic review the reporting of the distribution of scores of the major measures of illness severity was poor in neonates with and without sepsis.

6.1.3. Methods Pertaining the Reference Standard and the Index Test; Statistical Methods

Studies of diagnostic accuracy have a common basic structure. One or more tests are evaluated, with the purpose of detecting (or excluding) a target condition. However, the way the index test, i.e., PCT, works depends on which variables [the site, type and extent of the infectious (and noninfectious) insult; the nature and magnitude of the host response including severity of illness and risk status; onset or progression of the host inflammatory response; the degree of concomitant organ dysfunction] are in play. Thus in evaluating the quality of reporting of PCT accuracy studies for diagnosing EONS, the choice of an optimal reference standard to determine the disease status is crucial. Test sensitivity may vary with the stage of the target condition.

Consequently, another important objective is to describe how eligible subjects were identified. In this respect, differential verification is also a key issue in any diagnostic accuracy study [85]. Differential verification bias occurs when the reference standard is not the same for all patients [77]. Some of the index test results are verified by one type of reference

standard and other results by a different standard [75]. Differential verification is of great concern when reference standards differ in accuracy. They may not have the same degree of error and may not identify the same segment of the disease spectrum [86]. Differential verification may also pose a problem when the choice of reference standard relates to the results of the index test [75]. This usually occurs when patients testing positive on the index test receive a more accurate reference standard than those testing negative. Thus, it is of concern that in most of PCT diagnostic accuracy studies, the reference standard used to diagnose or exclude EONS and to verify index test results was not the same for all patients.

A further step, important in the analysis, is to assess whether the reference standard was independent of the index test and the comparator of the index test [80, 87]. When the index test or its comparator forms part of the reference standard, incorporation bias may occur. Unfortunately, more than half of the studies included in our systematic review incorporated the comparator of the index test such as CRP in the reference standard panel. In such situations, it is likely that persons interpreting the results of the comparator will have knowledge of the results of the other test (index test and reference standard) [77].

Also, measures of diagnostic accuracy will be biased if not all participants receive the reference standard and if not all participants are included in the analysis [75, 77]. Consequently, participants included in the analysis may have differed in systematic ways from participants who were not included and this may distort the test accuracy. It is therefore important to describe (and illustrate through a flow diagram [78, 88]) how many participants satisfying inclusion criteria failed to undergo the index test and/or the reference standard and the reasons for failing to do so. This was reported in a minority of the publications included in our analysis.

6.1.4. Test Results: Pre-Hoc versus Post-Hoc Definition of Cutoff Values: Estimates

A crucial index test domain is the timing in which criteria, in particular the cutoff values, were determined [78]. If the cutoffs are selected based on analyzing the completed study results, their post hoc selection may

artificially inflate observed diagnostic test accuracy [89, 90]. Cutoffs should be determined and defined before the start of the study, and this must be stated in the report. Unfortunately, only one study used pre-specified cutoff points (as upper limits of the reference sample over the first 48 h of age) on the index test [32].

Reporting quality should also address blinding, which is not commonly included in the diagnostic study designs, although expected in clinical trial designs [90]. Blinding is essential to ensure that knowledge of the reference standard or the index test results did not influence interpretation of the other test, which may lead to the distortion of measures of diagnostic accuracy [77, 80]. Information about blinding was reported in very few evaluated reports. Notably information concerning PCT reproducibility as well as the number and training of the persons executing and evaluating the index test and reference standard, was also reported in a minority of reports included in our systematic review. This is particularly alarming because limited test reproducibility, an effect of instrumental and/or observer variability, can introduce imprecise or biased accuracy estimates of the sensitivity and specificity [91], while lack of knowledge of the numbers, training, and experience of operators can mask true estimates of the repeatability in different settings [78]. Adherence to STARD also includes reporting how indeterminate index test or reference standard results were handled, and how missing data on the index test or reference standard were handled, since indeterminate results should be considered missing data, and missing data may bias conclusions [90]. Again, this item was reported in few studies of our analysis.

