Final:September 8, 2009
Revised: February 12, 2013
Version 2.0

A Multi-center Randomized Trial of Laparotomy vs. Drainage as the Initial Surgical Therapy for ELBW Infants with Necrotizing Enterocolitis (NEC) or Isolated Intestinal Perforation (IP): Outcomes at 18-22 months Adjusted Age.

Short Title: Necrotizing Enterocolitis Surgery Trial (NEST)

Martin Blakely, MD, MS, Jon Tyson, MD, MPH, Kevin Lally, MD, MS, Robin Ohls, MD, Michelle Walsh, MD, Brenda Poindexter, MD, Ivan Franz, MD, Susan Hintz, MD, Barbara Stoll, MD, Brenda MacKinnon, Georgia McDavid, Abhik Das, PhD, Richard R. Ricketts, MD, , and Rosemary Higgins, MD

ABSTRACT

Background. The outcome of ELBW infants with NEC or IP is poor, and the appropriate initial surgical treatment (laparotomy or drainage) is highly controversial. The mortality rate after surgical therapy for ELBW infants approaches 50% and more than half of survivors have been shown to have some element of neurodevelopmental impairment.[1] Until recently, all studies evaluating laparotomy and drainage have focused on death and other outcomes occurring in the immediate post operative period. Three recent prospective studies (the Network observational study[2,3] and two randomized trials [Moss[4] and Rees[5]]) found that the death rate in the immediate post operative period does not appear to be significantly influenced by the surgical procedure chosen. Although drainage is widely used and simpler than laparotomy, the Network observational study suggested that laparotomy may improve long term outcome (risk adjusted odds ratio for death or disability (0.55; 95% CI= 0.18 – 1.67). This observational study is the only prospective study addressing neurodevelopmental outcome according to different surgical management strategies for NEC or IP patients. The Network is best positioned to conduct a trial to resolve this major treatment dilemma, because of an established trial infrastructure and established routine follow up programs. An editorial in the New England Journal of Medicine[6] pointed out the limitations of current studies evaluating surgical therapies in NEC and highlighted the importance of measuring neurodevelopmental outcome beyond nursery discharge. Over the past several years, each of the above-referenced studies have been very valuable in providing different pieces of evidence regarding which surgical therapy should be considered most effective in ELBW infants with NEC or isolated intestinal perforation. The trial proposed herein is a natural extension of these studies and will provide critically needed information.

Hypotheses. The primary hypothesis is that survival without neurodevelopmental impairment (NDI) at 18-22 months adjusted age will be greater with initial laparotomy rather than with initial drainage.

Methods. Inclusion criteria are 1) birth weight of ≤ 1,000 g, 2) a decision to perform surgery for suspected NEC or IP, 3) the infant is less than or equal to 8 weeks of age (8 0/7 weeks or less) at the time of eligibility assessment, and 4) patient is at a center able to perform both laparotomy and drainage. Exclusion criteria are: 1) major anomaly which influences likelihood of developing primary outcome or affects surgical treatment considerations; 2) congenital infection; 3) prior NEC or IP; 4) prior laparotomy or peritoneal drain placement; and 5) follow-up unlikely and 6) infant for whom full support is not provided (including surgical treatment). To facilitate enrollment, the trial will be discussed with the parents of potentially eligible infants when NEC or IP is first suspected. Consent will be diligently sought for randomization utilizing members of the multidisciplinary research team (i.e. research coordinator, neonatologist, pediatric surgeon). For infants that are not randomized (due to refusal of consent or unwillingness of the neonatologist and/or surgeon to enroll the infant), consent will be sought to enroll the infant into the study as a member of the "preference cohort". These infants will be treated according to the preference of the surgeon and neonatologist and will have similar data collection as for randomized infants. ***Enrollment in the preference cohort was terminated on February 14, 2013.*** Infants will be enrolled by calling RTI International as soon as a decision to perform surgery is made (or via a computer randomization process). Randomized infants will be stratified according to two variables: center and according to the overall risk for death or NDI (higher / lower). The risk stratification formula was primarily developed in the prospective observational model and includes the following variables: birth-weight, gestational age, vasopressor requirement (yes / no), if infant is on high frequency (oscillating or jet) ventilation or not (any modality of conventional ventilation or no mechanical ventilation), pH, FiO2, and preoperative diagnosis (NEC or IP). Treatment beyond the initial surgical management will be unaffected by the trial. Data collection (beyond that for the generic data base) will be limited, and the forms will be based on those used in the observational study. Outcome at 18-22 months will be assessed by evaluators who are masked to the details of the operative intervention(s) that have been performed using standard Network assessments.

Analysis & Interpretation. The primary outcome (death or NDI at 18-22 months corrected age) will be analyzed in standard logistic regression equations in which the predictor variables include center, and level of risk at enrollment as determined by the method described above. Separate analyses will be used to assess whether there is an interaction between treatment (laparotomy or drainage) and disease (NEC or IP). Secondary analyses of randomized patients will be performed using Bayesian methods to assess the (posterior) probability that laparotomy reduces death or NDI relative to that with drainage.

Consent to randomization may be unobtainable before surgery for a sizable proportion of infants. Providing parents at least agree to collection of the same baseline and outcome data as for randomized infants, we will relate surgical treatment to risk adjusted outcome. A finding that the risk adjusted relative risk in the nonrandomized patients is similar to the relative risk observed in the randomized patients will support the validity of the trial results and generalizability of the results among randomized patients to nonrandomized patients. (This would be expected based on a prior meta analysis of studies using a comprehensive cohort design [JAMA 2005]). Whether use of propensity scoring also supports the results among randomized patients will also be assessed.

Sample Size and Resources. Based on our observational study, we hypothesize that 80% of drain infants and 65% of the laparotomy infants will die or be impaired (NNT = only 7 to gain one unimpaired survivor); 300 infants (150 / group) whose primary outcome is determined will be needed (alpha =0.05 & beta = 0.20, two sided Fisher exact test). Enrollment will likely require 4 years (assuming a consent rate of 50%) unless the consent rate is higher than projected. We propose to initiate this trial within Network sites initially given the importance and urgency of the clinical question. Enrollment will be prospectively monitored and if found to be low after 6 months (after 80% of participating centers have IRB approval and are actively participating in the trial), consideration will be given to add additional sites. Despite the difficulties anticipated with the proposed trial, the total expense and effort required will be small relative to other Network trials, and the trial may facilitate future trials involving urgent/emergent therapies, rare diseases, or surgical interventions.

