Trajectory of Metabolic Derangement in Infants with Necrotizing Enterocolitis Should Drive Timing and Technique of Surgical Intervention

Joseph J Tepas III, MD, Cynthia L Leaphart, MD, Donald Plumley, MD, Renu Sharma, MD, Brian G Celso, PhD, Pamela Pieper, PhD, ARNP, PNP-BC, Jennifer Quilty, ARNP, Veronica Esquivia-Lee

BACKGROUND: Seven clinical metrics of metabolic derangement (MD7) have improved the timing of surgical intervention in infants with necrotizing enterocolitis (NEC). We compared surgical NEC outcomes based on MD7 at our center (unit S) with a similar center (unit B) that based its intervention on abdominal radiograph.

STUDY DESIGN: Premature infants undergoing surgical care for NEC were evaluated. MD7 included positive blood culture, acidosis, bandemia, hyponatremia, thrombocytopenia, hypotension, and neutropenia. Surgical recommendations were stratified as observation or intervention. Good outcomes included full enteric feeding by discharge and poor outcomes were death or dependence on parenteral nutrition. For unit S and unit B, the frequency, median, and mode of MD7 component per case were determined for observation and intervention. Mann-Whitney U test and Wilcoxon matched pairs were used to compare positive MD7 frequency for observation with intervention. Institutional mortality was compared and metabolic severity of unit cohorts was evaluated by incidence of MD7 in each.

RESULTS: From March 2005 to July 2008, forty-one infants at unit S underwent 62 surgical evaluations. Observation was elected in 38 (median 1 MD7 per case, mode 0). Operative intervention occurred in 24 (median 4 MD7 per case, mode 4). Proportional MD7 difference between observation and intervention was significant (p = 0.018, U = 6). From February 2007 to December 2008, sixty-five unit B infants received 81 evaluations, recommending 37 observations (median 2 MD7 per case, mode 2), and 44 interventions (median 3 MD7 per case, mode 3). MD7 proportions between observation and intervention were not significant (p = 0.318, U = 16). Poor outcomes rates for unit S and unit B infants were 24% and 66%, respectively (p = 0.0001). Severity of MD7 did not differ between institutions (p = 0.53, U = 19).

CONCLUSIONS: These data demonstrate variability in surgical approach to NEC. The MD7 panel describes the trajectory of metabolic derangement, defines more timely surgical intervention, and demonstrates that waiting for free air is too late. (J Am Coll Surg 2010;210:847–854. © 2010 by the American College of Surgeons)

During the last 3 decades, necrotizing enterocolitis (NEC) has emerged as one of the most common surgical emergencies encountered in the neonatal period.1,2 As a multifactorial disease primarily afflicting very low birth weight premature infants, effective control of this neonatal catastrophe yields immediate return in better quality of life for survivors and long-term societal gain by decreasing chronic gastrointestinal dysfunction.3-5 Although better control primarily means more effective prevention, improved outcomes imply a need for more precise determination of the natural progression of the disease and optimal timing of operative intervention, when indicated. To this end, numerous investigators are exploring all aspects of etiology, early diagnosis, and more effective intervention, including identification of biomarkers and various immunologic indicators that reflect acuity or progress of disease.6-11 Currently, none of these are readily available Federal Drug Administration–approved
bedside resources that can provide immediate clinical decision support.