Like measures of effect in therapeutic trials, estimates of diagnostic accuracy are also subject to chance variation [76], and if sensitivity and specificity estimates are reported without a measure of precision, clinicians cannot know the range within which the true values of the indices are likely to lie [92]. Confidence intervals, which are important to judge the reliability of the estimates of diagnostic accuracy, were reported in more than half of the included studies. However, most of the studies included in

our analysis failed to report and cite the statistical methods used to calculate the reference intervals.

Overall, our analysis of PCT accuracy studies for diagnosing EONS shows 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 the variation between studies, is less than optimal. For this reason, assessment of the validity of the included studies (in terms of complete and accurate reporting) should be an essential component of any critical appraisal of the competence with which a primary study on PCT use in the diagnosis of EONS has been undertaken. This obviously would ensure that all relevant information has been provided to guide decisions on the use and interpretation of PCT test results in the management of patients with EONS.

6.1.5. Nosocomial Neonatal Sepsis and Quality of Reporting of PCT Accuracy Studies

There are also several points that should be taken into consideration when evaluating the quality of reporting of PCT accuracy studies for diagnosing late-onset neonatal sepsis. We will provide some examples but this is not a complete list. First, when interpreting studies of the performance of PCT as marker, consideration needs to be given to the assay used. For instance, Auriti et al. [57] and Isidor et al. [59] used a fast immune-chromatographic method (PCT-Q; BRAHMS). This semi-quantitative method, however, is insufficiently accurate in the measurement of PCT at four predefined reference intervals (<0.5, 0.5 to <2.0, 2.0 to <10.0, and ≥10 ng/mL). In addition, falsely increased PCT has been reported for this assay due to interference by human antimouse antibodies. In view of these limitations it is clear that any assessment based on analysis of receiver operating characteristic curves generated using PCT-Q test is useless [57, 59]. Second, a disease definition is critical for the selection of patients for clinical studies that will examine diagnostic and prognostic testing methods across the entire NICU length of stay. Unfortunately, very often investigators used different reference standards

for diagnosing nosocomial sepsis and verifying index test results [55-61] (differential verification bias). This is especially a problem if the reference standards differ in their definition of the target condition. Most studies established a diagnosis of nosocomial sepsis combining positive or negative culture results with clinical and/or laboratory features possibly compatible with late-onset neonatal sepsis [55-61, 63]. However, in only two studies definition of culture-positive sepsis required that the isolated bacteria were not due to contamination or central line colonization [54, 55]. Furthermore, in most studies the definition of clinical septicemia consisted of subtle, ambiguous (and non-validated) clinical signs that overlap with the extremely or very premature infant's struggle to survive after leaving the protective environment of the uterus too early [55-59, 61]. Unfortunately, in all studies the presentation of "persistently" abnormal clinical signs or inflammatory biomarkers to substantiate the presence of SIRS/sepsis were not recorded or included. On the other hand, it is bothersome to think that while in some studies patient controls were not included [56], in other investigations controls consisted of healthy newborns with normal clinical and laboratory findings [60], very low birth weight infants with asymptomatic status [61], or symptomatic babies with culture-negative blood cultures [54, 62] or uncertain clinical and laboratory findings for excluding infection [58]. Third, clinical investigations have focused on the accuracy of a single measurement of PCT [55-57, 59, 61-63]. However, in the case of a moving target such as nosocomial sepsis, less focus has been placed on the usefulness of PCT to identify and monitor the clinical and therapeutic outcome. However, since neonatal sepsis, in particular nosocomial neonatal sepsis, is a dynamic, complex and heterogeneous condition [93], the change in clinical status(clinical and laboratory signs), illness severity and therapy-based illness severity, as well as in PCT host response over time can be very informative for the full potential clinical use of this biomarker in babies requiring intensive care. Last, no studies that provided evidence of the effect of the choice of the PCT threshold value were identified. We conclude that PCT as a marker of neonatal nosocomial infection has been assessed and reported with a

striking lack of consideration of what we want the biomarker to add regarding relevant information on diagnostics and outcome.