TRIAL OBJECTIVES

1) To assess, as rigorously as possible, whether the initial surgical treatment of ELBW infants who have NEC or IP and require surgical treatment should be a laparotomy or percutaneous drain.

2) To assess whether outcomes at 18-22 months is affected by an interaction between surgical treatment (laparotomy and drainage) and diagnosis (NEC or IP) (that is, whether the more desirable surgical procedure may differ by diagnosis).

3) Determine the accuracy of the surgeon's preoperative diagnosis (NEC or IP) in patients that undergo a laparotomy.

4) To assist in designing future Network trials, to gain experience with three strategies that have been proposed for use when it is difficult or impossible to conduct conventionally powered randomized trials: Bayesian statistical analyses, propensity scoring, and comprehensive cohort methods. (See below).

TRIAL JUSTIFICATION. 1) The outcome of ELBW infants with NEC or IP is poor: 50% die; overall, 72% die or are disabled[1]; **2)** Drainage is commonly used, partly because it is easily performed, defers a laparotomy in critically ill infants, and may avoid a laparotomy altogether in a minority of patients. Yet, initial laparotomy may improve long term neurodevelopmental outcomes (potentially by reducing the maximum severity or duration of adverse inflammatory responses); **3)** The majority of published comparisons of the two surgical approaches focus on predischarge outcomes. We directed a recent prospective Network study of 156 ELBW infants with NEC or IP followed to 18-22 mo corrected age.[2,3] This observational study suggested comparable predischarge outcomes. However, the findings at follow-up were compatible with major benefit of laparotomy over drainage at follow-up (risk adjusted odds ratio for death or disability = 0.55) although the 95% confidence interval was wide (0.18 – 1.67). These findings indicate the importance of performing the most rigorous feasible trial to assess outcomes at ≥18 mo., the earliest age when cerebral palsy rates may be reliably assessed; **4)** Such a trial would be most feasible in the Network because of its large population, high follow-up rate, well standardized assessments, collaborative relationships between neonatologists and pediatric surgeons,[2] and a successful track record with difficult trials. No other current research network has a well functioning follow-up evaluation for neonatal trials. For this reason alone, there may be no other research network that would be willing and able to undertake such a trial. An NIH funded multicenter trial addressing only predischarge outcomes[4] (L Moss, PI) encountered difficulty in enrolling patients and did not reach its enrollment target despite a 5-year enrollment period. An editorial accompanying the publication of the Moss trial also highlighted the importance of neurodevelopmental follow up and the need for a trial with longer term outcomes as a focus.[6] An even more recent RCT in this same patient population also was stopped early due to poor recruitment (UK trial, Rees et al.)[5] **5)** Even in the Network, adequate power would be difficult to achieve to assess 18 month outcomes in a conventional randomized trial, largely because few ELBW infants (~5%) develop NEC or IP, and because valid consent is difficult to obtain for urgent/emergent therapies. **6)** For similar kinds of reasons, almost all therapies used in emergencies and therapies for rare neonatal diseases have been studied only in purely observational studies or in small trials assessing only predischarge outcomes.[7] Yet, there is a long, unfortunate experience in relying on such studies.[8] As with the recent experience with postnatal steroids,[9] reliance only on predischarge outcomes may be misleading even in large multicenter trials. **7)** The trial will have other important strengths: a) It will address the diagnosis and prognosis as well as treatment for individual ELBW infants presenting with signs of NEC or IP; b) The trial will not compete with and may complement the proposed Probiotics Trial (for which our trial might provide useful information on occurrence of NEC vs. IP according to treatment group, severity of clinical illness in affected infants, nature or extent of disease identified at laparotomy, and surgical co-interventions affecting outcome); c) The experience in conducting the trial we proposed may facilitate future trials involving uncommon diseases, urgent/emergent therapies, or other surgical interventions; d) It will be relatively inexpensive as it involves little extra data collection, use of forms and procedures developed in the prior observational study, and a smaller sample size than most Network trials; e) potential collaboration with non-Network center(s) may increase the generalizability of the results and facilitate broad future collaborative projects.

HYPOTHESES

Primary Hypothesis

1. Initial laparotomy rather than drainage will result in a higher rate of survival without NDI at 18-22 months adjusted age among ELBW infants presenting with signs of NEC or IP.

Secondary Hypotheses:

1. The generalizability of the results in randomized patients to nonrandomized patients will be supported by analyses using conventional risk adjusted relative risk among nonrandomized patients (as in a comprehensive cohort design) and by analyses using propensity scoring.

2. The laparotomy group will have a similar or better outcome than the drainage group with respect to other outcomes during the neonatal period (death; death or prolonged parenteral nutrition; death or prolonged hospital stay) and at follow-up (e.g., each component of NDI; anthropometry);

3. The proportion of infants who have surgical complications (e.g. procedure-related liver injury, wound dehiscence, intestinal stricture or fistula) will be similar in the two treatment groups;

4. There will be no significant statistical interaction between treatment and diagnosis.

5. The trial will provide improved methods to distinguish ELBW infants with NEC from those with IP;

6. The trial will provide improved methods to assess the prognosis of ELBW infants presenting with signs of NEC or with IP;

SPECIFIC AIMS.

Primary specific aim:

1. Using conventional frequentist analyses, to assess the relative risk for survival without NDI with initial laparotomy relative to initial drainage among ELBW infants who undergo surgical treatment of NEC or IP.

Secondary specific aims:

1. To determine whether analyses using propensity scoring and using conventional risk adjusted relative risk support the relative risk among randomized infants and the generalizability of the results to nonrandomized infants.

