Teamwork and Error in Neonatal Intensive Care

Eric J. Thomas, MD, MPH

Robert L. Helmreich, PhD

J. Bryan Sexton, PhD

Gwen Sherwood, RN, PhD, FAAN

Robert Lasky, PhD

Thomas Kallarackal

Sharon Crandell, MD

Jennipher Mulhollem

II. Structured Abstract

Purpose: To determine the relationships between teamwork and error during neonatal resuscitation.

Scope: Neonates are vulnerable patients and often the unfortunate victims of medical errors. We translated an aviation model of teamwork and error to the setting of neonatal resuscitation.

Methods: To better understand provider perspectives about working together, we conducted focus groups of NICU providers. We then used those results, plus survey data and preliminary viewing of videos of neonatal resuscitations, to define 10 behavioral markers of teamwork for neonatal resuscitation. We then analyzed videos of 132 resuscitations. Each tape was independently reviewed for "errors" (deviations from Neonatal Resuscitation Program [NRP] guidelines) and behavioral markers. We also directly observed 75 resuscitations and the admissions associated with those resuscitations (if admitted to the NICU).

Results: Providers reported three broad themes that affect the way they work together: Provider Characteristics, Workplace Factors, and Group Influences. We analyzed 132 videotaped resuscitations and recorded the frequency of 10 teamwork behaviors (behavioral markers). The tapes were independently analyzed for noncompliance with NRP guidelines. We also directly observed 75 additional resuscitations. We found an overall error rate during resuscitations of 17%. Independently observed errors and behaviors were correlated with each other: Inquiry, workload management, vigilance, and overall teamwork were negatively correlated with errors during the preparation and initial steps of resuscitations. In conclusion, we found that these behaviors can be systematically observed and quantifiably related to process-of-care measures. Planning for a clinical trial is underway.

Key words: error, teamwork, neonatal resuscitation, patient safety

III. Purpose

Medical errors and the adverse events they lead to are common and expensive. A recent Institute of Medicine (IOM) report suggested that healthcare organizations "promote effective team functioning" as one of its five principals for creating safe systems of care delivery. This emphasis on teamwork resonates with the experience of the aviation industry and healthcare providers and is supported by some research on measuring and improving teamwork during trauma resuscitations and in the operating room. However, fundamental questions remain about the relationships among teamwork and error.

Teamwork is a complicated phenomenon but, based upon Dr. Helmreich's aviation and medical research, we studied specific *error management behaviors* that comprise teamwork.

We hypothesized that error management behaviors were correlated with errors in delivering care to infants during their initial resuscitation and during the first minutes of care in the neonatal intensive care unit (NICU). To test these hypotheses we:

1) Adapted the aviation model of teamwork and error to NICU teams by conducting focus groups with NICU personnel; and

2) analyzed videotapes of a prospective cohort of preterm infants recorded during the initial resuscitation of infants, and we trained nurses to directly observe an additional set of resuscitations and the initial admission process of these infants.

3) We also used the data derived from aims 1 and 2, with AHRQ-funded investigator Lou Halamek at Stanford University, to design and plan an intervention to improve teamwork during neonatal resuscitation.

IV. Scope

Errors are common in healthcare, especially in environments like NICUs, where there are sick and vulnerable patients, high workload, and changing priorities and demands. Neonates are vulnerable to errors, and errors occur frequently in NICUs. Improved teamwork could play a role in preventing errors in NICUs.

Evidence for the importance of teamwork is ubiquitous, and researchers have documented some general relationships between teamwork and error. Expert groups, such as the IOM, NIH, and Joint Commission and American Association of Critical Care Nurses, believe that teamwork (or "collaboration") is important.

The existing research is limited by several biases, such as reliance upon self-report of error, small samples, and use of the same observers to measure both team behaviors and errors, thus rendering them susceptible to hindsight bias. There is little understanding of how specific team-related behaviors (especially those amenable to improvement) relate to specific errors. Methods to measure these team-related behaviors are rudimentary. Surprisingly, outside simulation, no one has evaluated teamwork and error in NICUs.

Our ultimate goal is to improve health outcomes and decrease costs. For this goal, the NICU is an ideal setting. Very-low-birthweight infants (VLBW; $\leq$ 1,500 grams) contribute substantially to three national burdens: infant mortality, childhood handicap, and medical expenditure. Preventing errors in NICUs presents a rich opportunity to improve health outcomes and decrease costs.

Our study was conducted in the Neonatal Intensive Care Unit (NICU) of Memorial Hermann Children's Hospital. The unit contained 80 intensive care beds and was staffed by approximately 50 RNs (of which 11 had additional training in neonatal resuscitation and transport); 13 faculty members who served as attending physicians; and other providers, including respiratory therapists, licensed vocational nurses, clinical pharmacists, and nutritionists. Fellows and residents also rotated through the unit each month. A small proportion of the daily

work in this unit is guided by protocols or clinical guidelines. The unit is usually conducting a project to improve safety and quality, and certain quality measures are routinely monitored by unit and hospital management. Videotaping and direct observation occurred in the delivery rooms.

Participants

       Focus groups. Prior to beginning the study, we decided to conduct one focus group for each of the major provider types in the unit plus two groups of mixed provider types. Therefore, we conducted seven groups that represented the primary provider types in this NICU: one group each of transport nurses (n = 3), residents (n = 6), attending physicians (n = 3), staff nurses (n = 7), and fellows (n = 6) and two mixed-provider focus groups (n = 4 and n = 7), for a total of 36 participants. Each of the primary provider types in the unit participated in the focus groups, but we did not attempt to make the sample size of each provider type quantitatively proportional to the number of that provider type working in the unit. Respiratory therapists, clinical nutritionists, and pharmacists are important team members in this unit, but their participation is focused on specific types of patients, primarily the sickest. We did not include them, in order to derive more generalizable information.

       Resuscitations. Participants in videotaped and directly observed resuscitations included the transport nurses, residents, fellows, and faculty from the NICU.

V. Methods

       The methods are divided into four sections: focus groups, the development of behavioral markers, the videos, and the direct observation.

       Focus groups. Each of the seven groups met once in a quiet conference room in the hospital. After obtaining informed consent, participants recorded their age, gender, education, certification, and work experience on a demographic form. To reduce investigator bias, all questions were open ended, and the experienced focus group facilitator (G.S.) had little knowledge of and no direct experience with the aviation teamwork model or the providers themselves. Two researchers developed a guided interview format for the study using a series of open-ended questions to elicit descriptions of how providers work together. The phrase "work together" is a less-specific phrase than teamwork, therefore allowing a broader range of responses by the participants. Interviews, lasting 60 to 90 minutes, were recorded and transcribed verbatim. A research assistant checked the accuracy of all transcripts. Field notes were also recorded. The study was approved by the Human Subjects Committee. Participants received $40.00 for their participation. Qualitative data analysis methods were used. Two investigators (G.S., J.M.) read and re-read the transcripts to comprehend the whole and highlight relevant data bits. Synthesis revealed three common themes emerging from the data. Further analysis identified descriptive elements for each theme. Themes and their descriptive elements were revised as analysis proceeded to accurately present the participants' perspectives. Investigators searched for and considered data to compare and contrast with the established themes to verify inclusion. Given that some groups were composed of just one provider type and other groups, several provider types, themes that arose from the different groups could be compared and contrasted in order to derive more valid and generalizable themes regarding perceptions of how providers worked together. The validity of the analysis was enhanced with independent reviews of the transcripts by other investigators (E.J.T., J.B.S., R.J.H.) that resulted in revision of the themes.