REFERENCES

[1] Squire, E., Favara, B. and Todd, J. (1979). Diagnosis of neonatal bacterial infection: hematologic and pathologic findings in fatal and nonfatal cases. *Pediatrics*, 64(1):60-64.

[2] Chiesa, C., Panero, A., Osborn, J.F., Simonetti, A.F. and Pacifico, L. (2004). Diagnosis of neonatal sepsis: a clinical and laboratory challenge. *Clinical Chemistry*, 50(2):279-287.

[3] Lukacs, S. L. and Schrag, S. J. (2012). Clinical sepsis in neonates and young infants, United States, 1988-2006. *Journal of Pediatrics*, 160(6):960-965.

[4] Young Infants Clinical Signs Study Group (2008). Clinical signs that predict severe illness in children under age 2 months: a multicentre study. *Lancet*, 371 (9607):135-142.

[5] Johnson, C. E., Whitwell, J. K., Pethe, K., Saxena, K. and Super, D. M. (1997). Term newborns who are at risk for sepsis: are lumbar punctures necessary? *Pediatrics*, 99(4):E10.

[6] Wynn, J. L. (2016). Defining neonatal sepsis. *Current Opinion in Pediatrics*, 28(2):135-140.

[7] Franz, A. R., Kron, M., Pohlandt, F. and Steinbach, G. (1999). 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. *Pediatric Infectious Diseases Journal*, 18:666–671.

[8] Franz, A. R., Steinbach, G., Kron, M. and Pohlandt, F. (1999). Reduction of unnecessary antibiotic therapy in newborn infants using interleukin-8 and C-reactive protein as markers of bacterial infections. *Pediatrics*, 104:447–453.

[9] Nupponen, I., Andersson, S., Jarvenpaa, A-L., Kautiainen, H. and Repo H. (2001). Neutrophil CD11b expression and circulating

interleukin-8 as diagnostic markers for early-onset neonatal sepsis. *Pediatrics*, 108:E12.

[10] Buck, C., Bundschu, J., Gallati, H., Bartmann, P. and Pohlandt, F. (1994). Interleukin-6: a sensitive parameter for the early diagnosis of neonatal bacterial infection. *Pediatrics*, 93:54–58.

[11] Panero, A., Pacifico, L., Rossi, N., Mancuso, G., Stegagno, M. and Chiesa, C. (1997). Interleukin 6 in neonates with early and late onset infection. *Pediatric Infectious Diseases Journal*, 16:370–375.

[12] Benitz, W. E., Han, M. Y., Madan, A. and Ramachandra, P. (1998). Serial serum C-reactive protein levels in the diagnosis of neonatal infection. *Pediatrics*, 102(4):E41.

[13] Royston, P. (1991). Constructing time-specific reference ranges. *Statistics in Medicine*, 10(5):675-690.

[14] Coffin, C. M., Hamilton, M. S., Pysher, T. J., Bach, P., Ashwood, E., Schweiger, J., Monahan, D., Perry, D., Rogers, B. B., Brugnara, C., Rutledge, J., Weiss, R., Ash, O., Hill, H., Meikle, W., Roberts, W. and Geaghan, S. (2002). Pediatric laboratory medicine: current challenges and future opportunities. *American Journal of Clinical Pathology*, 117(5):683-690.

[15] Van Lente, F. and Pippenger, C. E. (1987). The pediatric acute care laboratory. *Pediatric Clinics of North America*, 34(1):231-246.

[16] Chiesa, C., Osborn, J.F., Pacifico, L., Natale, F. and De Curtis, M. (2011). Gestational- and age-specific CRP reference intervals in the newborn. *Clinica Chimica Acta*, 412(19-20): 1889-1890.

[17] Schelonka, R. L., Yoder, B. A., des Jardins, S. E., Hall, R. B. and Butler, J. (1994). Peripheral leukocyte count and leukocyte indexes in healthy newborn term infants. *Journal of Pediatrics*, 125(4):603-606.