2. Comparing initial laparotomy to initial drainage, to assess the relative risk for secondary outcomes assessed during the neonatal period (e.g., death; death or prolonged parenteral nutrition; specific surgical complications) and at follow-up (e.g., each component of NDI; Bayley III scores);

3. Comparing initial laparotomy to initial drainage, to assess the relative risk for surgical complications (e.g. procedure-related liver hemorrhage; wound dehiscence; intestinal stricture, requiring operation; intestinal fistula);

4. To assess whether there is evidence that the preferred surgical treatment should differ by preoperative diagnosis (NEC or IP) by evaluating whether there is an interaction between surgical treatment and preoperative diagnosis;

5. To develop improved methods to distinguish ELBW infants with NEC from those with IP;

6. To develop improved methods to define the prognosis of individual infants presenting with signs of NEC or IP;

7. To provide information that would be deemed useful by the subcommittee for any other simultaneous clinical trials (to be determined perhaps for the proposed Probiotics Trial).

BACKGROUND/PREVIOUS STUDIES (See also trial justification above.)

Critical Design Issues. We are proposing a conventional randomized trial to be done in the Neonatal Network. The hypotheses for this trial were developed in a Network prospective observational study whose primary goal was to determine the need for and help design such a trial. Two subsequent randomized trials evaluating laparotomy versus drainage in ELBW infants have also been critically evaluated and support the hypotheses and the justification for doing the trial. Because this is a rare disease and the anticipated consent rate is 50%, we are proposing as a secondary objective to explore the utility of various approaches that may be used when the usual sample size requirements are not reasonably attainable[9,11-21] During the design phase of this trial, we also carefully considered other potential innovations, including acceptance of wider confidence intervals,[11,12,14,19] and the use of an adaptive study design.[16] (For example, various adaptive study designs[16] that might be used were investigated with an extensive series of Monte Carlo simulations based on the findings of our observational study. These simulations were used to determine whether an adaptive design would reduce the total number of infants who would need to be studied, the total number who would be assigned to the inferior treatment method, or the total duration of study. Within a broad range of treatment effects [with the relative risk for the primary outcome with laparotomy relative to drainage being as low as 0.5 or as high as 1.5] there appeared to be little if any scientific or ethical advantage to adaptive designs for the particular therapies to be compared.)

We are proposing to conduct this trial purely within the Network as long as adequate enrollment can be achieved. If enrollment is below the predicted rate that would allow completion of the trial in 4 years, we would propose to add additional sites as needed to allow adequate enrollment. Both randomized trials published in this area have been halted prior to achieving the desired sample size. It would be unwise to ignore this history. Thus, we propose to critically evaluate actual versus desired enrollment at 6 months and carefully consider the need / desirability of recruiting additional sites.

Comprehensive cohort design. This design differs from that of a conventional randomized trial by including not only patients whose treatment is randomly assigned but also any whose treatment is based on the preference of the physician or patient (or surrogate). When properly performed, every effort is made in these trials to promote clinical equipoise and obtain informed consent to be randomized. If such consent is refused, consent is sought only for data collection. Whenever possible, the determination of primary outcome is carefully standardized and assessed by masked evaluators. Baseline data are carefully collected and controlled during analysis to adjust for baseline differences between patients who receive different therapies.

This approach has the potential advantage of increasing the number of patients studied and the external validity (generalizability) of the results by involving patients and physicians who otherwise would not be involved.[10] However, an increase in external validity would be important only if the internal validity of the results is not undermined. Based on surveys of purely observational studies, one would expect that inclusion of preference groups would bias the results. This problem might be avoided or minimized because of measures to promote equipoise prior to the study, stringent features of study design and analysis as noted above,

exclusion of centers with such strong biases that they are unwilling to randomize any patients, and the balancing of biases within and between centers that do participate. A systematic review published in JAMA this year[10] assessed all trials using a comprehensive cohort design. The review concluded that "Differences in outcome between randomized and preference groups were generally small, particularly in large trials and after accounting for baseline measures of risk".

In the proposed study, we will use risk adjusted relative risk to assess whether the results for randomized patients are generalizable to nonrandomized patients. If the findings for randomized and nonrandomized patients favor the same surgical therapy (as we would expect from the findings noted above for comprehensive cohort studies), we would presume the results are generalizable to the nonrandomized patients. If the findings are different, we could not be certain whether the difference resulted from chance, true differences in effectiveness for randomized and nonrandomized patients, residual confounding that could not be adequately adjusted in the analysis, or other reasons. However, the last reason would be most likely if as in prior studies, nonrandomized patients treated with drains are at systematically higher risk than those treated with laparotomy (due to severity of illness, time of day, or other factors). Such risk factors may be quite difficult to adequately assess.

The analytic approach noted above is a different and more conservative analytic plan compared to a true comprehensive cohort study. However our trial provides an excellent opportunity for the Network to explore whether analysis conducted as if our trial used a comprehensive cohort design would provide similar results as for a conventional randomized trial in a setting where there is considerable effort to minimize bias.

Analyses Using Propensity Scoring. Propensity scoring (which indicates the probability that individual patients will receive different therapies) has been advocated for use assessing treatments in observational studies when randomization is not feasible. There is ongoing controversy about the extent to which estimates of treatment effect obtained using propensity scoring may be biased by their use. [22-25] We will use this trial as an opportunity to compare the relative risk for the primary adverse outcome with laparotomy among nonrandomized patients with that among randomized patients. As for conventional risk adjusted analyses noted above, results using propensity scoring for nonrandomized patients that favor the same surgical therapy found to be superior among randomized patients would support the generalizability of results obtained for randomized patients. If the findings are different, we could not be certain whether the difference resulted from chance, true differences in effectiveness for randomized and nonrandomized patients, or residual confounding that was not avoided by propensity scoring .

Involvement of Pediatric Surgeons. This trial will build upon the effective working relationships between pediatric surgeons and neonatologists established during the observational study, as evident in the high compliance in obtaining the data needed for the primary manuscript of the observational study. These relationships were promoted in part by the participation of these surgeons in the authorship on a major manuscript from this study published in the Annals of Surgery. Drs. Blakely, Lally, and Ricketts are directing the trial from a "surgical perspective" and are co-authors of the trial proposal. Prior to the proposed trial, Drs. Blakely, Lally, or Ricketts will visit participating sites as needed to meet with the pediatric surgeons to promote their equipoise and secure their agreement and enthusiasm for the trial. If any important problems develop with data collection, compliance or equipoise among the surgeons in any of the centers, Drs. Blakely and/or Lally will telephone the involved surgeons. If necessary they will revisit the Center. The budget for the trial will include expenses not only for these trips but also for a half-day meeting of the surgeons each year at the annual meeting of the American Pediatric Surgical Association or the Surgical Section of the American Academy of Pediatrics meeting. These meetings are well attended and would very likely have surgeons from each participating center present.