       Behavioral marker development. Although there are generalizable aspects of team behavior in managing safety-critical situations, we believe it is premature to assume that the same behaviors that manage threat and error in a cockpit will also manage error while resuscitating an infant. Therefore, we tried to translate the behaviors to this new setting instead of just directly applying them. Subsequent research will determine which markers are useful for preventing and managing errors.

In aviation, these teamwork-related behaviors are called "threat and error management behaviors," because they have been shown to help manage threats and errors in commercial aviation cockpits. In this paper, we use the term "behavioral markers" instead of "threat and error management behaviors." Behavioral markers are observable, nontechnical behaviors that contribute to performance within a work environment. We use this terminology because it is more generic and does not assume that these behaviors have been proven useful for managing threat and error during the resuscitation of newborns.

Our focus groups were used in the process of developing the behavioral markers. The full focus group results are presented below in the Results section of the report; in brief, providers in the NICU identified three broad themes that affect the way they work together: Providers, Workplace Factors, and Group Influences. The provider theme was related to provider personalities, reputations, and "egos." Workplace factors included issues such as staffing and the organization of care processes. Group influences included communication styles, relationships, and team functioning. Notably, teamwork was mentioned infrequently, and there were contrasting views about it. Many providers talked about teams as being organized along provider types (a nursing team) or care processes (resuscitation team). But others talked about teams as being more inclusive and functioning as a "family" and a "well-oiled machine." Focus group participants did not generate behavioral marker themes per se; however, numerous participants mentioned several LOSA behavioral markers. These included information sharing, teamwork, assertion, inquiry, workload distribution, leadership, and evaluation of plans (all defined below). These findings supported inclusion of these behavioral markers, but the focus groups also emphasized the unique complexity of this work environment and the numerous factors that influence how providers work together.

We also reviewed our previous surveys of these providers and other providers. Significant percentages of providers consistently reported that it is difficult to speak up; that conflicts are not appropriately resolved; that it is not easy to ask questions; that it is difficult to disagree "up the hierarchy"; and that staffing levels were not sufficient to handle the number of patients. These attitudes supported inclusion of behavioral markers, such as inquiry, assertion, leadership, and workload management.

Full results of the video recordings are presented below, but a small sample of tapes was reviewed to develop the behavioral markers. A team of MD and PhD researchers reviewed an initial group of five videos and determined that behaviors could indeed be identified on the videotapes. Subsequently, two investigators (E.J.T., J.B.S.) viewed 132 events together. While viewing each event, the investigators discussed the care process and looked for observable behaviors. During this process, we did not explicitly refer to LOSA error management behaviors, although both investigators were familiar with them. We then developed a list of candidate behaviors that was refined as additional events were observed. The initial list was derived from the focus group and survey results and from initial viewings of the videos. Through further review and refinement, the behavioral marker list was developed into a one-page (single-sided) data collection instrument. Notes were taken during the development process and organized to help define the behaviors and give examples. This process resulted in The University of Texas Behavioral Marker Audit Form (BMAF).

The UT BMAF for neonatal resuscitation is a one-page form with three sections (event demographics, behavioral markers, and threats to patient care). Two scales, observability and frequency, are used to rate each behavior. There is not a scale to rate the quality of the behavior, because we do not yet assume to know whether or how these behaviors prevent or manage errors. We used the form to record observations made while viewing the video recordings. The form is also being tested to record observations made during direct observation of an event in real time.

Observability Rating Scale

We could not always hear everything that providers were saying because of multiple providers talking at once or background noise from alarms or pagers. Therefore, we used the observability rating to allow us to indicate how well we could observe the behavior. If the behavior was absent, observability = 0. If the behavior was present, a rating of 1, 2, 3, or 4 was used to indicate how observable the behavior was. For example, poor audio or a simple inability to interpret verbalizations may lead a rater to choose an observability rating of 1 or 2 for a given behavior. If the behavior was explicit and easily observed, the reviewer may choose a 3 or 4 for observability. These ratings allow identification of clear examples of certain behaviors. A rating of 3 or 4 means that it would serve as an example of the behavior (not necessarily a "good" or "bad" example, just an observable example). The observability rating is a marked departure from LOSA, for which, after thousands of observations, a definitive list of behaviors exists. In this setting, we wanted to make sure that observers had the opportunity to note that a behavior may not be present or observable. The observability rating also allows observers to qualify the confidence of their observations. This may be done either because of technical problems (poor sound) or due to uncertainty regarding whether a specific behavior meets study definitions.

Frequency Rating Scale

Each behavior is also rated on a 1-4 scale based upon its frequency. Unlike LOSA, these ratings are not anchored at the upper or lower end by any reference to either the quality of the behavior or to safety or errors. For example, LOSA uses a rating of 1 to indicate poor performance that is a clear detriment to safety. In the setting of neonatal resuscitation, there is not empiric data to tell us what a "poor" behavior is. We do not know if a poor rating of one of these behaviors is correlated with decreased safety (increased errors). Therefore, instead of anchoring the scale as being detrimental (score of 1) or beneficial (score of 4) to safety, we rated the frequency of the behavior. If there were examples of a behavior carried out in an obviously "poor" or unhelpful manner, this was noted in the comments section or in the individual rating section.

Use of this frequency rating emphasizes that the UT BMAF for neonatal resuscitation is currently a screening instrument in early stages of development. More detailed analyses of the video recordings, including comparisons with independent ratings of the events for the presence of errors, will help us identify behaviors that can be called error management behaviors and subsequently taught to providers.

Behavioral Markers

The next section of the UT BMAF for neonatal resuscitation contains the 10 behavioral markers that we are currently rating. Below, we define each marker and discuss how it compares to LOSA markers.

1. Information sharing. This includes verbalization of information that relates to the assessment of the baby. For example, verbalization of heart rate, color, tone, vocal cord visualization, statements of opinion, advocating of views in noncritical moments, and other relevant observations or impressions about the baby's status. Discussions about Apgar scores are examples of information sharing. Information sharing may take the form of "baby talk." For example, "we like it when you cry," "you are nice and pink" (assessment of color), or "you are a strong little baby" (assessment of tone). Information sharing is usually observable; therefore, the observability rating is often a 2 or 3. Comparison to LOSA: There were several LOSA markers that appear to be either not applicable to this setting or were so infrequent that we collapsed them into a new marker called information sharing. Information sharing is not a LOSA marker, but it is a more general concept that includes LOSA behaviors, such as monitoring and cross checking, contingency management, briefings, automation management, and communication environment.