[18] Manroe, B. L., Weinberg, A. G., Rosenfeld, C. R. and Browne, R. (1979). The neonatal blood count in health and disease. I. Reference values for neutrophilic cells. *Journal of Pediatrics*, 95(1):89-98.

[19] Escobar, G. J., Li, D. K., Armstrong, M. A., Gardner, M. N., Folck, B. F., Verdi, J. E., Xiong, B. and Bergen R. (2000). Neonatal sepsis

workups in infants ≥ 2000 grams at birth: A population-based study. *Pediatrics,* 106(2 Pt 1):256-263.

[20] Christensen, R. D., Henry, E., Jopling, J. and Wiedmeier, S. E. (2009). The CBC: reference ranges for neonates. *Seminars in Perinatology,* 33(1):3-11.

[21] Peevy, K. J., Grant, P. H. and Hoff, C. J. (1982). Capillary venous differences in neonatal neutrophil values. *American Journal of Diseases of Children,* 136(4):357-358.

[22] Kayiran, S. M., Ozbek, N., Turan, M. and Gürakan, B. (2003). Significant differences between capillary and venous complete blood counts in the neonatal period. *Clinical and Laboratory Haematology,* 25(1):9-16.

[23] Schmutz, N., Henry, E., Jopling, J. and Christensen, R.D. (2008). Expected ranges for blood neutrophil concentrations of neonates: the Manroe and Mouzinho charts revisited. *Journal of Perinatology,* 28(4):275-281.

[24] Hasan, R., Inoue, S. and Banerjee, A. (1993). Higher white blood cell counts and band forms in newborns delivered vaginally compared with those delivered by cesarean section. *American Journal of Clinical Pathology,* 100(2):116-118.

[25] Da Silva, O., Ohlsson, A. and Kenyon, C. (1995). Accuracy of leukocyte indices and C-reactive protein for diagnosis of neonatal sepsis: a critical review. *Pediatric Infectious Diseases Journal,* 14:362–366.

[26] Gutteberg, T. J., Haneberg, B. and Jorgensen, T. (1984). Lactoferrin in relation to acute phase proteins in sera from newborn infants with severe infections. *European Journal of Pediatrics,* 142:37–39.

[27] Forest, J-C., Lariviere, F., Dolce, P., Masson, M. and Nadeau, L. (1986). C-Reactive protein as biochemical indicator of bacterial infection in neonates. *Clinical Biochemistry,* 19:192–194.

[28] Schouten-Van Meeteren, N. Y. N., Rietveld, A., Moolenaar, A. J. and Van Bel, F. (1992). Influence of perinatal conditions on C-reactive protein production. *Journal of Pediatrics,* 120:621–624.

[29] Chiesa, C., Signore, F., Assumma, M., Buffone, E., Tramontozzi, P., Osborn, J. F. and Pacifico, L. (2001). Serial measurements of C-reactive protein and interleukin-6 in the immediate postnatal period: reference intervals and analysis of maternal and perinatal confounders. *Clinical Chemistry*, 47(6):1016-1022.

[30] Chiesa, C., Natale, F., Pascone, R., Osborn, J. F., Pacifico, L., Bonci, E. and De Curtis, M. (2011). C reactive protein and procalcitonin: reference intervals for preterm and term newborns during the early neonatal period. *Clinica Chimica Acta*, 412(11-12):1053-1059.

[31] Monneret, G., Labaune, J. M., Isaac, C., Bienvenu, F., Putet, G. and Bienvenu, J. (1997). Procalcitonin and C-reactive protein levels in neonatal infections. *Acta Paediatrica*, 86(2): 209-212.

[32] Chiesa, C., Panero, A., Rossi, N., Stegagno, M., De Giusti, M., Osborn, J. F. and Pacifico, L. (1998). Reliability of procalcitonin concentrations for the diagnosis of sepsis in critically ill neonates. *Clinical Infectious Diseases*, 26(3):664-672.

