Gut Online First, published on April 5, 2017 as 10.1136/gutjnl-2016-312819

ORIGINAL ARTICLE

Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants

Chloe A Autran,[1] Benjamin P Kellman,[1,2] Jae H Kim,[1] Elizabeth Asztalos,[3] Arlin B Blood,[4] Erin C Hamilton Spence,[5] Aloka L Patel,[6] Jiayi Hou,[7] Nathan E Lewis,[1,2,8] Lars Bode[1]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2016-312819).

For numbered affiliations see end of article.

Correspondence to
Dr Lars Bode, Department of Pediatrics, University of California San Diego School of Medicine, 9500 Gilman Dr., Mail code 0715, San Diego, CA 92093, USA; lbode@ucsd.edu

CAA and BPK contributed equally.

Received 7 August 2016
Revised 7 March 2017
Accepted 8 March 2017

ABSTRACT

Objective Necrotising enterocolitis (NEC) is one of the most common and often fatal intestinal disorders in preterm infants. Markers to identify at-risk infants as well as therapies to prevent and treat NEC are limited and urgently needed. NEC incidence is significantly lower in breast-fed compared with formula-fed infants. Infant formula lacks human milk oligosaccharides (HMO), such as disialyllacto-N-tetraose (DSLNT), which prevents NEC in neonatal rats. However, it is unknown if DSLNT also protects human preterm infants.

Design We conducted a multicentre clinical cohort study and recruited 200 mothers and their very low birthweight infants that were predominantly human milk-fed. We analysed HMO composition in breast milk fed to infants over the first 28 days post partum, matched each NEC case with five controls and used logistic regression and generalised estimating equation to test the hypothesis that infants who develop NEC receive milk with less DSLNT than infants who do not develop NEC.

Results Eight infants in the cohort developed NEC (Bell stage 2 or 3). DSLNT concentrations were significantly lower in almost all milk samples in NEC cases compared with controls, and its abundance could identify NEC cases prior to onset. Aggregate assessment of DSLNT over multiple days enhanced the separation of NEC cases and control subjects.

Conclusions DSLNT content in breast milk is a potential non-invasive marker to identify infants at risk of developing NEC, and screen high-risk donor milk. In addition, DSLNT could serve as a natural template to develop novel therapeutics against this devastating disorder.

Significance of this study

What is already known on this subject?

- Necrotising enterocolitis (NEC) is one of the most frequent and often fatal intestinal disorders in premature infants.
- Breast-fed infants are at a 6-fold to 10-fold lower risk of developing NEC than formula-fed infants.
- Human milk oligosaccharides (HMO), complex glycans that are highly abundant in breast milk but not in infant formula, prevent NEC in a neonatal rat model.
- Of the more than 150 HMO described to date, a single oligosaccharide, disialyllacto-N-tetraose (DSLNT), is responsible for the beneficial effects in neonatal rats.

What are the new findings?

- A prospective cohort study on 200 mothers and their very low birthweight infants that were predominantly human milk-fed supports the preclinical results in the neonatal rat model.
- Infants who developed NEC received less DSLNT with the milk than infants who did not develop NEC.
- Even though DSLNT concentration is occasionally low in individual milk samples consumed by control infants, the ability to discriminate between NEC cases and control infants is enhanced when samples from multiple consecutive days are averaged.

How might it impact on clinical practice in the foreseeable future?

- Low DSLNT concentrations in the mother's milk might become a non-invasive biomarker to identify breast-fed infants at risk of developing NEC.
- Donor milk and human milk fortifiers and other products might be screened for low DSLNT concentrations to avoid feeding them to infants at risk to develop NEC.
- DSLNT might be a natural template to develop novel therapeutics to help prevent NEC.

BACKGROUND

Necrotising enterocolitis (NEC) is one of the most common and devastating intestinal disorders in preterm infants.[1] It affects 5%–10% of all very low birthweight infants (VLBW; birth weight under 1500 g) and leads to a severe and often fatal destruction of the infant's intestine. More than a quarter of the affected infants die from NEC, and the survivors are often faced with long-term neurological complications.[1 2] Markers to identify at-risk infants as well as therapies to meet the clinical needs for this special and highly vulnerable population are extremely limited, and urgently needed.

Human milk-fed infants are at a 6-fold to 10-fold lower risk of developing NEC than formula-fed infants.[3–5] Several different human milk components attenuate NEC in preclinical models in rodents or piglets,[6–8] but it is not known whether

To cite: Autran CA, Kellman BP, Kim JH, *et al*. *Gut* Published Online First: [*please include* Day Month Year] doi:10.1136/gutjnl-2016-312819

these results translate to benefit the human neonate. Data from in vitro models and our own preclinical studies in neonatal rats suggest that human milk oligosaccharides (HMO) contribute to the reduced NEC incidence in human milk-fed infants.[9–11] HMO are complex glycans that represent the third most abundant component of human milk, but are currently not present in infant formula.[12] Feeding neonatal rats with HMO significantly improves survival and attenuates NEC pathology scores.[11] While more than 150 different HMO have been described so far, we identified one specific HMO called disialyllacto-N-tetraose (DSLNT) that was most effective in preventing NEC in neonatal rats. Closely related HMO like sialyllacto-N-tetraose (LSTb) or lacto-N-tetraose (LNT) that lack one or both sialic acid residues had no effect,[11] suggesting a highly structure-specific mechanism.