2. Inquiry. Providers ask questions of each other. For example, "What's the heart rate?" "What do you think of his tone?" "Should we intubate her?" Like information sharing, inquiry is usually

observable (2 or 3). Statements may also serve as questions, depending upon the inflection of the voice. Comparison to LOSA: This marker exists in LOSA, and its definition was not modified.

3. Assertion. An individual provider asserts their opinion (through questions or statements of opinion) during *critical times*. Assertion *does not* include routine statements or questions about a baby's heart rate, tone, color, and respirations (these are part of information sharing or inquiry). The observability rating for assertion is often 0, meaning that the majority of the events we have observed do not contain examples of assertion. Assertion may be indirect, through the use of repeated questions (e.g, "Do you really think the endotracheal tube is in the esophagus?"). Assertion can overlap with other behaviors. For example, the previous quote is also an example of inquiry. An example of assertion and workload management is "Is that your phone ringing? Let me help." Comparison to LOSA: This marker exists in LOSA, and its definition was not modified.

4. Intentions shared. A provider states their intentions before deviating from the norm. Statements about following routine Neonatal Resuscitation Program guidelines are not examples of intentions shared. Thus defined, this behavior may be infrequently observed (observability rating often may be 0). Comparison to LOSA: The LOSA marker is called "plans stated." We changed to it to "intentions shared," because it more accurately described what we saw. "Plans stated" implies explicit statement of an explicit plan. In neonatal resuscitation, we saw much more sharing of general intentions/next steps but not explicit statements of multi-step plans.

5. Teaching. Teaching is observed during the observation. This may be short and informal information exchanges. It may occur between any of the providers (e.g., nurses can teach residents and vice versa). Teaching is often not seen (observability = 0). Comparison to LOSA: There is not a comparable LOSA marker, because rates of on-the-job training in commercial aviation are far lower than in medicine, especially in hospitals with residency programs and those affiliated with medical schools.

6. Evaluation of plans. An explicit discussion about the status of the baby and the decisions made to get to the current situation. This is often not observed, probably due to the standardization of neonatal resuscitation. Comparison to LOSA: This marker exists in LOSA, and its definition was not modified.

7. Workload management. The workload is distributed among those present at the resuscitation. Tasks are appropriately prioritized. This behavior is usually observed but not heard on tape. Therefore, observability is usually a 2, 3, or 4 (with 4 reserved for examples in which there are explicit verbal decisions made about workload management).
Comparison with LOSA: This marker exists in LOSA and its definition was not modified.

8. Vigilance/environmental awareness. Providers remain alert and focused on the resuscitation. This is usually observable based upon nonverbal behaviors (rated 2-4), and 4 is reserved for some examples in which vigilance is explicitly addressed. The frequency rating is usually a 3 or 4. It is 1 or 2 for clear examples of loss of focus on the task, often due to interruptions such as social conversations, phone calls, twin births, and parental interruptions. These interruptions are recorded separately below. Comparison with LOSA: The same marker exists in LOSA.

9. Teamwork overall. This is a global assessment based upon the ratings of behaviors 1-8 above. It may also include other observations, such as dynamic "give and take" among team members and nonverbal communication that are not explicitly defined in the behavioral markers. Comparison with LOSA: In LOSA the comparable marker is "overall crew effectiveness rating." Because the goal of the current study was exploratory and descriptive, we were not able to judge effectiveness, so the LOSA term was not appropriate.

10. Leadership. This may be rated for any provider at the resuscitation. Leadership activities may include sharing of a mental model, assigning tasks, or sharing of information and opinion. There is usually not a clear leader (either in deed or word), so the observability rating is often 0. Lack of leadership was not obviously detrimental to the process of care. This may be related to the standardized nature of these resuscitations and because most babies were not very ill.

Leadership is often attributed to one person, but two people may exhibit leadership; if so, the rating would represent the combined leadership activities. Comparison to LOSA: The LOSA marker focuses on leadership exhibited by the captain. But the neonatal resuscitations we observed were notable for not having one explicitly identified leader. Moreover, leadership roles were fluid and highly dependent upon team composition and experience level.

After the 10 behavioral markers, the form allows reviewers to rate individuals who differed from the team, to rate threats, and to make additional comments.

Individual ratings. Was there an individual who differed in one of the above markers in a manner that set them apart from the rest of the providers? Note that one person may exhibit leadership, but that is rated under leadership and not as an individual rating.

Comparison to LOSA: LOSA provides an opportunity to rate each member of the crew on technical proficiency and a composite assessment of CRM skills. In the past, LOSA collected individual ratings if an individual differed substantially from the rest of the crew, but current LOSA methodology has removed this step.

Threats. These may include a) under- or overstaffing; b) environmental distractions such as phone calls, twin birth, social conversations, alarms, parental interruptions, etc; and c) other (including the complexity of the medical situation, illness of infant, twins, congenital malformations, etc). Each threat is rated for observability. It is also rated (yes/no) for whether it was managed. The threat or problem can be managed without being solved. For example, the providers may believe they are understaffed and an attending is needed. The attending may never come, but the threat (understaffing) could still be managed. Recognizing a threat as a threat (e.g., telling a caller you will call back when done with the resuscitation) can qualify as managing the threat. Threat severity is used to indicate the degree to which the threat interfered with the resuscitation. Although we use the term "threat," it is not yet clear that these events can be empirically demonstrated to "threaten" the infant.

Other comments. Space for any observations not already included on the data form.

Candidate for additional review. Events with good examples of behaviors or complicated events (e.g., babies that require intubation) with a lot of interaction among team members are candidates for additional review. Observers then write in the comment section why it is a candidate for additional review. In this study we identified a case as candidate for additional review if 1) the infant was intubated; 2) one or more markers were rated as 4 for observability; 3) the infant was born with meconium; and 4) at the discretion of the reviewer.

Videos. We video recorded 132 resuscitations of infants born by cesarean section between January and December 2002 and used the Behavioral Markers Audit Form to rate observability and frequency of 10 behaviors and some "threats" to patient care. Approximately half of the videos were reviewed independently by Sexton and Thomas; then, they met for consensus. The other half were reviewed independently with no consensus. Of these independent ratings, the weighted kappas ranged from 0.4 (fair agreement) to 0.8 (very good agreement) except for two markers (workload management kappa = 0.2 and vigilance kappa = 0.3).

"Error" was defined as noncompliance with Neonatal Resuscitation Program (NRP) guidelines. Videos were reviewed independently by two NRP instructors for compliance with NRP guidelines. The reviewers then met for consensus. We then grouped errors by stage of NRP process (Preparation and Initial Steps [15 items]; Oxygen Administration [3 items]; Bag/Mask Ventilation [7 items]; Chest Compressions [5 items]; Intubation [14 items]; Drugs) and by error type (Assessment, Action, Timing, Re-evaluation, and Communication).

We report frequencies (%) of patient characteristics, behavioral markers, and errors. Spearman rho was used to measure correlation among behavioral markers and error.