[33] Turner, D., Hammerman, C., Rudensky, B., Schlesinger, Y., Goia, C. and Schimmel, M. S. (2006). Procalcitonin in preterm infants during the first few days of life: introducing an age related nomogram. *Archives of disease in childhood. Fetal and neonatal edition*, 91(4):F283-286.

[34] Hahn, W. H., Song, J. H., Park, I. S., Kim, H., Park, S. and Oh, M. H. (2015). Reference intervals of serum procalcitonin are affected by postnatal age in very low birth weight infants during the first 60 days after birth. *Neonatology*, 108(1):60-64.

[35] Fukuzumi, N., Osawa, K., Sato, I., Iwatani, S., Ishino, R., Hayashi, N., Iijima, K., Saegusa, J. and Morioka, I. (2016). Age-specific percentile-based reference curve of serum procalcitonin concentrations in Japanese preterm infants. *Scientific Reports*, 6:23871.

[36] Garg, B., Sharma, D. and Farahbakhsh, N. (2018). Assessment of sickness severity of illness in neonates: review of various neonatal illness scoring systems. *Journal of maternal, fetal &neonatal medicine*, 31(10):1373-1380.

[37] Chiesa, C., Pacifico, L., Osborn, J. F., Bonci, E., Hofer, N. and Resch, B. (2015). Early-onset neonatal sepsis: still room for improvement in procalcitonin diagnostic accuracy studies. *Medicine (Baltimore)*, 94(30): e1230.

[38] Morse, S., Groer, M., Shelton, M. M., Maguire, D. and Ashmeade, T. (2015). A systematic review: the utility of the revised version of the score for Neonatal Acute Physiology among critically ill neonates. *Journal of Perinatal & Neonatal Nursing*, 29(4):315-344.

[39] Richardson, D. K., Gray, J. E., McCormick, M. C., Workman, K. and Goldmann, D. A. (1993). Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. *Pediatrics*, 91(3):617-623.

[40] Richardson, D., Tarnow-Mordi, W. O. and Lee, S. K. (1999). Risk adjustment for quality improvement. *Pediatrics*, 103 (1 Suppl E):255-265.

[41] Pacifico, L., Osborn, J. F., Natale, F., Ferraro, F., De Curtis, M. and Chiesa, C. (2013). Procalcitonin in pediatrics. *Advances in Clinical Chemistry*, 59:203-263.

[42] Vouloumanou, E. K., Plessa, E., Karageorgopoulos, D. E., Mantadakis, E. and Falagas, M. E. (2011). Serum procalcitonin as a diagnostic marker for neonatal sepsis: a systematic review and meta-analysis. *Intensive Care Medicine*, 37(5):747-762.

[43] Nylen, E. S., Whang, K. T., Snider, R. H. Jr., Steinwald, P. M., White, J. C. and Becker, K. L. (1998). Mortality is increased by procalcitonin and decreased by an antiserum reactive to procalcitonin in experimental sepsis. *Critical Care Medicine*, 26(6):1001-1006

[44] Chiesa, C., Pellegrini, G., Panero, A., Osborn, J.F., Signore, F., Assumma, M. and Pacifico, L. (2003). C-reactive protein, interleukin-6, and procalcitonin in the immediate postnatal period: influence of illness severity, risk status, antenatal and perinatal complications, and infection. *Clinical Chemistry*, 49(1):60-68.

[45] Assumma, M., Signore, F., Pacifico, L., Rossi, N., Osborn, J. F. and Chiesa, C. (2000). Serum procalcitonin concentrations in term

delivering mothers and their healthy offspring: a longitudinal study. *Clinical Chemistry*, 46(10):1583-1587.

[46] Gendrel, D. and Bohuon, C. (1997). Procalcitonin, a marker of bacterial infection. *Infection*, 25(3):133-134.