As surgical experience with NEC has evolved, primary peritoneal drainage has emerged as an alternative to formal exploration. Like many units, we evaluated our experience with these surgical alternatives and developed a multivariate regression model to identify epidemiologic, demographic, and physiologic variables that contributed substantially to outcomes.\textsuperscript{12-14} From this effort, 7 readily available clinical metrics of metabolic derangement (MD7) were noted to reflect severity of metabolic derangement and were initially used for risk adjustment in outcomes analysis. Continued experience tracking these metrics demonstrated a meaningful correlation between their appearance and disease severity, suggesting a threshold of metabolic derangement and a trajectory of deterioration that warranted immediate surgical exploration.\textsuperscript{15,16} Although these indicators of metabolic derangement were not applied as a formal protocol in the surgical evaluation of infants with NEC, continued focus on them as components of collaborative research between neonatologists and pediatric surgeons indirectly introduced their use in the process of assessment of infants with NEC.\textsuperscript{17} The purpose of this study was to assess the impact of MD7 by comparing outcomes of surgical management of NEC in our unit (S) to a similarly designated and functioning unit (B) in which NEC management was determined by established experience based primarily on radiological evidence of free air. We hypothesized that MD7 could be applied to clinical decision support for assessment and timing of operative intervention in infants with suspected NEC and no radiological evidence of free air.

**METHODS**

After appropriate Institutional Review Board approval from both participating institutions, the records of premature infants referred for surgical evaluation of NEC were reviewed to determine timing and presence of MD7, defined in Table 1. Infants undergoing surgical evaluation for NEC at the neonatal ICU (unit S), where application of MD7 in assessment of NEC had become routine, were compared with a population encountered in a similar neonatal ICU (unit B), to which the concept of tracking the trajectory of metabolic derangement using MD7 had not been disseminated. Because the profile of MD7 was not applied as such in unit B, laboratory results obtained within 6 hours of surgical evaluation were used to compute what the MD7 profile would have been for each clinical assessment performed in unit B. Good outcomes included full enteral feeding by discharge; poor outcomes were death or continued dependence on parenteral nutrition.

**Table 1. Criteria for Components of Metabolic Derangement**

<table>
<thead>
<tr>
<th>Metabolic derangement</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCPos</td>
<td>Blood culture positive within 96 hours</td>
</tr>
<tr>
<td>Acidity</td>
<td>pH $&lt;7.25$ or receiving bicarb/Tris-hydroxymethyl aminomethane</td>
</tr>
<tr>
<td>Bandemia (I/T $&gt;0.2$)</td>
<td>Absolute no. of immature neutrophils/total</td>
</tr>
<tr>
<td>Na $&lt;130$ meq/L</td>
<td>Within last 24 hours</td>
</tr>
<tr>
<td>Pts $&lt;50$ K</td>
<td>Platelet count $&lt;50,000$/mm$^3$</td>
</tr>
<tr>
<td>BP- on pressors</td>
<td>MAP $&lt; $ gestational age OR on any pressor</td>
</tr>
<tr>
<td>Neut # $&lt;2$ K</td>
<td>Absolute neutrophil count $&lt;2,000$/mm$^3$</td>
</tr>
</tbody>
</table>

BP, blood pressure; I/T, ratio of immature to total neutrophils; MAP, mean arterial pressure.

Evaluations were stratified by recommendation of the surgical consultant as observation or intervention. To assess the interactions of each component of MD7 and of the group as a whole, Mann-Whitney U test was used to compare the frequency of presence of each component for observation with intervention. For each neonatal ICU, the median and mode of MD7 component per case was determined for observation and intervention. Unpaired t-test was used to evaluate the role of MD7 in surgical decision making by comparing number of MD7 per case for observation with intervention. Overall outcomes of both institutional samples, as well as infants receiving only observation and those undergoing surgical intervention, were compared using chi-square with $\alpha$ set at 0.05. Within the groups of infants who were initially observed, those who underwent surgical exploration after subsequent assessment were evaluated using Wilcoxon matched pairs to assess difference in frequency of the MD7 components between the initial decision to observe and the subsequent recommendation to operate.

Disease severity of infants undergoing operative exploration was assessed by comparing incidence and outcomes of infants undergoing enterostomy with resection of diseased gut with those who had no resection and were merely explored for confirmation of unsalvageable disease.

Finally, to demonstrate a second application of the MD7 profile for risk adjustment, the overall severity of metabolic derangement of the 2 populations was evaluated by Mann-Whitney U test to compare the frequency of occurrence of each MD7 component in each cohort.