METHODS/PROCEDURES

Inclusion Criteria: The inclusion criteria proposed are meant to be broad to allow as many eligible infants to participate as possible. Some surgeons and neonatologists have treatment preferences for some subsets of this broad group of patients. For example, some surgeons are not comfortable randomizing an infant with NEC but no pneumoperitoneum even after a decision has been made that operative intervention is warranted. Some surgeons are not comfortable randomizing larger subgroups of patients (e.g. those with isolated perforation). This trial is designed to allow randomization of as many eligible infants as is feasible, while recognizing that treatment preferences exist and will limit the proportion of infants that are randomized. There are no convincing data driving the treatment of these subgroups; however, preferences do exist and do impact management and will impact agreement for randomization. With these guidelines in mind, inclusion criteria are:

1) birth weight of $\leq$ 1,000 g,
2) a decision by the attending pediatric surgeon to perform surgery for suspected NEC or IP (The indications for surgery for infants with NEC or IP vary among surgeons and sites),
3) the infant is less than or equal to 8 weeks of age (8 0/7 weeks or less) at the time of eligibility assessment, and
4) patient is at a center able to perform both laparotomy and drainage.

There are no evidence-based guidelines to rely on regarding which indications for surgery are valid. The most common indications for surgery in this patient population are pneumoperitoneum discovered on plain abdominal radiographs, and clinical deterioration despite maximal medical management. Since there are no proven indications for surgery, we will rely on the treating neonatologist and pediatric surgeon to continue to decide when operative therapy is warranted, just as they do outside of a study setting. Consistent with these thoughts, the presence of a pneumoperitoneum is not a requirement for eligibility, as many patients with perforation will not have evidence of pneumoperitoneum on preoperative imaging.

Exclusion Criteria: 1) Major anomaly which influences likelihood of developing primary outcome or affects surgical treatment considerations; 2) congenital infection; 3) Prior laparotomy or peritoneal drain placement, 4) Prior NEC or IP, 5) Follow-up unlikely (e.g., mother incarcerated, or currently resides (or plans to move) far from any Network center.), and 6) infant for whom full support is not provided (including surgical treatment).

Consent, Enrollment, and Treatment Assignment. To facilitate enrollment, the trial will be discussed with the parents of potentially eligible infants as soon as feasible after the parent(s) have been informed by the medical staff that their infant has NEC or IP, whether or not there has been a decision to perform surgery. (Depending on the center and circumstances, this discussion might include the coordinator, attending neonatologist, fellow, and/or pediatric surgeon). In addition to printed informational materials, the development of a brief explanatory video or DVD is being considered which would explain the rationale for the trial and provide a standardized method of starting the informed consent process. Informed consent will be diligently sought for both randomization and collection of predischarge and follow-up data. If refused, consent will then be sought only for data collection. We propose to allow variability in the consent process at individual centers with the overall goal in common of achieving as a high a consent rate as possible. One method we are proposing is to have provisional consent for the trial obtained by the neonatologist or Network research coordinator initially. If the family has given this provisional consent, the attending pediatric surgeon would then assess the infant for inclusion in the study. If the surgeon agrees to randomize the infant then enrollment would occur. If the family has not agreed to the initial / provisional consent, the surgeon could also discuss the trial with the family and consent might be obtained at that time. Infants will be enrolled by calling the RTI International (RTI). During "standard hours" this would ideally be done by the Network coordinator. To avoid baseline differences, randomized infants will be stratified according to two variables: center; and according to the overall risk for death or NDI (higher / lower) at enrollment. The risk stratification formula was primarily developed in the prospective observational model and includes the following variables: birth-weight, gestational age, vasopressor requirement (yes / no), if infant is on high frequency (oscillating or jet) ventilation

or not (any modality of conventional ventilation or no mechanical ventilation), pH, FiO2, and preoperative diagnosis (NEC or IP). (There is no intent to power the trial to compare all subgroups). A variable, permuted, block size will be used.

Stratification according to preoperative diagnosis. A unique design feature of this trial (as compared to the two prior RCTs in this patient population) is the stratification utilizing the presumed preoperative diagnosis (NEC or IP). We feel it is critical to make this distinction, to the best of our ability, to ensure maximal comparability of treatment groups and to have the ability to test for interactions between treatment and diagnosis. Based on the Network observational study and other data, the following patient characteristics support a preoperative diagnosis of NEC: presence of pneumatosis on KUB, older age at operation (mean age at operation from observational study = 27.8 days), and presence of portal vein air. A history of enteral feedings prior to disease occurrence has been suggested as supportive of a diagnosis of NEC in other publications. In the Network observational study, only 46% of patients with proven NEC (determined at laparotomy) had a pneumoperitoneum on preoperative imaging.[3]

The following characteristics support the diagnosis of IP: younger age at operation (mean age at operation in Network observational study = 7.4 days), absence of pneumatosis, and presence of pneumoperitoneum on preoperative KUB. Other characteristics that have been used to favor a diagnosis of IP include the radiographic finding of "gasless abdomen" and the physical exam finding of a "blue abdomen". In the Network observational study, the last two characteristics were not statistically different between NEC and IP infants, but the findings supported using these as secondary determinants to help distinguish the two diagnoses.

In the Network observational study, the correlation of the preoperative diagnosis (NEC or IP) with intraoperative diagnosis in patients undergoing laparotomy was 95%. While recognizing that the distinction is not perfect preoperatively, we feel that this is a positive design feature and preoperative diagnosis of the pediatric surgeon will be recorded. Accuracy of this distinction will again be measured in those undergoing laparotomy.

Baseline and Intraoperative Data. Baseline variables include: BW, GA, postnatal age, sex, and multiple measures of illness severity related to outcome in our prior study or other studies (blood pressure, vasopressor therapy, platelet count, pH, FiO2, ventilator type, and settings), whether randomization was refused by the surgeon or parents, and if so, the stated reason. Preoperative and intraoperative observations will be recorded by the pediatric surgeon as in the observational study. Location of operative intervention (NICU or operating room) will be recorded.