[47] Blommendahl, J., Janas, M., Laine, S., Miettinen, A. and Ashorn, P. (2002). Comparison of procalcitonin with CRP and differential white blood cell count for diagnosis of culture-proven neonatal sepsis. *Scandinavian Journal of Infectious Diseases*, 34(8):620-622.

[48] Janota, J., Stranák, Z., Bělohlávková, S., Mudra, K. and Simák, J. (2001). Postnatal increase of procalcitonin in premature newborns is enhanced by chorioamnionitis and neonatal sepsis. *European Journal of Clinical Investigation*, 31(11):978-983.

[49] Bonac, B., Derganc, M., Wraber, B. and Hojker, S. (2000). Interleukin-8 and procalcitonin in early diagnosis of early severe bacterial infection in critically ill neonates. *Pflugers Arch: European journal of physiology*, 440(5 Suppl):R72-74.

[50] Lapillonne, A., Basson, E., Monneret, G., Bienvenu, J. and Salle, B. L. (1998). Lack of specificity of procalcitonin for sepsis diagnosis in premature infants. *Lancet*, 351(9110): 1211-1212.

[51] van Rossum, A. M., Wulkan, R.W. and Oudesluys-Murphy, A. M. (2004). Procalcitonin as an early marker of infection in neonates and children. *Lancet Infectious Diseases*, 4(10): 620-630.

[52] Bell, S. G. (2017). Procalcitonin and Neonatal Sepsis: Is This the Biomarker We Are Looking For? *Neonatal Network*, 36(6):380-384.

[53] Tsai, M. H., Hsu, J. F., Chu, S. M., Lien, R., Huang, H. R., Chiang, M. C., Fu, R. H., Lee, C. W. and Huang Y. C. (2014). Incidence, clinical characteristics and risk factors for adverse outcome in neonates with late-onset sepsis. *Pediatric Infectious Diseases Journal*, 33(1):e7-e13.

[54] López Sastre, J. B., Pérez Solís, D., Roqués Serradilla, V., Fernández Colomer, B., Coto Cotallo, G. D., Krauel Vidal, X., Narbona López, E., García del Río, M., Sánchez Luna, M., Belaustegui Cueto, A., Moro Serrano, M., Urbón Artero, A., Alvaro Iglesias, E., Cotero Lavín, A., Martínez Vilalta, E., Jiménez Cobos, B. and Grupo de

Hospitales Castrillo. (2006). Procalcitonin is not sufficiently reliable to be the sole marker of neonatal sepsis of nosocomial origin. *BMC Pediatrics*, 6:16.

[55] Jacquot A., Labaune, J. M., Baum, T. P., Putet, G. and Picaud, J. C. (2009). Rapid quantitative procalcitonin measurement to diagnose nosocomial infections in newborn infants. *Archives of disease in childhood. Fetal and neonatal edition,* 94(5):F345-348.

[56] Fendler W. M. and Piotrowski, A. J. (2008). Procalcitonin in the early diagnosis of nosocomial sepsis in preterm neonates. *Journal of Paediatrics and Child Health*, 44(3):114-118.

[57] Auriti, C., Fiscarelli, E., Ronchetti, M. P., Argentieri, M., Marrocco, G., Quondamcarlo, A., Seganti, G., Bagnoli, F., Buonocore, G., Serra, G., Bacolla, G., Mastropasqua, S., Mari, A., Corchia, C., Prencipe, G., Piersigilli, F., Ravà, L. and Di Ciommo, V. (2012). Procalcitonin in detecting neonatal nosocomial sepsis. *Archives of disease in childhood. Fetal and neonatal edition,* 97(5):F368-F370.

[58] Turner, D., Hammerman, C., Rudensky, B., Schlesinger, Y. and Schimmel, M. S. (2006). The role of procalcitonin as a predictor of nosocomial sepsis in preterm infants. *Acta Paediatrica*, 95(12):1571-1576.