Direct Observation. Three nurses from the university's Clinical Research Center were trained to use the behavioral marker form and to record five process-of-care measures.

Time until surfactant and number of intubations were recorded for the resuscitations in the delivery room. Time until IV placement, time until x-ray, and time until orders written were recorded for the infants who were admitted to the NICU. These are all measures of the quality of care in this setting. It was not feasible for the nurses who were observing resuscitations in real time to determine compliance with NRP guidelines. The observers also attempted to observe unexpected codes of infants in the NICU. The nurses carried the code beeper and attended as many events as possible between October 1, 2003, and April 16, 2004.

We report frequencies (%) of behavioral markers and time until event for our process-of-care measures; Spearman rho was used to measure correlation among behavioral markers and time until event measures of the process of care.

VI. Results

Focus groups

Demographic characteristics are shown in the table below. The average age of the participants was 34.4 years, with a range of 25-57 years. There were nine male participants and 27 female. Participants included seven African Americans, 16 Caucasians, four Hispanics, six Asians, and three from other ethnic groups. Years of experience ranged from 3 months to 26 years, with an average of 8 years. Twenty-three held medical degrees (residents, fellows, attending physicians), nine were registered nurses, two were LVNs, and two were unlicensed patient care assistants (PCAs). The nurses and PCAs were full-time employees of the hospital.

	RNs N = 9	LVNs N = 2	Unlicensed N= 2	Physicians N = 23	Total N = 36
Male	0	0	0	10	10
Female	9	2	2	13	26
Mean years experience	9.3	18.5	4	6.8	9.7
Mean age	34.4	45	26.5	33.6	34.9
Ethnicity:					
AA	3	0	2	2	7
Caucasian	6	1	0	9	16
Hispanic	0	1	0	2	3
Asian	0	0	0	7	7
Other	0	0	0	3	3

Provider responses to questions about working together centered around three major themes: 1) Provider Characteristics; 2) Workplace Factors; and 3) Group Influences. To elaborate the themes, each was defined by several descriptive elements and participant quotes.

Theme 1. Provider Characteristics

Provider Characteristics included the descriptive elements of personal attributes, reputation, and expertise. *Personal Attributes* The personal attributes that providers mentioned included power, competitiveness, collegiality, values, beliefs, gender, confidence, vulnerability, flexibility, and attitudes about their work and other providers. One nurse summarized the importance of personal attributes: "If the nurse doesn't like the physician for whatever reason, even if it's on a personal level, well that can affect the nursing." However, some providers overcame differences in personal attributes. A nurse reported that, "in spite of personality differences, people adopt roles, know where they fit in to help best, and come together to get the work done."

Reputation, composed of previous experiences together, perception of the other, trust, and respect, was the second descriptive element of the Provider Characteristics theme. Reputation often preceded a new provider's arrival and was purportedly difficult to change. A nurse said, "[if] you make a decision based on something that you knew from a previous experience that is not what is usually done in this institution, 50 people will know within 30 minutes." When asked how that influences working together, the nurse responded "It's very influential, because some people really care about what everybody is saying. And for other people, they don't really care." Regarding respect and trust, another said, "The degree of confidence in coworkers is based on prior interactions and often a quick assessment of the provider's assertiveness that would be needed in a particular circumstance." A provider's reputation was derived from prior secondhand information as well as judgments in particular situations.

Expertise was the third descriptive element in Provider Characteristics and included a colleague's knowledge, skills, certifications, and education. Expertise went beyond seniority (e.g., "the best expert may not be the most senior"), but seniority plays a critical role (e.g., a fellow said that, as "a back-up plan if someone can't perform a skill, you call the next person in the hierarchy."). A transport nurse described how the interaction of reputation, expertise, and trust affected working together after an attending physician observed her successfully working with a fragile infant during a transfer: "Whereas before I was kind of shaky on our relationship [needing to prove myself], now it's gone." Positive clinical experiences among providers from different disciplines could demonstrate expertise and build respect.

Theme 2. Workplace Factors

Workplace Factors, the second theme, had three descriptive elements: staffing, work organization, and work environment. *Staffing* levels and provider mix were reported as being highly variable. This affected staff workload and their ability to complete work in a timely manner. The dynamic mix of permanent and transient providers caused by rotation of physicians and use of agency personnel created an often complicated and unpredictable work environment. A resident stated, "There's a different set of residents every month, … others, like staff nurses, RTs [respiratory therapists], and the ancillary staff, are there all the time, so I think there's a lot of trust that's built up between the different people that you're working with. It's so fast and it goes by so fast, by the end of the month when you finally learn the patient and you're learning how things work, you leave." Staffing impacted both the fluidity of group membership from day to day as well as continuity of care (e.g., a physician said, "it is best when the same nurse cares for the patient"). The number of healthcare providers present also affected the workplace and how work was done; having too many was potentially as complex and detrimental as having too few. With too many, "there may be no prioritization, too many orders being shouted," but they also reported "it is hard to say no to extra help." Consensus views were that appropriate levels of staffing involved a delicate balance. Seniority, whether granted from position title or from experience, played a role. Junior staff members (particularly residents) felt challenged when confronting new situations without a more senior provider to assist and guide. In summary, descriptions of staffing pertained to the number of providers, their qualifications, their familiarity with one another, and the ability to efficiently and effectively complete tasks.

Work Organization was defined by the nature and complexity of the work and also the roles, responsibilities, and accountabilities of staff. The NICU environment was described as "stressful" and "complex with many influences." Patient care protocols specified roles and responsibilities for some situations, such as resuscitations. In other situations, practices among physicians differed, sometimes substantially. One physician stated, "I mean, everyone has their own way of practicing medicine." Emergencies were reported as challenging, "attracting too many people, which creates a chaotic environment. The senior person there takes command, which means some assess the situation, assign staff, and make sure that things are being done, not just one more person yelling one more set of orders." Although care protocols assigned "predefined responsibility, recognition of the role you are playing, and its seniority," behavior

varied among providers based on "assertiveness, competence, and reputation." Position titles defined the chain of command so that "the senior person takes charge." Checks and balances among the providers was reportedly "a two-edged sword, serving to ensure accuracy" and producing conflict when perceived as others "checking up [on me]."

Workplace Environment was defined as the unit atmosphere and provider responses to the work place. Staff nurses described the unit as functioning best when there was a climate of collaboration, "togetherness, like a sense of community, though we are all different, pride in our job is a common denominator. When we don't have community, we get polarization, we don't get to bring things to the table." As trust builds in a healthy unit, providers "lean on each other for different observations that others may notice." The importance is signified by the desire for "a good environment to think in a better way. NICU is a little overwhelming at first with all the monitors, and a lot of things going on with the patient…and the workload and the hours…until you establish rapport with the people working there."