While the data are encouraging, the validity of available preclinical NEC models in rodents or piglets is limited.[13] Animals are exposed to external hypoxic and/or hypothermic insults that are rather artificial, and the use of animals itself is a limitation due to interspecies differences in GI development, anatomy and physiology. Thus, advancing a potential therapeutic like DSLNT from controversial preclinical models to clinical treatment trials carries a tremendous risk of failure. To help close the gap between animal models and clinical intervention studies, we used an intermediate approach and conducted a multicentre clinical prospective cohort study with mothers and their VLBW infants fed predominantly human milk. The study is based on the observation that some infants still develop NEC despite receiving predominantly human milk. HMO composition in human milk varies between women and over the course of lactation. This led us to hypothesise that human milk fed to infants who develop NEC contains less DSLNT than human milk fed to infants who do not develop NEC.

METHODS

Cohort and samples

We recruited 200 mothers and their VLBW infants (birth weight under 1500 g) at five different sites in North America (University of California, San Diego (UCSD), California, USA; Sunnybrook Health Sciences Centre, Toronto; Loma Linda University Children's Hospital, Loma Linda, California, USA; Cook Children's Health Care System, Forth Worth, Texas, USA; Rush University Medical Center, Chicago, Illinois, USA). Only infants who predominantly received human milk for at least the first 28 days of life were included. Infants were excluded if they received infant formula or had known congenital bowel anomalies such as gastroschisis. An aliquot of the milk that the infant received on a given day (not necessarily the milk that the mother produced that day) was collected every 2–3 days for the first 28 days of life, the time period when most NEC cases occur. No samples were collected if the infant was not fed that day or if there was insufficient milk to collect a 30 µL aliquot. Milk samples were stored at 4°C for <1 hour and at −20°C until shipment to the Bode lab at UCSD for HMO analysis. Basic demographic data were recorded for the mother (age, gravidity, parity, medications, medical diagnoses) and for the infant (gestational age, birth weight, sex, medications, medical diagnoses). NEC was diagnosed based on the modified Bell staging system.[14] Further clinical details for each NEC case are found in the Case Descriptions in the online supplementary appendix.

Once all 200 mother-infant pairs were recruited and all samples were collected, each NEC case was matched with five controls (mothers and their infants who did not develop NEC). Controls were selected from the same study site to minimise location effects. In addition to location, matching criteria included gestational age, birth weight, mode of delivery, race and ethnicity as well as availability of milk samples throughout the 28-day collection period. An example of case-control matching is shown in online supplementary figure S1 in the appendix. The institutional review boards at all five participating study sites approved the research protocol, and parents of all participants gave their written informed consent.

Human milk oligosaccharide analysis

HMO composition was analysed in all available milk samples from all NEC cases and the respective matched controls. Analytical staff was blinded to the clinical metadata associated with each sample. HMO analysis was performed by high-performance liquid chromatography (HLPC) after fluorescent derivatisation with 2-aminobenzamide (2AB) as previously described.[15 16] The non-HMO oligosaccharide raffinose was added to each milk sample as internal standard at the very beginning of sample preparation to allow for absolute quantification. The following individual HMO were detected based on retention time comparison with commercial standard oligosaccharides and mass spectrometry analysis: 2′-fucosyllactose (2′FL), 3-fucosyllactose, 3′-sialyllactose, LNT, lacto-N-neotetraose (LNnT), lacto-N-fucopentaose (LNFP)1, LNFP2 and LNFP3, sialyl-LNT b (LSTb) and LSTc, difucosyl-LNT (DFLNT), DSLNT, fucosyl-lacto-N-hexaose, difucosyl-lacto-N-hexaose, fucosyl-disialyl-lacto-N-hexaose and disialyl-lacto-N-hexaose. In addition to absolute concentrations, the proportion of each HMO per total HMO concentration (sum of all integrated HMO) was calculated and expressed as relative abundance (% of total). Secretor status was defined by the presence of 2′FL. Simpson's diversity index D was calculated as the reciprocal sum of the square of the relative abundance of each of the measured HMO. HMO equitability (evenness, E) was calculated by dividing the actual D index for each sample by Dmax (maximum D index in the theoretical case that all measured HMO have the same relative abundance).

Statistical analysis

Initial comparison of HMO levels was done with the Mann-Whitney U test, Wilcoxon test and Kruskal-Wallis test, since the distributions of HMO concentrations were non-normal according to Shapiro-Wilk tests. Univariate logistic regression models were used to prescreen clinical covariates, including gestational age, birth weight, mode of delivery, race, etc. To test if HMO significantly influenced the onset of NEC, we estimated its effect using generalised estimating equation (GEE) models to account for longitudinal measurement.[17] GEE allows non-linear relations between the outcome and covariates, and accounts for unknown correlation among repeated measurements from the same subject. Here we used GEE with logit link and exchangeable correlation structure, by assuming the within-subject correlation between any two time-points is ρ. To stabilise the variance and equalise the range, we standardised each HMO measurement, for example, DSLNT, LNFP1. We also used the square root of days post partum (DPP) to linearise the relationship over time. NEC status (ie, Bell stage) was used as outcome, and the Wald test was used to assess statistical significance of model components. To reduce variation and allow comparison between HMO, oligosaccharide concentrations were standardised by subtracting the mean and dividing by the SD.

We first prescreened each HMO using a univariate GEE model. Any HMO with $p<0.2$ were further analysed to assess their joint contribution to NEC onset in combination with

significant clinical covariates to construct final GEE model. Quasi-likelihood under the independence model criterion (QIC) was used to compare multivariate models. QIC and Wald statistics informed a backward elimination process of model selection. Detailed information on statistical analysis is provided in the online supplementary appendix.