**RESULTS**

From March 2005 to July 2008, 41 infants at unit S underwent 62 surgical evaluations. Observation was elected in 38 (median 1 MD7 per case, mode 0, range 0 to 5). Operative intervention occurred in 24 (median 4 MD7
per case, mode 4, range 1 to 6). Proportional difference of MD7 between observation and intervention (Fig. 1) was significant \( (p = 0.018, U = 6) \), as was number of MD7 per case \( (p < 0.0001, \text{unpaired } t\text{-test}) \). Between February 2007 and December 2008, unit B reported 81 evaluations of 65 infants, recommending observation for 37 (median 2 MD7 per case, mode 2, range 0 to 6), and 44 procedures (median 3 MD7 per case, mode 3, range 1 to 7). MD7 proportions between observation and intervention were not significant \( (p = 0.318, U = 16) \) (Fig. 2), however, number of MD7 per case was significant \( (p = 0.037) \), suggesting that infants undergoing operative care had more severe disease, but that presence of the MD7 components was probably not used to make the determination to operate. This difference in application of MD7 is further demonstrated by analysis of the frequency histograms of the number of MD7 present per case in observation and intervention groups for both units. The distribution of observation and intervention MD7 in unit S demonstrates a skew toward few components in observation versus greater MD7 frequency in infants undergoing surgery (Fig. 3). For unit B, both distributions are Gaussian, with a peak between 3 and 4 components (Fig. 4). Overall mortality for unit S and unit B infants was 34% and 75%, respectively \( (p = 0.0001) \), yet severity of metabolic derangement of each study population did not differ between institutions \( (p = 0.53, U = 19) \). When the populations were stratified by observation and intervention, similar differences in outcomes were noted (Table 2). Although poor outcomes were defined as death or need for continued parenteral nutrition, the disease severity of all infants with poor outcomes resulted in fatality.

The role of application of MD7 in defining the trajectory of metabolic derangement and timing operative intervention was then assessed by pairs analysis of the infants who were initially observed then subsequently explored. Two infants who were first drained then explored in unit S and 1 infant who underwent 2 operative procedures in unit B were excluded, leaving matched pairs of infants who were
initially recommended for observation then subsequently underwent operative exploration. Unit S encountered 9 infants who progressed to operative care, of which 5 (56%) had good outcomes. Unit B reported 14 paired evaluations, of which only 3 (21%) had good outcomes. Figure 5 illustrates the distribution of the MD7 profiles for observation then intervention recommendations in unit S, which, by Wilcoxon matched pairs evaluation, is significantly different (p < 0.016). In addition, these curves are almost parallel, indicating progression of all of the components of MD7. Figure 6 illustrates the same data for unit B, which appears almost random and, on Wilcoxon matched pairs comparison, shows no statistical difference (p = 0.84).

Finally, to provide some definition of disease severity encountered at operation, cases were stratified by resection with enterostomy versus exploration with no gut resection for unsalvageable disease. Table 3 illustrates this experience and indicates that unit S encountered more unsalvageable cases, but had considerably better outcomes for infants undergoing resection (8 of 11 [73%]). Unit B resected 32 of 41 patients, but achieved good outcomes in only 10 (31%), suggesting either attempted surgical rescue of unsalvageable disease or advanced disease that was refractory to emergent resection and diversion.

**DISCUSSION**

The initial identification of these 7 indicators arose from multivariate prospective analysis of outcomes comparing laparotomy with peritoneal drainage. Common demographic characteristics, such as birth weight, estimated gestational age, ethnicity, and gender played no role in predicting outcomes, defined in that study as survival. At that time, the process of risk stratification could have applied a complex mathematical calculation using the multivariate regression weights to each of the 7 metrics defined in Table 1 to assign a summary level of risk or disease acuity. The alternative was a simplified approach that accepted the contribution of each of the MD7 equally as independent

| Table 2. Outcomes Based on Recommendation at Initial Surgical Evaluation |
|-----------------------------|-----------------------------|-----------------------------|
| **Unit** | **Good, n** | **Bad** |
| **Total study population** | 31 | 10* |
| **S** | 16 | 8 |
| **B** | 11 | 31 |
| **Procedure recommended** | | |
| **S** | 15 | 2 |
| **B** | 11 | 12 |

*p < 0.0001, chi-square.