Theme 3. Group Influences

Three descriptive elements defined Group Influences: communication, relationships, and teams. *Communication* was defined by both verbal and written information sharing. Further descriptors included communication skill and style, accountability to others to share accurate information, verification of information and observations, questioning, and hierarchy. A physician illustrated the intricately critical nature of communication: "Balancing questions at all levels and sharing information is critical to what physicians do," especially given that "you must make decisions based on observations and assessments made by others" and that miscommunication is noted as a serious problem. Accountability for information sharing "gets to be a vital issue in patient care. There are people on the team who withhold information from the person who is supposed to be in charge of care, but they are waiting for the person who they feel like really knows what is going on, because they don't think the person in charge is adding to the patient's care." Still, physicians explained that "information helps save lives."

Hierarchy also influenced communication and was "embedded in title, not expertise." A physician said that "even attendings are careful questioning another physician. I counsel junior levels how to address another physician appropriately, especially if they are from outside the facility." A nurse said, "questions can only go down the chain of command, and you are very careful about questions going up the chain of command." However, another nurse stated, "Questions done in a timely and respectful manner can be beneficial," and there was a physician who "usually welcomes questions, because I feel like it could be something that I'm missing." Participants agreed that the way questions were asked and information was shared was critical to effective communication.

Communication within a hierarchy may skip the order in the chain of command, sometimes leading to conflict: "We are in a system with few senior people with the majority of the care delivered by an army of junior people at all levels…with a duality that you have a nurse here, a physician that should be above the nurse, yet the physician is so inexperienced that the nurse knows more of what to do and that causes a lot of friction…as opposed to my asking is there any value in what you are saying." A transport nurse said, "I don't think because you have the authority means you have the answer."

Relationships were influenced by interactions among providers, types of providers, and time working together. Interactions among providers depended upon the types of providers present and were influenced by behaviors such as being dominant, authoritative, or unilateral and uncompromising. These behaviors affected opportunities for others to engage in care activities and provider communication. One nurse described it as "the ability to know depends on being there day after day to have good a communication system with other nurses and physicians." Sorting through relationships often left some providers feeling caught in the middle: "the residents come down on [the nurses] and say this isn't right and has to be this way, then they get it from their attending and they are in the middle and nobody knows where to go."

Teams comprised the third descriptive element for the Group Influences theme. One resident noted that "*team* referred to a group of people who work together to care for patients in that environment." Most participants described teams according to disciplines and the work ascribed to each role. Teams were also described through processes of care, such as resuscitation, admissions, and deliveries. Broader definitions of teams were also provided. An attending physician said that teamwork is a "complex set of interactions and activities with defined roles to accomplish a task, so that the total of the collective efforts surpasses the success achieved by an individual." Nurses described the role of team members as "helping each other be the best we can be." Shared responsibility and accountability was another factor describing teams, as stated by one attending: "there may be some things about that patient's care that could be jeopardized because you are making the decision to not involve everybody on the team. Even though you may feel like this person is a weaker link, we are all still part of the team." How teams worked together depended on the personalities, attitudes, knowledge, and skill of the people involved. The team role was built from working together and getting to know one another based on skill and expertise.

Teams were considered necessary for patient care coordination: "Collective minds offer better care, when decisions are made together" (but one provider observed how rarely the entire team made patient care rounds together). Others used metaphor to capture the definition of team, including "a well-oiled machine," a "sports team needing a coach," "a lot of good musicians but no director," or a "family where members come and go." A physician summarized it as "once you have established a system within your team, and there is consistency every day, then it tends to run smoothly, and if there is a role that nobody plays and agrees on and assigns some responsibility and the respect for that person's opinion, then it tends to run. But when there is one member or one person in the system not working, then that teamwork does not happen."

Results of Video Tape Analyses.

The infants captured on video were relatively healthy, and their characteristics are listed in the table below. The sample from the tapes is similar to all infants born by cesarean section who require resuscitation in this hospital.

Race:
Black	46%
White	31%
Hispanic	16%
Asian	7%

Gender:
Male	47%
Female	53%

|	
Mean gestational age	36.6
Median gestational age	38 (range 24-42)
Mean birthweight	2789
Median birthweight	2967 (range 657-4445)
1-min Apgar (mean/median)	7.6 / 8
5-min Apgar	8.6 / 9

Regarding the process of neonatal resuscitation, 132 received preparation and initial steps; 117 had oxygen administered; 20 had bag/mask ventilation; 11 were intubated; two received chest compressions; and one received epinephrine. Of all relevant steps, 17.5% were done incorrectly or not done at all. The error rates for each section of resuscitation and by type of error are shown below.

NRP Section	%
Prep/initial steps (15 items)	15.9
Oxygen administered (3 items)	30.3
Bag/mask ventilation (7 items)	25.5
1st intubation (14 items)	24.2
2nd intubation (6 items)	54.5

NRP "Error" Types	%
Assessment	<1
Action	11.7
Timing	16.9
Re-evaluation	74.7
Communication	64.5

The following figures illustrate the frequency of the behavioral markers. They show that Information Sharing, Inquiry, Vigilance, and Workload Management were always observable. Assertion and Teaching were observed in about 20% of cases. Evaluation of Plans was observed in 9% of cases. Leadership and Intentions Stated were observed in 5% of cases.