RESULTS

Eight of the 200 recruited infants developed NEC Bell stage 2 or 3, and two infants were diagnosed with NEC Bell stage 1. Table 1 shows the case-control characteristics stratified by study site A to E, and Case Descriptions in the online supplementary appendix provide detailed clinical information for each NEC case. We analysed the HMO composition in a total of 636 milk samples (15 samples from Bell stage 1, 38 from stage 2, 25 from stage 3 and 558 from controls).

DSLNT concentrations are significantly lower in milk fed to NEC cases compared with controls

We first compared the concentrations of each HMO from all 636 milk samples using the Mann-Whitney U test, Wilcoxon test and Kruskal-Wallis test. Figure 1A shows that DSLNT concentrations in NEC cases (Bell stage 2–3) were significantly lower than DSLNT concentrations in controls. However, there was no significant difference between NEC cases and controls when looking at the sum of all HMO (figure 1B) or any of the individual HMO other than DSLNT (see online supplementary figure S2 in the appendix).

While the first analysis combined HMO results from all NEC cases and compared it with results from all controls, we next examined HMO concentrations for each separate NEC case and its associated controls. For each milk sample, the fold change for each HMO was calculated relative to its average concentration in the associated control group (figure 2). Controlling for known clinical factors through our case-control matching further suggested that DSLNT exhibits a consistently lower concentration in all NEC cases, throughout the length of the study (figure 2A). Furthermore, the magnitude of deficiency worsened with Bell stage. Indeed, NEC Bell stage 3 cases showed the lowest DSLNT concentration compared with controls, and Bell stage 1 cases exhibited the weakest effect. The oligosaccharides LSTb and LNT (structurally similar to DSLNT but with reduced sialylation) did not show consistent differences in the case-control matching (figure 2B, C). Similarly, total HMO concentration did not differ considerably between cases and their clinically matched controls (figure 2D).

A robust statistical assessment of all measured HMO shows DSLNT as the primary contributor

To test for associations between clinical covariates and NEC onset, we used univariate logistic regression model to prescreen each clinical covariate measured at birth. None of the clinical covariates (ie, study site, delivery mode, race/ethnicity or gender) or characteristics of the infant at birth (ie, gestational age or birth weight) was significantly associated with NEC onset in our cohort (figure 3A). The insignificance of birth characteristics indicated that our cohort was successfully matched for variance due to birth weight, gestational age and location, and potential confounding based on these covariates was minimised. Our analysis concentrated on Bell stage 2 and 3 as Bell stage 1 is generally considered less specific for NEC diagnosis.

We then assessed the contribution of each HMO to the onset of NEC using a univariate GEE to account for repeated measurements. In this analysis, DSLNT clearly contributed ($p<0.001$; figure 3B), with an OR of 0.86, suggesting that lower concentrations increase the risk of developing NEC. Additionally, HMOs such as LNFP1, LNFP3 and DFLNT were selected ($p<0.2$) from the univariate screening for further examination. A final multivariate model demonstrated that DSLNT has a significantly protective contribution (OR 0.84; 95% CI 0.79 to 0.88; $p=0.001$), while LNFP1 and DFLNT are associated with a decreased (OR 0.91; 95% CI 0.84 to 0.97; $p=0.006$) and increased (OR 1.15; 95% CI 1.01 to 1.28; $p=0.022$) risk of NEC, respectively, albeit to a much lesser

Table 1 Case-control characteristics

Case ID	Bell stage	Gestational age (weeks+days)	Birth weight (g)	Gender	Delivery mode	Race/ethnicity
A003	2	29+2	1350	M	C	A
Controls		29+0.6±4.4	1120±234	(M, M, M, M, F)	(C, C, C, C, C)	(H, H, H, Ca, A)
A029	2	27+3	1069	M	C	Ca
Controls		27+2.1±3.5	1022±103	(M, F, M, M, F)	(V, V, C, C, V)	(H, NS, NS, AA, AA)
A066	1	26+6	961	M	C	H
Controls		26+4.9±0.7	950±173	(F, M, M, M, F)	(C, C, C, C, C)	(Ca, Ca, Ca, H, Ca)
B032	3	31+1	1160	F	V	AA
Controls		31+6.4±7.6	1382±96	(F, F, M, F, M)	(V, V, C, C, V)	(AA, A, H, Ca, AA)
C005	3	26+3	1040	M	C	AA
Controls		26+0.4±8.2	766±228	(F, F, F, F, M)	(V, C, V, C, V)	(H, AA, AA, Ca, AA)
C024	2	27+1	830	F	C	AA
Controls		26+4.6±3.2	928±120	(M, F, F, M, F)	(C, V, C, C, C)	(AA, Ca, H, Ca, AA)
C027	2	29+5	1370	F	C	AA
Controls		29+2.8±2.7	1276±142	(M, F, M, M, F)	(V, C, C, C, C)	(Ca, AA, Ca, AA, H)
D015	1	26+1	813	M	V	NS
Controls		26+3.2±2.4	771±224	(M, M, M, M, M)	(C, C, C, C, C)	(NS, NS, NS, NS, NS)
D024	2	26+2	770	F	C	NS
Controls		26+0.8±5.0	790±116	(F, F, F, F, F)	(V, C, V, C, C)	NS
E002	3	25+0	1040	M	V	H
Controls		26+5.1±17.5	823±197	(F, F, M, M, F)	(C, C, C, V, V)	(Ca, AA, H, H, Ca)

A, Asian; AA, African-American; C, caesarean; Ca, Caucasian; H, Hispanic; NS, not specified; V, vaginal.