Primary outcome variable, as defined above, is death or neurodevelopmental impairment at 18-22 months as assessed in the generic follow-up of ELBW infants in Network centers.

Secondary outcomes variables will include surgical complications, such as wound dehiscence, intestinal stricture or fistula, procedure-related liver hemorrhage; number of surgical procedures; sepsis episodes; duration of parenteral nutrition; parenteral nutrition associated cholestasis; length of hospital stay; rehospitalizations; and each component of the primary outcome.

Treatment. Initial drainage will involve placing a ¼ inch Penrose drain in the lower abdomen with local anesthesia and sedation. The drain will be fed into the peritoneal cavity toward the pelvis and sutured to the skin. There are differences in this technique that are used across the multiple sites (e.g. irrigation or not, right or left sided, etc...) and no specific techniques have been shown to correlate with either effectiveness of drainage or with the primary outcome of our study. Therefore, no formal guidelines regarding the drainage procedure are felt necessary. A standard dressing will be applied at the surgeon's discretion.

Subsequent laparotomy will be allowed after initial drainage, since this trial is designed to test two treatment strategies (i.e. initial drain versus initial laparotomy), rather than a strict comparison of drain only vs. laparotomy only. The study protocol recommends reevaluation of the patient 12 – 24 hours after drain placement to consider the need for subsequent laparotomy. If the patient is improved, no subsequent

laparotomy would be recommended. However, the decision regarding subsequent laparotomy is the responsibility of the treating physicians and not dictated by this study. Factors to be considered in performing a subsequent laparotomy will include the usual scenarios in current practice. Reasons surgeons use to perform a laparotomy after initial drainage include: continued deterioration despite adequate drainage (e.g. increased pressor requirement, worsened coagulation profile, increasing ventilator support, decreasing urine output, increasing or persistent pneumoperitoneum, worsened abdominal wall discoloration or abdominal mass, or others). Alternatively, some surgeons proceed with a subsequent laparotomy if the patient improves after initial drainage, arguing that with improvement the risk for laparotomy has fallen to an acceptable level. This will also be allowed and recorded in this study. The reasons that study patients undergo subsequent laparotomy will be recorded. Other than randomizing the original treatment option, the trial will attempt to allow standard current practice to facilitate enrollment and to increase generalizability.

Initial laparotomy will be performed under general anesthesia in the NICU or operating room. Standard procedures will be used, including inspection of the bowel with resection of diseased areas, creation of stoma(s), and other procedures felt to be indicated by the surgeon. Intraoperative data recorded by the surgeon will include complications, intestinal length removed and remaining, presence of ileocecal valve, etc. As appropriate in a management trial,[21] use of surgical and medical cointerventions will be determined by the attending pediatric surgeon and neonatologist. Patients undergoing initial laparotomy may also undergo subsequent laparotomy as indicated.

Issues Regarding Subsequent Laparotomy after Initial Surgical Treatment (either drain or laparotomy). Data regarding the value of subsequent laparotomy are minimal. In a Pubmed search using "salvage laparotomy and necrotizing enterocolitis" only 6 articles were identified. (2/12/09) In one study specifically addressing this issue, 17 patients were treated with drainage over a 5 year period and only 4 of these had a subsequent laparotomy.[26] These 4 did have a 100% mortality rate vs. 46% mortality in the 13 patients treated with drain only, but these are very small patient numbers from which to make treatment recommendations. Goyal et al., reviewed an 18 year experience and had 42 patients treated with initial drainage. Eleven patients had what was considered to be a poor response to drainage, 6 of these had a rescue laparotomy and 3 survived. These authors considered rescue laparotomy after drainage to be beneficial.[27] Sharma et al. reported 32 infants treated with initial drainage, 8 of whom had an early subsequent laparotomy with 5 survivors and these authors also considered early subsequent laparotomy to be an important and useful therapy.[28]

In the prospective observational study done by the NICHD Neonatal Research Network, 80 infants had an initial drain placement. Only 18 of these (22.5%) had a subsequent laparotomy. In the group treated with drainage only (n=62), the mortality rate was 55%. The mortality rate in the group initially treated with drainage, followed by a subsequent laparotomy, was 68%. The low percentage of patients treated with a subsequent laparotomy after initial drainage provides some evidence that drains are often used as planned definitive therapy, rather than a temporizing measure.[2,3] The high mortality rate in the group of patients that did have a subsequent laparotomy (68%) should give caution to surgeons choosing to employ this treatment strategy.

The UK randomized trial (Rees et al) allowed infants randomized to initial drainage to undergo a subsequent laparotomy >/= 12 hours later if the infant's condition had deteriorated (at the discretion of the surgeon).[5] Infants who were improved or were considered stable did not have a subsequent laparotomy. In this trial, 35 infants were randomized to initial drain placement and 26 did undergo subsequent laparotomy (74%). The median time of the subsequent laparotomy was 2.5 days (range: 0.4 – 21 days). Indications for the laparotomy were clinical deterioration (14), increasing / persistent pneumoperitoneum (12), palpable mass and continued pneumoperitoneum (1), stool drainage from drain site (1), intestinal obstruction (2) (some had more than one of these indications). The relative risk for mortality in the Drain + Laparotomy subgroup was 1.4 (95% CI: 0.6-3.4; p=0.4) compared to the initial laparotomy group. The survival curves show the drain + laparotomy group to be intermediate between the initial laparotomy group and the drain group, with no significant differences between any of the 3 subgroups analyzed. The drain only group in this trial had a mortality rate of 49% compared with 36% for laparotomy. Drain + subsequent laparotomy had an intermediate mortality rate of approximately 42% (exact data not published).

The Moss trial protocol discouraged subsequent laparotomy after initial drainage, stating that infants that survive after initial drainage often deteriorate early but then improve.[4] The protocol does allow early subsequent laparotomy for persistent metabolic acidosis, hemodynamic instability, and respiratory failure, and later laparotomy for obstruction. This trial randomized 55 infants to drainage, 5 of whom had subsequent laparotomy for clinical deterioration (9%) between days 2 and 45 after the initial procedure. Four of these 5 survived. Sixteen additional patients had delayed laparotomy.