Frequency of Behavioral Markers

Frequency of Behavioral Markers
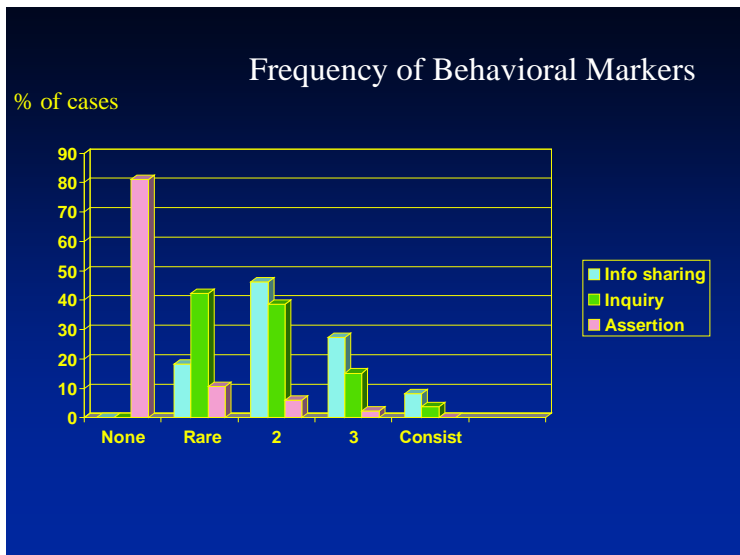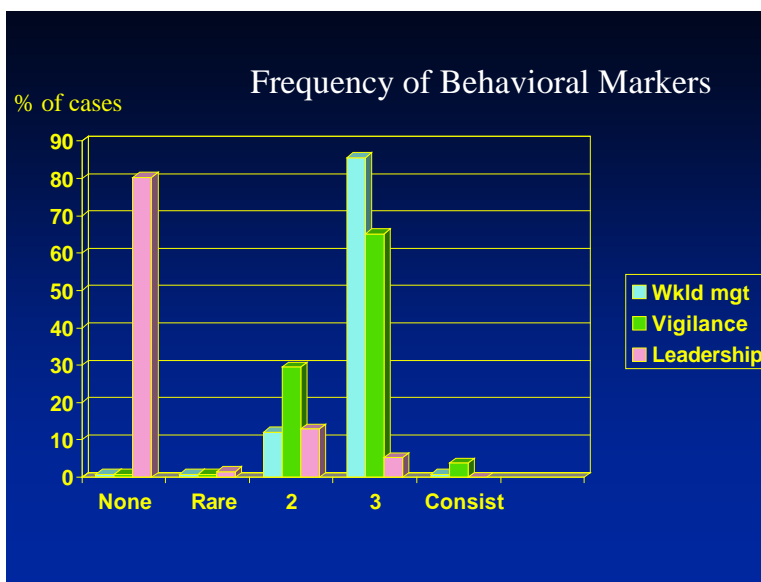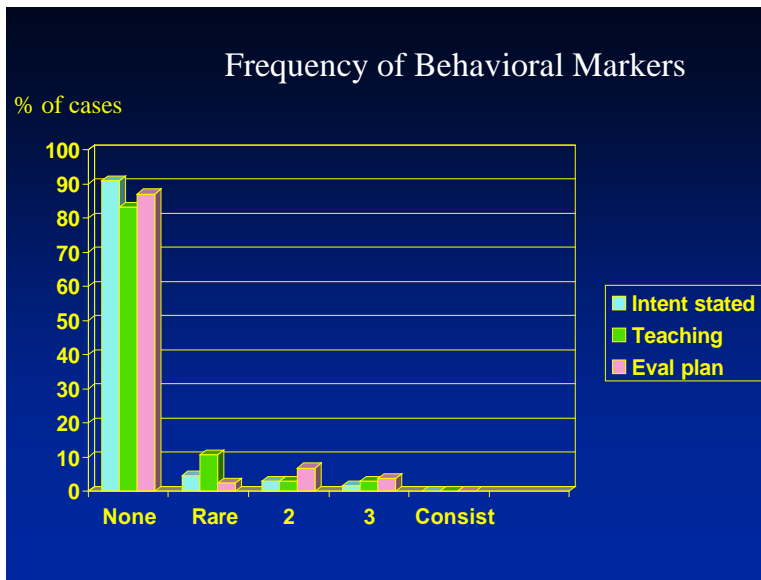
Of the 132 infants who received the preparation and initial steps section of NRP, we found significant correlations (P < 0.01) between the overall error rate and inquiry (-.265), workload management (-.228), teamwork overall (-.265), and a summary behavioral marker variable (-.293). Vigilance was also correlated, with a correlation coefficient of -.217 and P < 0.05. Although these correlations are relatively small in absolute value, the findings suggest that teams who exhibit more frequent inquiry, workload management, vigilance, and overall teamwork commit fewer errors during the initial steps of neonatal resuscitation. For the NRP sections of oxygen administration, bag/mask ventilation, and chest compression, we found no significant correlations with behavioral markers. Of the infants who had two intubation attempts, information sharing was strongly correlated with errors during the second attempt (-.853, P < 0.05). This suggests that more information sharing was correlated with fewer errors.

We observed 75 resuscitations. Of these, 14 were admitted to the NICU and observed. We observed two unexpected codes in the NICU. The frequency of the behavioral markers during the resuscitations is shown below.
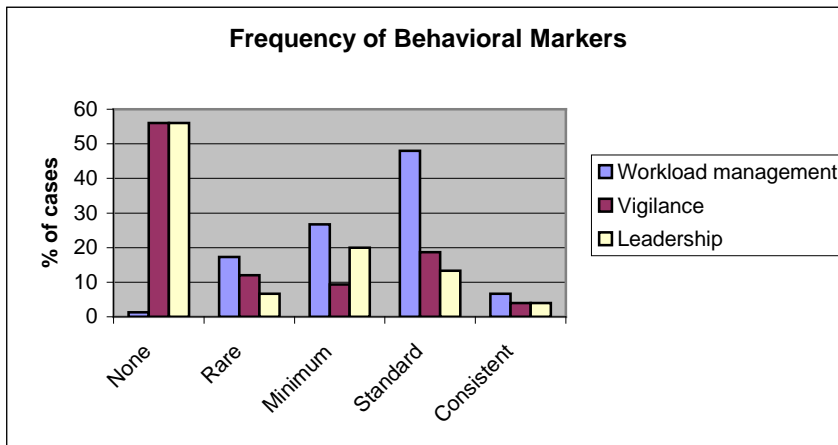
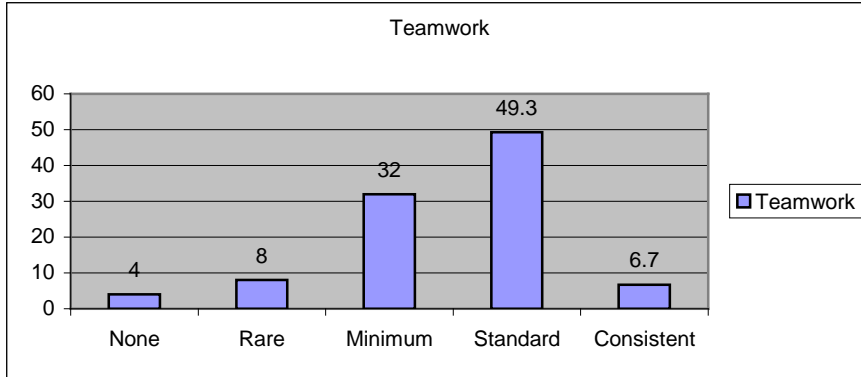## Resuscitation: Frequency of information sharing, inquiry, and assertion
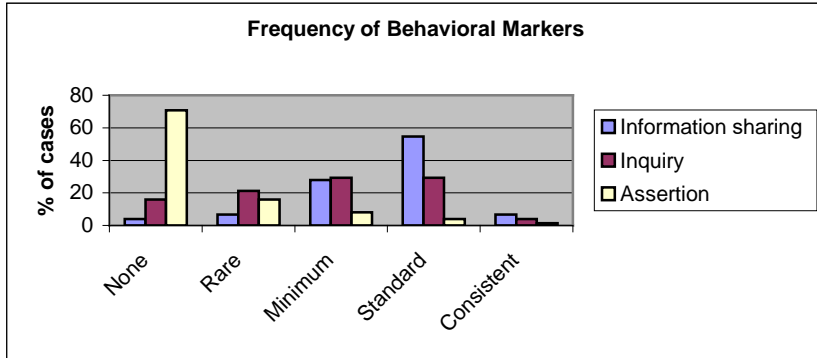
Resuscitation: Frequency of intentions stated, teaching, and evaluation of plans
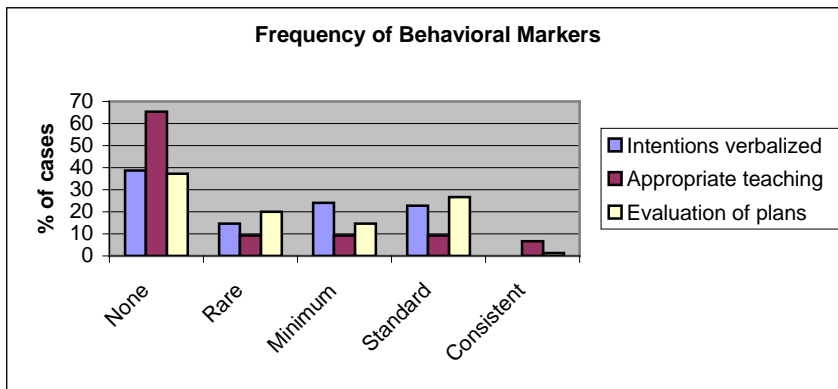
Resuscitation: Frequency of workload management, vigilance, and leadership