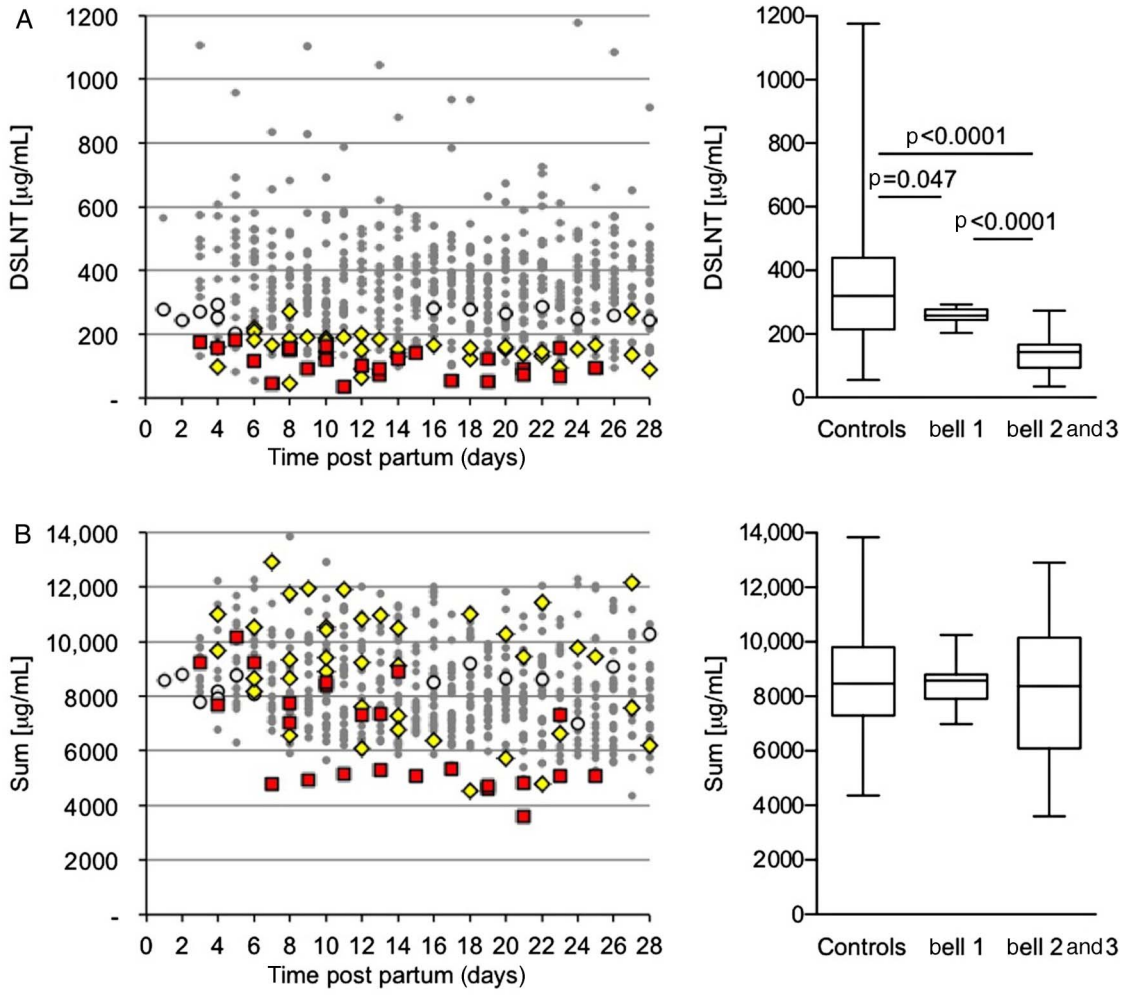


Figure 1 Disialyllacto-N-tetraose (DSLNT) concentrations are consistently lower in necrotising enterocolitis (NEC) cases. Concentrations of DSLNT alone (A) and all human milk oligosaccharides (HMO) combined (B) in human milk were plotted as a function of days post partum when the milk was given to infants who either developed NEC (red squares: Bell stage 3; yellow diamonds: Bell stage 2; grey circles: Bell stage 1 or did not develop NEC and served as controls (small grey circles). Box plots on the right show median with 25/75 quartiles and whiskers for min/max values. The Bell stage groups were compared by Mann-Whitney U test and Kruskal-Wallis test. Distribution normality was rejected by the Shapiro-Wilk test ($p<0.001$). DSLNT concentrations in human milk were significantly lower in infants who developed NEC (stage 2 and 3 combined) when compared with controls. However, there was no significant difference in total HMO concentration (sum of all integrated individual HMO), and there were also no differences in any HMO other than DSLNT (see online supplementary figure S2).

extent than DSLNT, as detailed in figure 3C and online supplementary table S1. LNFP3 did not significantly contribute to univariate or multivariate models (see online supplementary table S1) and its removal did not considerably increase QIC (see online supplementary table S2). Therefore, LNFP3 was excluded from the final multivariate model. A minimal model accounting only for DSLNT and DPP, provided only a small increase in QIC relative to the final model with three HMO, further supporting the dominant contribution of DSLNT to NEC (see online supplementary table S2).