Conclusions regarding subsequent laparotomy after initial drainage:

1. When US surgeons are allowed to decide initial treatment (i.e. observational study), the use of subsequent laparotomy is uncommon (approximately 20%).
2. If the initial treatment is chosen by randomization, rather than surgeon preference, the subsequent laparotomy rate may be higher. This is due to the fact that some infants would be randomized to drainage that the surgeon would have treated with laparotomy if chosen by preference.
3. In two randomized trials, some infants were apparently "salvaged" by subsequent laparotomy that was performed due to clinical deterioration as determined by the surgeon.
4. In the UK trial, the majority of initial drain infants (74%) did undergo subsequent laparotomy for a variety of reasons and the survival rate of this group was higher (although not statistically significant) than the drain only group. Based on data from this trial there is no evidence that subsequent laparotomy is harmful.
5. In the Moss trial, only 9% of drain infants had early subsequent laparotomy, but 4 of 5 that did survived. This also provides some evidence that in this trial early subsequent laparotomy was not harmful.
6. Surgeons have access to older retrospective data regarding "salvage" laparotomy, the more recent prospective data summarized above, and most have established practice patterns regarding this question. There are no truly clear, convincing data supporting benefit or harm of this therapy. For sure, some infants do survive after early subsequent laparotomy.
7. Disallowing subsequent laparotomy does not appear to be a realistic option from an ethical standpoint.
8. Limiting subsequent laparotomy to instances in which there is clinical deterioration seems appropriate.

Based on the above literature review, we strongly discourage subsequent laparotomy after initial drainage during the first 12 hours post-drain. Trial infants will be very carefully and continually assessed for their response to the initial surgical therapy and if the patient's condition is deteriorating further, subsequent laparotomy is allowed in the trial. For each operative intervention, the indications for the operation as well as details of the intervention will be recorded prospectively.

Follow-up Evaluation. Infants will be assessed at 18 to 22 months of age, corrected for prematurity, according to the NICHD Neonatal Research Network Follow-up Study Protocol (rev. 10.31.2007) and Manual (rev. 2.28.2008). The follow-up visit includes the following assessments or structured inquiries: medical history including neurosensory outcomes, physical and neurological examination, neurodevelopmental evaluation with the Bayley Scales of Infants Development III, social/behavioral evaluation with the Brief Infant Toddler Social Emotional Assessment (BITSEA), and demographics and socioeconomic status questionnaires. Examiners are certified according to the procedures described and approved by the NICHD Neonatal Research Network.

Neurodevelopmental impairment (NDI) is defined as meeting any of the following criteria: moderate to severe cerebral palsy (CP) with Gross Motor Function Classification System (GMFCS) Level >=2, a Bayley III Cognitive score <85, < 20-200 bilateral vision, and/or permanent hearing loss that does not permit the child to understand the directions of the examiner and communicate despite amplification.

The components of NDI are defined as follows:

Cerebral Palsy: Definite abnormalities in classical neuromotor exam, including tone, deep tendon reflexes, coordination and movement, coupled with a delay in motor milestones with a disorder of motor function. Severity of CP is classified according to the GMFCS level.[30] Moderate CP is defined as GMFCS level 2 or 3; severe CP is defined as GMFCS level 4 or 5.

Psychometric testing: Cognitive outcome at 18-22 months will be assessed by certified psychologists or psychometricians using the BSID III. For this study, the cognitive subtest (mean±SD: 100±15) will be administered. Scores on the cognitive subtest of <70 (>2 SD below the mean) is consistent with severe cognitive impairment.

Sensory deficits are identified through specific question concerning vision and hearing. Children with no useful vision in either eye is consistent with a refraction definition of <20/200 (legally blind). Permanent hearing impairment that does not permit the child to understand the directions of the examiner and communicate despite amplification is consistent with severe auditory impairment and would be adequate to meet the definition of NDI.

An important issue (for this trial and for neurodevelopmental assessment of premature infants in general) is the potential difference in sensitivities of the Bayley II and III products in defining NDI. Some authors have reported recently that the Bayley III likely underestimates NDI, compared to the Bayley II.[31] Because of these issues, for this trial a Bayley III score of <85 will be one of the conditions defining NDI (in addition to those above).

Blinding and Reliability. Blinded collection of predischarge data would not be reasonably feasible for our research nurses. Follow-up personnel who are not told the treatment group will assess all follow-up findings. Because patients in each treatment group are likely to require laparotomy after the initial surgical procedure, it will be difficult to predict treatment group assignment at 18-22 months by simply examining the child, as infants in both groups may have undergone multiple abdominal procedures. The final assessment of impairment will be reviewed by three subcommittee members without knowledge of treatment group who will adjudicate the outcome of any infants whose outcome is uncertain (as was done Hypothermia trial).

Statistical Analyses. The primary analysis will assess the effect of treatment on death or NDI at 18-22 months corrected age. This analysis will be performed using standard statistical analyses often used in multicenter trials: logistic regression equations in which the predictor variables include treatment group and the stratifying variables (in this trial, center, and risk stratification formula). The analyses will include an assessment of whether there is an interaction between treatment (laparotomy or drainage) and disease (NEC or IP). Bayesian analyses[11, 15] of the primary outcome will also be performed because statistical power will be limited to identify treatment effects that are clinically important but smaller than hypothesized and because Bayesian analyses can be used to estimate the (posterior) probability that laparotomy reduces death or NDI at 18-22 months.

To avoid bias in selecting the final regression model for the risk adjusted analyses of the nonrandomized patients, the baseline predictors will be determined before the logistic regression analyses are performed. Selection will be based on biologic plausibility, the results of our prior observational study, and univariate analyses of trial results. Likewise the baseline variables used in the final models for propensity scoring will be determined before any of these analyses are performed. Diagnosis & prognosis will be assessed using methods developed in our prior observational study. Nominal p values will be calculated for secondary outcomes without adjustment for multiple comparisons. Exploratory subgroup differences will be performed.