[59] Isidor, B., Caillaux, G., Gilquin, V., Loubersac, V., Caillon, J., Roze, J. C. and Gras-le Guen, C. (2007). The use of procalcitonin in the diagnosis of late-onset infection in neonatal intensive care unit patients. *Scandinavian Journal of Infectious Diseases,* 39(11-12):1063-1066.

[60] Ucar, B., Yildiz, B., Aksit, M. A., Yarar, C., Colak, O., Akbay, Y. and Colak, E. (2008). Serum amyloid A, procalcitonin, tumor necrosis factor-alpha, and interleukin-1beta levels in neonatal late-onset sepsis. *Mediators of Inflammation,* 2008:737141.

[61] Vazzalwar, R., Pina-Rodrigues, E., Puppala, B. L., Angst, D. B. and Schweig, L. (2005). Procalcitonin as a screening test for late-onset sepsis in preterm very low birth weight infants. *Journal of Perinatology*, 25(6):397-402.

[62] Bohnhorst, B., Lange, M., Bartels, D. B., Bejo, L., Hoy, L. and Peter C. (2012). Procalcitonin and valuable clinical symptoms in the early detection of neonatal late-onset bacterial infection. *Acta Paediatrica*, 101(1):19-25.

[63] Enguix, A., Rey, C., Concha, A., Medina, A., Coto, D. and Diéguez, M. A. (2001). Comparison of procalcitonin with C-reactive protein and serum amyloid for the early diagnosis of bacterial sepsis in critically ill neonates and children. *Intensive Care Medicine*, 27(1):211-215.

[64] Gray, J. E., Richardson, D. K., McCormick, M. C., Workman-Daniels, K. and Goldmann, D. A. (1992). Neonatal therapeutic intervention scoring system: a therapy-based severity-of-illness index. *Pediatrics*, 90(4):561-567.

[65] Oygur, N., Ongun, H. and Saka, O. (2012). Risk prediction using a neonatal therapeutic intervention scoring system in VLBW and ELBW preterm infants. *Pediatrics International*, 54(4):496-500.

[66] Wu, P. L., Lee, W. T., Lee, P. L. and Chen, H. L. (2015). Predictive power of serial neonatal therapeutic intervention scoring system scores for short-term mortality in very-low-birth-weight infants**.** *Pediatrics and Neonatology*, 56(2):108-113.

[67] Yu, Z., Liu, J., Sun, Q., Qiu, Y., Han, S. and Guo, X. (2010). The accuracy of the procalcitonin test for the diagnosis of neonatal sepsis: a meta-analysis. *Scandinavian Journal of Infectious Diseases*, 42(10):723-733

[68] Pontrelli, G., De Crescenzo, F., Buzzetti, R., Jenkner, A., Balduzzi, S., Calò Carducci, F., Amodio, D., De Luca, M., Chiurchiù, S., Davies, E. H., Copponi, G., Simonetti, A., Ferretti, E., Di Franco, V., Rasi, V., Della Corte, M., Gramatica, L., Ciabattini, M., Livadiotti, S. and Rossi P. (2017). Accuracy of serum procalcitonin for the diagnosis of sepsis in neonates and children with systemic inflammatory syndrome: a meta-analysis. *BMC Infectious Diseases*, 17(1):302.

[69] Robinson, P. and De S. K. (2018). How to use… Procalcitonin. *Archives of disease in childhood. Fetal and neonatal edition*, Epub ahead of print January 17.

[70] Benitz, W. E. (2010). Adjunct laboratory tests in the diagnosis of early-onset neonatal sepsis. *Clinics in Perinatology*, 37(2):421-438.

[71] Meem, M., Modak, J. K., Mortuza, R., Morshed, M., Islam, M. S. and Saha, S. K. (2011). Biomarkers for diagnosis of neonatal infections: A systematic analysis of their potential as a point-of-care diagnostics. *Journal of Global Health*, 1(2):201-209.