NEC cases exhibit consistently perturbed HMO concentration over time

In addition to highlighting HMO that are associated with NEC, the univariate DSLNT model and multivariate models were capable of identifying individual milk samples associated with NEC cases (see online supplementary figure S3). Indeed, the models could identify potentially problematic milk samples or identify infants who may benefit from intervention prior to the possible onset of NEC. Furthermore, we note that DSLNT concentration tends to be significantly perturbed in NEC cases for multiple consecutive days (see online supplementary figure S4 in the appendix). Thus, even though DSLNT concentration is occasionally diminished in individual milk samples consumed by control infants (see online supplementary figure S4A), the ability to discriminate between NEC cases and control infants is enhanced when samples from multiple consecutive days are averaged (figure 4, see online supplementary figure S4B and S5 in the appendix).

DISCUSSION

For decades, efforts have been made to understand why human milk-fed infants are at significantly lower risk to develop NEC. While human milk components have previously been shown in preclinical models to provide a protective effect,[6–8 11] none of these findings, until now, has been validated in human infants. In this study, we demonstrated in a clinical cohort that DSLNT may provide a significant protective effect against the onset of NEC. These results validate our earlier observation of the protective effect of DSLNT on neonatal rats.[11] DSLNT deficiency was identified as a major contributor to NEC from a panel of the 16 most