We plan to collect the same data for eligible, non-randomized infants (with consent) as for randomized infants. This will allow comparison of the patient characteristics and outcomes in the randomized and non-randomized cohorts and perhaps assess the external validity of our study findings (generalizability). As a secondary analysis, the risk adjusted relative risk will be determined in the randomized & preference groups and the value of combining these cohorts will be evaluated. This approach, known as a comprehensive cohort design, has been recommended in a recent meta-analysis for rare diseases or for evaluation of emergency therapies.[10] Whether this approach is useful or not will be a secondary aim and these analyses will not affect the primary analysis of the randomized cohort.

Currently, there is unavoidable uncertainty about differences between the Bayley II and III outcomes. Based on Network GDB data available at this time, it appears that a Bayley III score <85 best approximates that of a Bayley II <70 which has been previously used by the Network in the definition of neurodevelopmental impairment. As this trial progresses, an increasing body of data will accumulate regarding the differences in

the Bayley tests. We plan to critically examine these data (GDB data, pooled data from NEST study subjects, and publicized data from other studies) at yearly intervals during the trial enrollment period. The sample size determination will be reevaluated, and may be revised upwards as the uncertainty regarding the Bayley III lessens (2 years after study commencement). These decisions will be made by the Steering Committee based upon a recommendation by the NEST subcommittee, and without any knowledge of the difference between the two treatment groups in this trial with regard to the proportion of impaired children.

Sample Size. Allowing for a smaller difference between treatment groups in the trial than in our observational study, we hypothesize that 80% of the drain group and 65% of the laparotomy group will die or be impaired at 18-22 mo. (Number needed to treat = only 7 to gain one unimpaired survivor from laparotomy). By PASS software, a total sample size of 300 infants who have their outcome assessed is needed with alpha = 0.05 and beta = 0.20 using 2-sided Fisher exact test.

In the observational study 156 infants were enrolled in 14 centers in 18 months. By extrapolating these numbers to a 16-center trial design, 120 eligible patients / year would be expected. Based on the latest GDB data, using the 16 centers prior to reformatting the Network, an estimated number of eligible babies of 190 is obtained. Our best estimate of the number of eligible patients in the new 16-center Network is 160 / year. We anticipate a 50% enrollment rate, $\geq$95% consent for data collection (as for generic data collection), equal use of initial laparotomy and drains; and a 50% mortality in each group. Assuming a consent rate of 50%, 320 patients would be enrolled in a 4-year period with 160 deaths and 160 survivors. With a 90% follow up rate, 144 survivors would have the primary outcome measured. This would yield 304 patients which will meet our sample size requirement during a 4-year enrollment period. If the consent rate is below 50%, the enrollment period would need to be extended.

We propose that the trial be stopped when any of the following are met: 1) DSMC might stop trial for large difference in neonatal mortality (secondary outcome), which is unlikely given equivalent survival in both observational study and two randomized trials; 2) 300 randomized patients are enrolled and have primary outcome assessed; or 3) a 5-year enrollment period has elapsed.

Interim Monitoring. Adverse events during the course of the trial treatment period will be prospectively monitored, as will clinical morbidities throughout hospitalization. The enrolled population is extremely high risk, and their hospitalization can produce a great number of expected adverse events and clinical diagnoses. Rates of these events, historically observed among similar extremely low gestation/birth weight infants, will be provided to the Data Safety Monitoring Committee (DSMC). Secondary outcomes of interest in this trial that can be compared across the two treatment groups include death; death or prolonged parenteral nutrition; specific surgical complications such as wound dehiscence, intestinal stricture or fistula, procedure-related liver hemorrhage; number of surgical procedures; sepsis episodes; duration of parenteral nutrition; parenteral nutrition associated cholestasis; length of hospital stay; rehospitalizations; and each component of the primary outcome (available only after the 18-22 month follow up visit). The set of common, serious neonatal morbidities as defined in the NICHD Neonatal Research Network (NRN) Generic Data Base (GDB) will also be collected and analyzed in all subjects. These include data on in-hospital growth, the incidence and severity of intraventricular hemorrhage and periventricular white matter damage, seizures, patent ductus arteriosus (PDA) and its treatment, nosocomial sepsis (and organisms), hearing impairment, in-hospital ROP, pneumothorax, and bronchopulmonary dysplasia.

Adverse events that are worse than mild or moderate, are unexpected or unanticipated, and possibly related to the study treatments will be reported within 48 hours of discovery to the Data Coordinating Center (DCC), to NICHD, and to the site Institutional Review Board (IRB). NICHD and the NRN DCC will determine if the event warrants expedited reporting to the Chair of the NRN DSMC.

The NRN DSMC will review and will need to approve the protocol before enrollment begins. The members of the DSMC are independently appointed by the NICHD and include representation from neonatologists, biostatisticians, clinical trialists, and ethicists. The DSMC reports to the head of NICHD.

The DSMC provides a summary of its conclusions after each meeting to each study site as prepared and distributed by the NRN DCC.

Note that the treatment in this trial will be conducted during the first few weeks of life for most study subjects, while the primary outcome of death or neurodevelopmental impairment will be assessed at 18-22 months of age. Thus, interim monitoring for this trial will be based primarily on monitoring adverse events (chiefly, death). For the first 120 randomized enrollees, the DCC will track rates of neonatal mortality and generate pre-specified tables after each 30 enrollees reach the Network Status (death, discharge, transfer or 120 days of age). Thereafter, if no trends of differences between groups were to develop, these comparisons will be done after every 60 enrollees reach Network status. The computed statistic at each of these safety looks will be compared to Pocock boundaries that are constructed beforehand so that an overall alpha level of 5% is maintained. If any trends of differences between groups were to develop, the DCC would notify the DSMC.

There will be annual DSMC meetings convened to monitor the progress of the trial and review the accruing safety data, starting one year after at least 11 sites have IRB clearance to start enrolling into the trial. Once follow up data start becoming available (which should occur from the 2nd year onwards) interim efficacy data will also be presented to the DSMC. To control for the inflation of Type I error associated with sequential testing, O'Brien-Fleming boundaries will be calculated. Since we cannot precisely predict beforehand what proportion of trial enrollees will have primary outcome data available before each of these meetings, we will use Lan DeMets spending functions to adjust the O'Brien-Fleming sequential monitoring bounds to account for unequally spaced interim analyses. In order to ensure the safety of participants, this monitoring regime ensures that adverse events are monitored more frequently than measures of efficacy, and the statistical bounds used to detect group differences in adverse event rates are more liberal than those used to determine efficacy. All data presented to the DSMC will be blinded to treatment group, though the report may be unblinded at the DSMC's request.