[72] Kordek, A., Torbé, A., Tousty, J., Łoniewska, B., Podraza, W., Nikodemski, T. and Rudnicki, J. (2017). The Determination of Procalcitonin Concentration in Early-Onset Neonatal Infection: a valuable test regardless of prenatal antibiotic therapy. *Clinical Pediatrics (Phila)*, 56(4):333-340.

[73] Cook, C., Cleland, J. and Huijbregts, P. (2007). Creation and Critique of Studies of Diagnostic Accuracy: Use of the STARD and QUADAS Methodological Quality Assessment Tools. *Journal of manual & manipulative therapy*, 15(2):93-102.

[74] Escobar, G. J. (2003). Effect of the systemic inflammatory response on biochemical markers of neonatal bacterial infection: a fresh look at old confounders. *Clinical Chemistry*, 49(1):21-22.

[75] Reitsma, J. B., Rutjes, A. W. S., Whiting, P., Vlassov, V. V., Leeflang, M. M. G. & Deeks, J. J. (2009). Assessing methodological quality. In: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. (Eds.), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0.0. Chapter 9:1-28. The Cochrane Collaboration. Available from: http://srdta.cochrane.org/.

[76] Coppus, S. F., van der Veen, F., Bossuyt, P. M. and Mol, B. W. (2006). Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertility and Sterility*, 86(5):1321-1329.

[77] Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M. and Kleijnen, J. (2003). The development of QUADAS: a tool for the

quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 3:25.

[78] Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C. and Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clinical Chemistry*, 49(1):7-18.

[79] Whiting, P. F., Weswood, M. E., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. N. and Kleijnen, J. (2006). Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology*, 6:9.

[80] Ransohoff, D. F. and Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New England Journal of Medicine*, 299(17):926-930.

[81] Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bonsel, G. J., Prins, M. H., van der Meulen, J. H. and Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*, 282(11):1061-1066.

[82] Rutjes, A. W., Reitsma, J. B., Vandenbroucke, J. P., Glas, A. S. and Bossuyt, P. M. (2005). Case-control and two-gate designs in diagnostic accuracy studies. *Clinical Chemistry*, 51(8):1335-1341.

[83] Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S. and QUADAS-2 Steering Group. (2013).A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *Journal of Clinical Epidemiology*, 66(10):1093-1104.

[84] Philbrick, J. T., Horwitz, R. I., Feinstein, A. R., Langou, R. A. and Chandler, J. P. (1982). The limited spectrum of patients studied in exercise test research. Analyzing the tip of the iceberg. *JAMA*, 248(19):2467-2470.

[85] Rutjes, A. W., Reitsma, J. B., Di Nisio, M., Smidt, N., van Rijn, J. C. and Bossuyt, P. M. (2006). Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal*, 174(4):469-476.

[86] Bossuyt, P. M. & Leeflang, M. M. (2008). Developing Criteria for Including Studies. In: Bossuyt, P. M. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 0.4. Chapter 6:1-7. The Cochrane Collaboration.

[87] Worster, A. and Carpenter, C. (2008). Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. *Canadian journal of emergency medical care*, 10(2):174-175.

[88] Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D. and de Vet, H. C. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Clinical Chemistry*, 49(1):1-6.

[89] Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., Irwig, L., Levine, D., Reitsma, J. B., de Vet, H. C. and Bossuyt, P. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*, 6(11): e012799.

[90] Hughes, J. M., Penney, C., Boyd, S. and Daley, P. (2018). Risk of bias and limits of reporting in diagnostic accuracy studies for commercial point-of-care tests for respiratory pathogens. *Epidemiology and Infection*, 146(6):747-756.

[91] Korevaar, D. A., van Enst, W. A., Spijker, R., Bossuyt, P. M. and Hooft, L. (2014). Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evidence based medicine*, 19(2):47-54.

[92] Harper, R. and Reeves, B. (1999). Reporting of precision of estimates for diagnostic accuracy: a review. *BMJ*, 318(7194):1322-1323.

[93] Lynn, L. A. (2014). The diagnosis of sepsis revisited - a challenge for young medical scientists in the 21st century. *Patient Safety in Surgery*, 8(1):1.