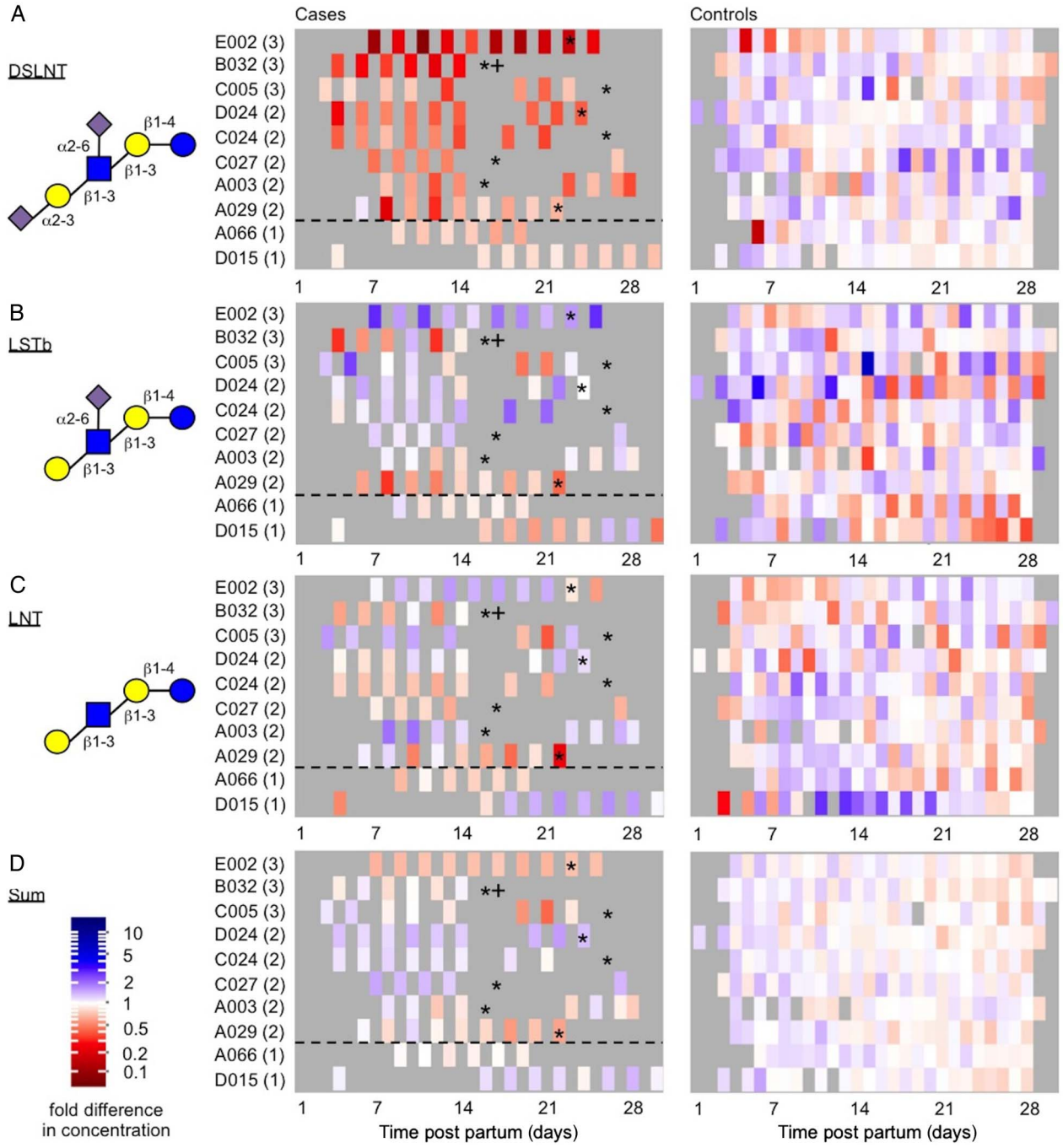


Figure 2 Disialyllacto-N-tetraose (DSLNT) concentrations are uniquely and consistently low in necrotising enterocolitis (NEC) cases (left) when compared with controls (right). Samples in each row are case-control matched by study site, gestational age, birth weight and other NEC-relevant factors. For each milk sample collected over the first 28 days post partum, the fold difference of human milk oligosaccharides (HMO) concentration relative to the associated matched control sample average is illustrated for DSLNT (A), sialyllacto-N-tetraose (LSTb) (B), lacto-N-tetraose (LNT) (C) and the sum of all integrated HMO (D). DSLNT was lowest in cases with a Bell stage of 3 and 2 at concentrations an order of magnitude lower than matched control averages. Bell stage 1 cases showed slightly lower concentrations than their matched controls. Structurally similar HMOs with reduced sialylation, such as LSTb and LNT, failed to exhibit consistent variations in concentration in NEC cases compared with matched controls. Number in parentheses after case codes denotes NEC Bell stage. (*) denotes the day of NEC onset, (+) denotes the day of death due to NEC. Oligosaccharide structure nomenclature: blue circles: glucose; yellow circles: galactose; blue squares: N-acetylglucosamine; purple diamonds: sialic acid.

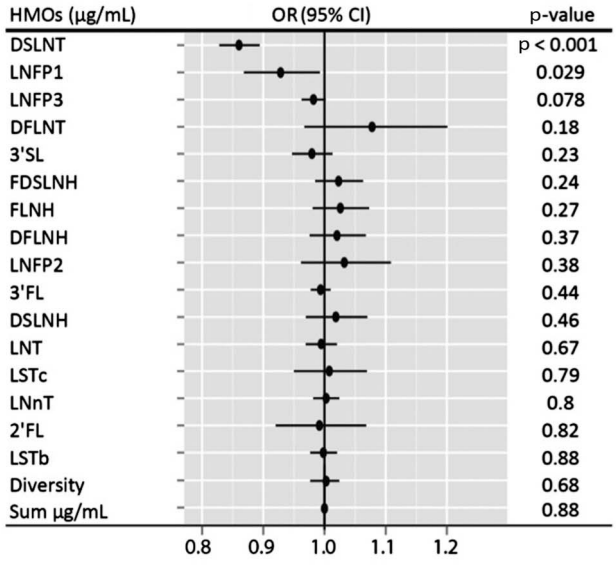
abundant HMO, which represent >95% of the HMO in human milk. In this panel, additional HMO (ie, LNFP1 and DFLNT) were also found to provide a lesser contribution. It is important to note that we did not measure total milk volume fed to each infant per feeding or per day. The analysis purely focuses on HMO concentrations and not on absolute HMO amounts received.

The role of DSLNT and other HMO were robustly quantified using GEE. DSLNT deficiency is significantly associated ($p=0.001$) with NEC onset with an OR of 0.84. LNFP1 and DFLNT were also found to have a significant protective (OR 0.91) and harmful (OR 1.14) contribution, respectively (figure 3C). These contributions were stable across all

A Univariate regression

Clinical covariates	p-value
Delivery	0.69
Gender	0.96
Secretor status	0.5
Location	0.96
Race/ethnic	0.34
Birth weight	0.28
Gestational age	0.77

B Univariate GEE model



C Final model

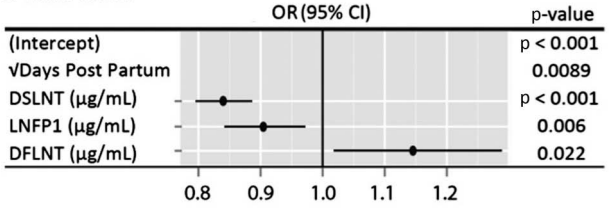


Figure 3 Univariate logistic regression screening of (A) birth characteristics (ie, birth weight and gestational age) were effectively controlled through case-control matching, and therefore, no clinical covariate showed a significant association with necrotising enterocolitis (NEC). (B) Univariate temporal logistic generalised estimating equation (GEE) screening of human milk oligosaccharides (HMOs) revealed several candidate associations, with disialyllacto-N-tetraose (DSLNT) being the predominant HMO. (C) The final multivariate temporal GEE model demonstrated that DSLNT, lacto-N-fucopentaose (LNFP1) and difucosyl-LNT (DFLNT) each contribute significantly to a final multivariate model. The OR is the exponentiated coefficient and the 95% CI describes the range of possible OR. For panel A, the p value represents the significance based on the χ^2 distribution, while p values in panels B and C were calculated from the Wald statistic of each coefficient, evaluated along a normal distribution.

multivariate models (see online supplementary table S1). Furthermore, these HMO were consistently dysregulated in NEC cases. When considering dysregulation over multiple consecutive days, the separation between cases and controls increased (see online supplementary figures S2–S4), suggesting that prolonged dysregulation of HMO is more indicative of NEC onset. Interestingly, NEC Bell stage 1 did not correlate with DSLNT deficiency, supporting the lack of specificity of Bell

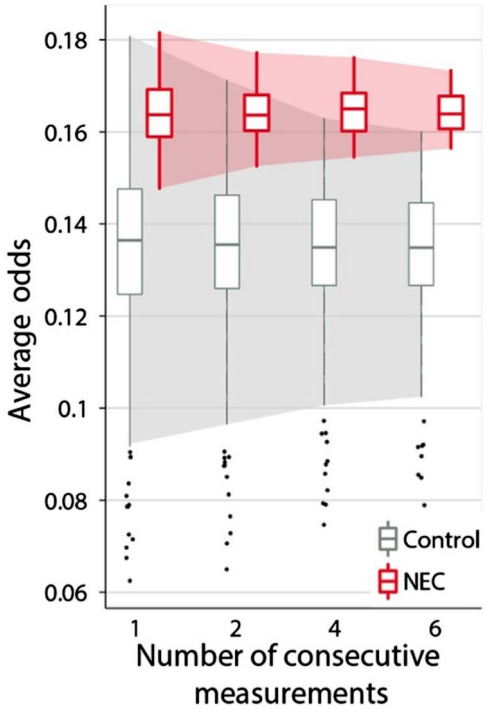


Figure 4 Aggregation of disialyllacto-N-tetraose (DSLNT) concentration for multiple days enhances the identification of high-risk infants. Infants who will develop necrotising enterocolitis (NEC) are more readily identifiable when DSLNT concentration from multiple consecutive milk samples for each subject is aggregated using the geometric mean. When we combine the computed odds for 2, 4 and 6 consecutive milk samples from the same individual, separation of cases and controls increased and variance in the average odds decreased.