PUBLICATION / AUTHORSHIP PLAN. The primary manuscript will report data regarding the primary outcome and important secondary measures and will have a neonatal, a neonatal follow-up, and a surgical author per site as long as criteria for authorship are satisfied as expected. A secondary analysis focusing on the impact of the extent of disease on the duration of parenteral nutrition requirement is also planned with a similar authorship plan as above. Another secondary manuscript focusing on sepsis and infectious outcomes between the two treatment groups is also planned.

RISKS/BENEFITS. The trial involves no experimental procedures or assessments that impose risk above that experienced by study candidates under nonstudy conditions. The potential benefits include the increased attention to the care and the predischarge and post-discharge evaluation associated with conducting his trial.

REFERENCES

1. Hintz SR, Kendrick DE, Stoll MJ, et al. Neurodevelopmental and growth outcomes of extremely low birth-weight infants after necrotizing enterocolitis. Pediatrics 2005; 115:696-703.

2. Blakely ML, Lally KP, McDonald S, et al. Postoperative outcomes of extremely low birth weight infants with necrotizing enterocolitis or isolated intestinal perforation: A prospective cohort study by the NICHD Neonatal Research Network. Ann Surg 241:984-994, 2005.

3. Blakely ML, Tyson JE, Lally KP et al. Laparotomy vs. Peritoneal Drainage for Necrotizing Enterocolitis or Isolated Intestinal Drainage in ELBW Infants: Outcomes through 18 Months Age. Pediatrics 2006;117:680-687.

4.. Moss RL, Dimmitt RA, Barnhart DC, et al. Laparotomy versus Peritoneal Drainage for Necrotizing Enterocolitis and Perforation. NEJM 2006; 354:2225-2234.

5. Rees CM, Eaton S, Kiely EM, et al. Peritoneal Drainage of Laparotomy for Neonatal Bowel Perforation? A Randomized Controlled Trial. Ann Surg 248:44-51, 2008.

6. Flake AW. Necrotizing Enterocolitis in Preterm Infants – Is Laparotomy Necessary? NEJM 354:2275-2276.

7. Sinclair JC, Haughton DF, Bracken MB, Horbar JD, Soll RF. Cochrane neonatal systematic reviews: A survey of the evidence for neonatal therapies. Clin Perinatol 2003;30:285-304.

8. Silverman WA. Human experimentation in perinatology. Clin Perinatol 1987;14:403-16.

9. AAP Committee on Fetus and Newborn, and Canadian Pediatric Society, Fetus and Newborn Committee. Postnatal corticosteroids to treat or prevent chronic lung disease in preterm infants. Pediatrics 2002;109:330.

10. King M, Nazareth I, Lampe F, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. JAMA 2005;293:1089.

11. Liford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. BMJ 1995:311:1621.

12. Edwards S, et al. Why "underpowered trials" are not necessarily unethical. Lancet 1997;350:804.

13. Lagakos SW. Clinical trials and rare diseases. N Engl J Med 2003;348:2455.

14. Dunnett CW, Gent M. An alternative to the use of two-sided tests in clinical trials. Stat Med 1996;15:1729.

15. Tan S, Dear K, Bruzzi P, et al. Strategy for randomized clinical trials in rare cancers. BMJ 2003;327:47..

16. Rosenberger W, Feifang H. Maximizing power and minimizing treatment failures in clinical trials. Clin Trials 2004;1:141-7.

17. Wilkin B. Rare diseases and the assessment of intervention: what sorts of clinical trials can we use? J Inherit Metab Dis 2001;24:291.

18. Halpern SD. Prospective preference assessment: a method to enhance the ethics and efficiency of randomized controlled trials. Control Clin Trials 2002:23: 274.

19. Schulz K. Sample size calculations in randomized trials: mandatory and mystical. Lancet 2005;365:1348.

20. Tyson, J. Scientific evaluation of the growth of preterm or LBW infants fed new infant formulas. An expanded view. FDA White Paper, 2002. http://www.fda.gov/ohrms/dockets/ac/01/briefing/3903b1.htm

21. Sinclair JC. Weighing risks and benefits in treating the individual patient. Clin Perinatol 2003;30:251.

22. Adamina M, Guller U, Weber WP, Oertli D. Propensity scores and the surgeon. Brit J Surg 2006;93:389.

23. Braitman LE, Rosenbaurm PR. Rare outcomes, common treatments: Analytic strategies using propensity scores. Annal Intern Med 2002;137:693.

24. Rubin D. Estimating causal effects from large data sets using propensity scores. Annals Int med 1997;127:757.

25. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity scores models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med 2007;20:734.

26. Dimmitt RA, Meier AH, Skarsgard ED, et al., Salvage laparotomy for failure of peritoneal drainage in necrotizing enterocolitis in infants with extremely low birth-weight. J Pediatr Surg 2000; 35:856-859.

27. Goyal A, Manalang LR, Donnell SC, et al., Primary peritoneal drainage in necrotizing enterocolitis: an 18-year experience. Pediatr Surg Int 2006; 22:449-52

28. Sharma R, Tepas JJ III, Mollitt DL, et al., Surgical management of bowel perforations and outcome in very low-birth-weight infants (< or =1,200 g). J Pediatr Surg 2004; 39:190-194.

29. Hintz SR, Kendrick DE, Vohr BR, et al. Changes in neurodevelopmental outcomes at 18-22 months among infants of less than 25 weeks gestational age born in 1993-99. Pediatrics 2005; 115:1645.

30. Palisano, R, Rosenbaum P, Bartlett D, Livingston M. GMFCS E&R ©. http://www.canchild.ca

31. Anderson PJ, Kelly EA, Charlton M, et al. Cognitive Outcome at 2 Years in Very Preterm or Extremely Low Birthweight Infants Born in Victoria in 2005, PAS Meeting, May 4, 2009, E-PAS2009:4353.408.