---

| Table 3. Outcomes of Infants Receiving Surgical Intervention |
|-----------------------------|-----------------------------|-----------------------------|
| **Unit** | **No resection (all died)** | **Resection/stoma** | **Good outcomes** |
| **n** | **%** | **n** | **%** | **n** | **%** |
| **S** | 13 | 57 | 11 | 43 | 8 | 73 |
| **B** | 9 | 22 | 32 | 78 | 10 | 31 |
variables contributing to the dependent variable of survival, and using the sum of indicators present per case as a risk adjuster. The second alternative was chosen because of easier potential clinical application. As experience with surgical management of high-risk premature infants with NEC continued to evolve, a Hawthorne effect began to emerge, as neonatologists became increasingly aware of these metabolic metrics and began to consult the surgical service as soon as any infant with early signs of enteritis was identified. Although we have not deployed a formal protocol defining which or how many metabolic derangements mandate operative intervention, awareness of these indicators has sensitized us to the proposed threshold of 3 metabolic derangements as indicative of the need for operative intervention. These data clearly demonstrate the ability of this profile of readily available clinical laboratory assessments to define a trajectory of metabolic derangement in high-risk, low birth weight infants with NEC.

The purpose of this study was comparison of surgical outcomes for treatment of NEC in 2 similar, state-designated regional perinatal intensive care centers to determine whether the increasing emphasis and awareness of the predictive ability of MD7 was, in fact, associated with better outcomes. The results speak for themselves, but must be tempered by the realization that unit S, which is actively engaged in numerous clinical investigations on prevention, early diagnosis, and treatment of NEC, does not use the MD7 profile as a management protocol to define need for operative intervention. Collaborators from unit B, when initially apprised of the study's findings, immediately noted that data from that unit indicated that surgical intervention, both in initial assessment and operative care, did not appear to be requested early enough in the disease process. This can be affirmed by comparison of the MD7 frequencies in S and B (Figs. 3 and 4), which clearly shows a propensity at unit S for early surgical evaluation of infants with minimal evidence of progression of disease.

These findings introduce 2 important questions. First, should a threshold of ≥3 metabolic derangements indicate need for exploration? Previous work from our unit seems to validate this concept, and appears to imply that any 3 of the 7 metrics, rather than an overwhelming dominance of 1 or 2 can be considered the sole indicators for immediate operative intervention. Second, can timing of assessment adequately define rapidly progressive disease? Analysis of operative care focusing on the different incidence of resection with enterostomy between the units suggests that both surgical teams still encountered infants with a severity of disease that was not survivable. Theoretically, this might imply that, in addition to a threshold of MD7 at which surgical intervention should be considered, the number of MD7 fewer than that threshold noted on initial evaluation might dictate the frequency of repeat evaluations. In addition to the established mantra of periodic radiological assessment for free air, arterial or venous gas, complete blood count, and serum electrolytes would be evaluated every 4, 8, or 12 hours, depending on the number of MD7 noted at initial evaluation. Finally, although Wilcoxon matched pair analysis suggests that these metrics are functioning as a single “profile” to define metabolic derangement, are all of these components critical contributors? As the pairs analyses demonstrate, all 7 components increased in frequency as infants progressed from initial observation to operative intervention. The least amount of change occurs in incidence of positive blood culture, which is understandable because of the time required for incubation. The major contribution of this component is on initial evaluation because bacteremia clearly indicates a source of sepsis. Each of the other factors represents rapidly changing physiology and, as such, defines a measurable trajectory of progressive derangement or response to therapy. Preoperative radiological evidence of free intraperitoneal air did not develop in any of the infants undergoing surgery at unit S, but all of the infants at unit B underwent operative care for either evidence of free air or progressive physiologic deterioration despite aggressive medical management. This might explain the advanced level of disease encountered in unit B's surgical cohort.