stage 1 in diagnosing NEC or suggesting that DSLNT deficiency only impacts infants' risk for more advanced NEC.

The underlying mechanisms of how HMO such as DSLNT attenuate NEC risk remain to be elucidated. Although HMO have profound effects on infant microbiota composition,[18–20] the importance of microbiota composition on NEC onset and development is poorly understood.[21–26] Whether microbial dysbiosis is a causative event or merely a marker of intestinal disease remains unknown.[27] Instead, HMO may have direct effects on infant intestinal epithelial or immune cells, which might directly attenuate NEC risk, and also indirectly alter microbiota composition. The observation that the effects of DSLNT are highly structure-specific (removal of just one sialic acid renders the oligosaccharide ineffective in neonatal rats[11] and these truncated oligosaccharides are no longer associated with NEC risk in the cohort study) indicates a potentially receptor-mediated mechanism.

The study recruited 200 mothers and their VLBW infants, of which 8 (4%) developed NEC Bell stage 2 or 3. NEC incidence in VLBW infants in North America typically varies between <5 and up to 10%, but that includes both human milk-fed as well as formula-fed infants. Since NEC incidence is 6-fold to 10-fold lower in predominately human milk-fed infants compared with formula-fed infants,[3–5] the 4% NEC incidence reported in this study is well within the anticipated range.

While the results from this study indicate that higher DSLNT concentrations in mother's milk lower the infant's risk to develop NEC, larger cohort studies with more detailed maternal data will be needed to identify maternal factors (genetics, nutrition, stress, etc) that influence DSLNT synthesis.

Although selection bias is a common limitation of case-control studies, this has been minimised in two ways: first by

prospective enrolment at five different locations before identification of cases or controls, and, second, by matching cases with controls from their own location, and therefore with the most similar unmeasured exposures.

Current practice aims to reduce the risk of NEC through the administration of human milk from the mother or a donor. While human milk in general reduces the risk of NEC, it remains vastly unknown how the natural variation in the concentration of specific human milk components contributes to NEC risk. Our data suggest that feeding preterm infants with human milk rich in DSLNT lowers NEC risk, while feeding human milk deficient in DSLNT increases NEC risk. While cohort association studies cannot exclude that DSLNT is simply a proxy for other maternal or infant markers, the data are consistent with results from preclinical studies showing that supplementation with DSLNT (alone and without any other confounders) improves outcome measures in neonatal rodents.[11] The combined datasets from preclinical study and mother-infant cohort increase our confidence that feeding human milk rich in DSLNT lowers NEC risk. However, larger cohort studies as well as clinical intervention studies are needed to validate the association of DSLNT with reduced NEC risk. If confirmed, low DSLNT concentrations in the mother's milk might become a non-invasive marker to identify breast-fed infants at risk of developing NEC. Donor milk, human milk fortifiers and other products might be screened for low DSLNT concentrations to avoid feeding them to infants at risk to develop NEC. In addition, DSLNT might serve as a natural template to develop novel therapeutics to help prevent NEC.[28]

Author affiliations
[1]Department of Pediatrics, University of California San Diego School of Medicine, La Jolla, California, USA
[2]Bioinformatics and Systems Biology Program, University of California, La Jolla, California, USA
[3]Department of Newborn & Developmental Pediatrics, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada
[4]Department of Pediatrics, Loma Linda University School of Medicine, Loma Linda, California, USA
[5]Cook Children's Health Care System, Forth Worth, Texas, USA
[6]Rush University Medical Center, Chicago, Illinois, USA
[7]Clinical & Translational Research Institute, University of California San Diego, La Jolla, California, USA
[8]Novo Nordisk Foundation Center for Biosustainability at the University of California San Diego School of Medicine, La Jolla, California, USA

Twitter Follow Benjamin Kellman @bkell1123

Acknowledgements We thank the following staff for their help and dedication with subject recruitment and sample collection: Renee Bridge RN and Lisa Lepis RN at the University of California, San Diego; Rosine Bishara RD MSc at Sunnybrook Health Sciences Centre, Toronto; Andrea Pinto at Loma Linda University School of Medicine; Virginia Campbell RN, Study Coordinator RN at Cook Children's Health Care System, Forth Worth, TX; Judy Janes RN, Emily White RN, Esmeralda Covarrubias IBCLC, Katherine Yucius RN, Katherine McGee RN IBCLC, Kathleen Dolan RN, Kathy Lombardo RN, Krystle Ekhoff IBCLC, LaShanna Kimmons IBCLC, Mayra Espinoza IBCLC at Rush University Medical Center, Chicago, Illinois, USA.