**Resuscitation: Frequency of overall assessment of teamwork**

The mean time until surfactant was administered for the eight infants who required surfactant was 8.4 minutes ($\pm$ 4.12). Seven infants required intubation, and the mean number of attempts was 1.75 ($\pm$ 0.89), with a range of 1-3. As shown below, there were no significant correlations among behavioral markers and time until surfactant or number of intubations.

	BM1	BM2	BM3	BM4	BM5	BM6	BM7	BM8	BM9	BM10
Surfactant Time[a]	.24	.48	-.47	-.42	-.21	-.27	-.40	-.37	-.22	.03
Intubation Attempts[b]	.24	.59	.21	.19	.28	.19	.27	.41	.24	.37

Note: All correlations are Spearman rho; none were significant at $P < 0.05$.
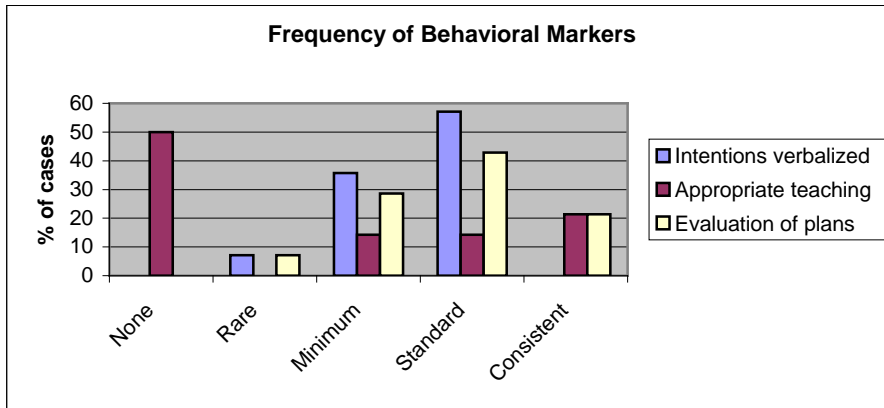
The frequency of the behavioral markers for the 14 admissions is shown in the following figures.

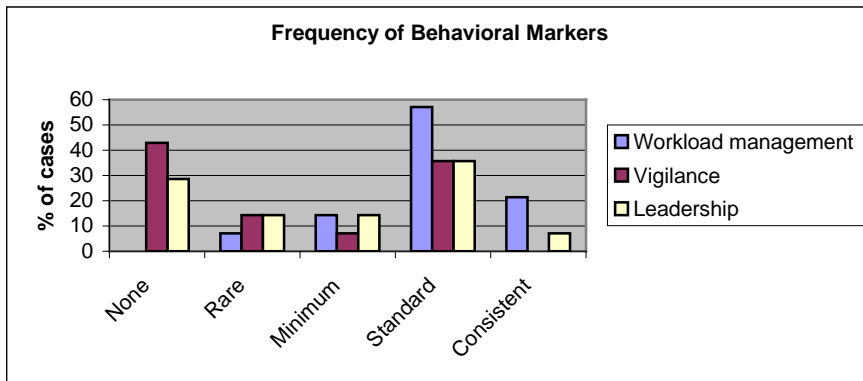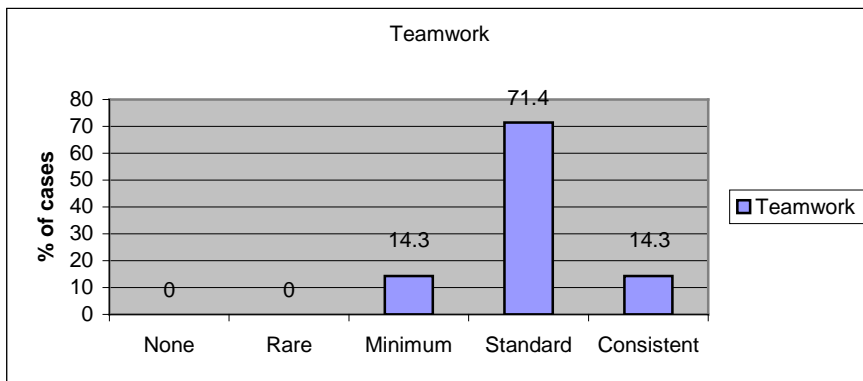**Admissions: Frequency of information sharing, inquiry, assertion**
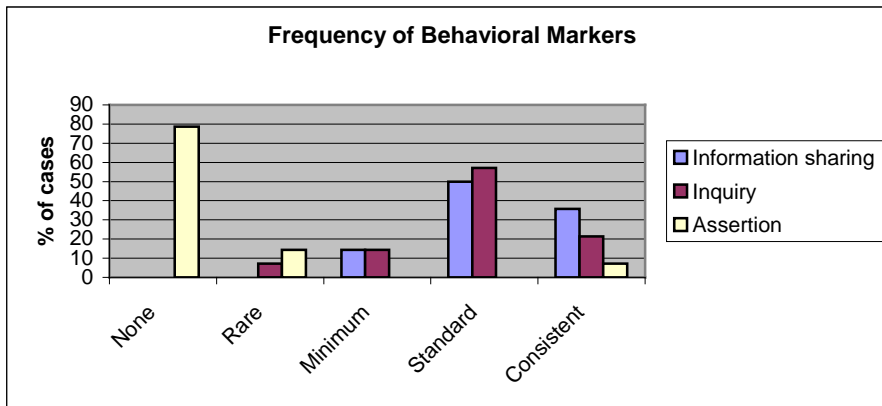
Admissions: Frequency of intentions stated, teaching, evaluation plans

Frequency of Behavioral Markers

Admissions: Frequency of workload management, vigilance, and leadership

Frequency of Behavioral Markers

Admissions: Frequency of overall teamwork

Teamwork

The mean time to IV placement was 27.6 minutes $\pm$ 12.5. The mean time to x-ray was 28.1 minutes $\pm$ 16.7. The mean time until orders were written was 9.9 minutes $\pm$ 8.9. Correlations among behavioral markers and these process-of-care criteria are shown below. These reveal that teams with more frequent assertions (BM3) had shorter times to IV placement. Teams with more frequent intentions shared and evaluation of plans (BM6) had shorter times to orders written.