The major limitation of this study is also the road map to its next chapter. This is an uncontrolled, retrospective comparison of 2 similar centers differing in their approach to NEC in technique of assessment only. The stark difference in outcomes of the unit applying readily available clinical laboratory data to define a trajectory of metabolic derangement and more timely intervention in progressing NEC mandates that the true validity of this approach be confirmed in a multi-institutional prospective observational study. Based on findings of such an exercise, a randomized trial of specific MD7 thresholds or profiles would be appropriate. As such a process of objective validation unfolds; it is likely that emerging technologies that define cell signaling and organ cross talk will combine with evolving nanotechnology to enable even more precise definition of metabolic derangement to be mapped at the bedside with a drop of the infant’s blood. When that day comes, the MD7 approach of tracking readily available clinical data in a manner that defines the trajectory of physiologic derangement and indicates opportunity for pre-emptive surgical disease control will simply provide more precise and sophisticated tools to a process of critical thinking that is nothing more than constant careful clinical surveillance of the pediatric surgeon’s most fragile and vulnerable patients.
Author Contributions

Study conception and design: Tepas, Leaphart, Plumley, Sharma

Acquisition of data: Tepas, Leaphart, Pieper, Quilty, Esquivia-Lee

Analysis and interpretation of data: Tepas, Celso, Pieper

Drafting of manuscript: Tepas, Pieper

Critical revision: Leaphart, Sharma

REFERENCES


Discussion

DR J ALEX HALLER (Baltimore, MD): I am particularly pleased to have the opportunity to open the discussion on Dr Tepas’ excellent paper. I am also pleased that the Council saw fit to include this paper on the program because it gives pediatric surgeons a chance to understand at least one of the papers during the program.

This paper focuses on a very special problem because as Dr Tepas has indicated, we are discussing patients who weigh less than a pound. To put that in proper perspective, you know, this is the size of a pound or a pound and a half small mouth bass. So we are dealing with very small structures as well as small patients.

These patients are also unstable from many standpoints. They are just beginning to breathe. They are just beginning to have any alimentation at all. As he pointed out, there are a number of different presentations of this devastating condition that we see in this age group.

Dr Tepas identified 3 different patterns of the disease. In my experience, 2 are enough. One group of patients includes those who just stop feeding and then get distended and have free abdominal air. They have very little evidence of the metabolic changes that he’s outlined. They don’t seem to have anything else wrong with them. And yet they have perforated.

The second group comprises those who do fit into the category that requires careful monitoring. They are the ones who seem to have a smoldering inflammatory process, mainly associated with ischemia, low oxygen, and in many of these babies, pulmonary disease as well. They are the ones in whom it is difficult, as we follow them, to determine if and when they need to be operated on.

Dr Tepas and his group have focused on this problem for more than 7 or 8 years, I believe, Joe, but most of their data have been presented at pediatric surgery meetings. And that’s another reason I’m delighted this is on this program, because I think it’s important that we compare this disease with anything that’s comparable in the adult population.

That is one of my questions, Joe. Is there a counterpart to this type of ischemic bowel disease that we see in adults? We are very familiar with ischemic bowel disease, but it seems to have a different pattern in babies.

The second question I have for you is, why did you decide to compare the 2? I realize there are 2 different intensive care units, but the unit that you use for comparison, that was not your own, had a very high percentage of perforation. As a matter of fact, most of their patients had free air.

Is that not selecting a pattern of disease in a group who has a different type of presentation from those who were largely present in your own group? Therefore, are they comparable, or are you looking at 2 different kinds of patients with different disease entities?