Note: * indicates significant at P < 0.05; *a* refers to n = 11; *b* refers to n = 7

	BM1	BM2	BM3	BM4	BM5	BM6	BM7	BM8	BM9	BM10
IV Time*a*	-.04	.08	-.56 *	-.48	.09	.35	.12	-.23	.33	-.06
X-Ray Time*b*	-.07	.36	.12	.41	-.16	.29	.00	.29	.00	.26
Orders Written Time*a*	-.41	-.22	.22	-.63 *	-.45	-.55 *	.14	.42	.28	-.49

VII. Conclusions
Focus groups

Providers in a NICU reported three broad themes that affect the way they work together: Provider Characteristics, Workplace Factors, and Group Influences. As expected in any work setting, our participants said that the personal attributes of the providers involved have a major influence on how they work together. These relatively unchangeable characteristics, such as personality, reputation, and expertise, were noted as important. Workplace factors influencing the way providers worked together included staffing, organization of work, and environment. Staffing and organization of work were both dynamic and often unpredictable factors influenced by numerous characteristics of the providers, the organization, and the patients.

Group influences, such as group communication and relationships, were noted by participants (especially physicians), with emphases on how much they rely on others for information. Hierarchy within groups had a powerful but complicated influence on the way providers communicated with each other. As described in healthcare and other settings (10) front-line personnel often found it hard to question those with authority. In the current study, some participants described the process of skipping immediate superiors to go above their head to someone with more knowledge, experience, or rapport. Other participants reported being comfortable asking or receiving questions from anyone.

Within the broader concept of groups, we found that providers occasionally, and without prompting, talked about teams. Teams were often described according to a process of care (the transport team) or discipline (nursing team). Providers noted that rounds in the NICU did not consistently include all team members even though this inconsistency hampered communication and work coordination. These descriptions of team processes in this NICU were in contrast to other subjects' more inclusive descriptions of teamwork that included phrases like a "collective mind" or a "well-oiled machine."

From these focus group findings, we conclude that providers working in a large NICU noted that Provider Characteristics, Workplace Factors, and Group Influences affect the way they work together. Although providers mentioned specific team-related behaviors found to be important in aviation teams, they expressed different definitions and perceptions of teamwork. Future efforts to measure and improve how providers work together should note that behavior may be affected by multiple issues external to the immediate work environment. Regarding day-to-day practice in NICUs, medical directors, nurse managers, and administrators should be cautious about using the words like team and teamwork, and they should consider a broad set of organizational factors when evaluating and trying to improve the way providers work together.

<u>Videotapes and observations.</u>

We translated a set of 10 behavioral markers from aviation to neonatal resuscitation. Many of these, such as information sharing, inquiry, workload management, and vigilance, were commonly observed; others (leadership, intentions stated) were very rare. In this standardized process of care, lack of a behavior may be due to nature of the process, not a problem with the team. For example, resuscitation is very protocol driven, so there is less need to state intentions and less need for active leadership.

We found that independently observed errors and behaviors were correlated with each other. Inquiry, workload management, vigilance, and overall teamwork were negatively correlated with errors during the preparation and initial steps of resuscitations. This raises the possibility that these behaviors were being used by providers to prevent and manage errors. Our observational study does not allow statements about cause and effect. It is likely that other significant correlations would have been detected in later stages of resuscitation, but our sample size became smaller as resuscitations progressed (e.g., only 11 patients required intubation). Interestingly, the magnitude of the correlation coefficients was much greater during intubations (.7-.9 instead of .1-.2), making it possible that the link between these behaviors and the quality of care is even stronger during these critical periods of patient care.

We found that these behaviors can be systematically observed and empirically related to process-of-care measures. Our data support plans to study the effect of training providers to use these behaviors and to understand the impact on quality and outcomes. We have now identified behaviors that are important in this environment, developed a method to measure them, and demonstrated some correlations among these behaviors and errors in the process of care.

The aims of a randomized trial to evaluation team training would be to 1) develop a new NRP curriculum that includes CRM skills (NRP-plus) and 2) randomize providers to receive either NRP or NRP-plus and then test the effectiveness of NRP-plus by observing team performance in a simulator. Primary outcome measures will include errors and teamwork. Our hypothesis is that teams randomized to NRP-plus will commit fewer errors (deviations from NRP guidelines), will exhibit more teamwork behaviors, and will better manage unexpected complications compared with teams composed of NRP-trained providers.

During the last few months, Dr. Thomas has been communicating with Lou Halamek at Stanford. Dr. Halamek is an AHRQ-funded patient safety investigator who has developed a neonatal simulator and a neonatal resuscitation training course that uses CRM skills and teamwork behaviors similar to those developed for this grant. In May 2004, Dr. Halamek invited Dr. Thomas to Stanford for a 1-day meeting and presentation. During that meeting, this type of randomized trial was discussed in detail. The setting would be the Stanford Neonatal Simulator. The subjects would include nurses from Bay-area institutions who need NRP recertification. The intervention would be NRP-plus based upon the curriculum already developed at Stanford with some modification to teach specific behavioral markers found to be important in this study. The teams would go through scenarios in the simulator and be videotaped. We hypothesize that provider groups who receive NRP-plus training will a) commit fewer errors during simulated resuscitations than provider groups who received standard NRP; b) use CRM-related behaviors more often then than provider groups who received standard NRP; and c) better manage unexpected complications.

The current study allows sample size calculations to be performed. The overall error rate was 17.5%. If we randomized 42 teams to NRP and 42 to NRP-plus, we would have 80% power to detect a 4.4% absolute difference (25% reduction) in the overall error frequency (e.g., 17.5% in NRP and 13.1% in NRP-plus). Our study also found that 36% of teams had no behavioral markers with a frequency greater than 2 (isolated or minimum examples). To detect a 50% increase in the number of teams with at least one behavioral marker with a frequency of 3 or greater, we would need 103 teams in each group.

Detailed planning of this study among Dr. Halamek, Dr. Thomas, and their groups is underway.

List of publications and products

1. Thomas EJ, Sherwood GD, Mulhollem JL, Sexton JB, Helmreich RL. Working together in the neonatal intensive care unit: provider perspectives. Journal of Perinatology, in press.

2. Eric J. Thomas, MD, MPH, J. Bryan Sexton, PhD, Robert L. Helmreich, PhD
Translating Teamwork Behaviors from Aviation to Healthcare: Development of The University of Texas Behavioral Markers for Neonatal Resuscitation. Under review at Quality and Safety in Healthcare. Can be accessed at www.UTPatientSafety.org
This paper includes the dataform we used to observe the behaviors. We consider this a product of the research. Other investigators have requested the documented and downloaded from our website.

3. A manuscript is under preparation to report the relationships among teamwork and error during neonatal resuscitation.

4. We are planning a clinical trial of team training in neonatal resuscitation.