Contributors LB and JHK designed research; CAA, LB, EA, ABB, ECHS, ALP and JHK recruited patients, collected samples and conducted research; CAA, BPK, NEL, JH and LB analysed data and performed statistical analysis; BPK, JH, JHK, NEL and LB wrote the manuscript; LB had primary responsibility for final content. All authors read and approved the final version of the manuscript.

Funding The project was supported by a grant from Abbott Nutrition, a division of Abbott Laboratories, and the National Institutes of Health, Grant R00DK078668 (LB) and UL1TR000100 (Clinical and Translational Research Institute), and support from the Novo Nordisk Foundation that had been provided to the Center for Biosustainability at the Technical University of Denmark (NEL). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests None declared.

Ethics approval University of California San Diego Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Neu J, Walker WA. Necrotizing enterocolitis. *N Engl J Med* 2011;364:255–64.
2 Hintz SR, Kendrick DE, Stoll BJ, *et al*., NICHD Neonatal Research Network. Neurodevelopmental and growth outcomes of extremely low birth weight infants after necrotizing enterocolitis. *Pediatrics* 2005;115:696–703.
3 Lucas A, Cole TJ. Breast milk and neonatal necrotising enterocolitis. *Lancet* 1990;336:1519–23.
4 Schanler RJ, Lau C, Hurst NM, *et al*. Randomized trial of donor human milk versus preterm formula as substitutes for mothers' own milk in the feeding of extremely premature infants. *Pediatrics* 2005;116:400–6.
5 Meinzen-Derr J, Poindexter B, Wrage L, *et al*. Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. *J Perinatol* 2009;29:57–62.
6 Caplan MS, Russell T, Xiao Y, *et al*. Effect of polyunsaturated fatty acid (PUFA) supplementation on intestinal inflammation and necrotizing enterocolitis (NEC) in a neonatal rat model. *Pediatr Res* 2001;49:647–52.
7 Lu J, Pierce M, Franklin A, *et al*. Dual roles of endogenous platelet-activating factor acetylhydrolase in a murine model of necrotizing enterocolitis. *Pediatr Res* 2010;68:225–30.
8 Shiou SR, Yu Y, Guo Y, *et al*. Oral administration of transforming growth factor-β1 (TGF-β1) protects the immature gut from injury via Smad protein-dependent suppression of epithelial nuclear factor κB (NF-κB) signaling and proinflammatory cytokine production. *J Biol Chem* 2013;288:34757–66.
9 Bode L, Kunz C, Muhly-Reinholz M, *et al*. Inhibition of monocyte, lymphocyte, and neutrophil adhesion to endothelial cells by human milk oligosaccharides. *Thromb Haemost* 2004;92:1402–10.
10 Bode L, Rudloff S, Kunz C, *et al*. Human milk oligosaccharides reduce platelet-neutrophil complex formation leading to a decrease in neutrophil beta 2 integrin expression. *J Leukoc Biol* 2004;76:820–6.
11 Jantscher-Krenn E, Zherebtsov M, Nissan C, *et al*. The human milk oligosaccharides disialyllacotse-N-tetraose prevents Necrotizing Enterocolitis in neonatal rats. *Gut* 2012;61:1417–25.
12 Bode L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* 2012;22:1147–62.
13 Tanner SM, Berryhill TF, Ellenburg JL, *et al*. Pathogenesis of necrotizing enterocolitis: modeling the innate immune response. *Am J Pathol* 2015;185:4–16.
14 Bell MJ, Ternberg JL, Feigin RD, *et al*. Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. *Ann Surg* 1978;187:1–7.
15 Bode L, Kuhn L, Kim HY, *et al*. Human milk oligosaccharide concentration and risk of postnatal transmission of HIV through breastfeeding. *Am J Clin Nutr* 2012;96:831–9.
16 Alderete TL, Autran C, Brekke BE, *et al*. Associations between human milk oligosaccharides and infant body composition in the first 6 mo of life. *Am J Clin Nutr* 2015;102:1381–8.
17 Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–30.
18 Sela DA, Chapman J, Adeuya A, *et al*. The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci USA* 2008;105:18964–9.
19 Marcobal A, Barboza M, Sonnenburg ED, *et al*. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* 2011;10:507–14.
20 Wang M, Li M, Wu S, *et al*. Fecal microbiota composition of breast-fed infants is correlated with human milk oligosaccharides consumed. *J Pediatr Gastroenterol Nutr* 2015;60:825–33.
21 Wang Y, Hoenig JD, Malin KJ, *et al*. 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *ISME J* 2009;3:944–54.
22 Mai V, Young CM, Ukhanova M, *et al*. Fecal microbiota in premature infants prior to necrotizing enterocolitis. *PLoS One* 2011;6:e20647.
23 Claud EC, Keegan KP, Brulc JM, *et al*. Bacterial community structure and functional contributions to emergence of health or necrotizing enterocolitis in preterm infants. *Microbiome* 2013;1:20.
24 Morrow AL, Lagomarcino AJ, Schibler KR, *et al*. Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome* 2013;1:13.
25 Torrazza RM, Ukhanova M, Wang X, *et al*. Intestinal microbial ecology and environmental factors affecting necrotizing enterocolitis. *PLoS One* 2013;8:e83304.
26 Zhou Y, Shan G, Sodergren E, *et al*. Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study. *PLoS One* 2015;10:e0118632.
27 Elgin TG, Kern SL, McElroy SJ. Development of the neonatal intestinal microbiome and its association with necrotizing enterocolitis. *Clin Ther* 2016;38:706–15.
28 Yu H, Lau K, Thon V, *et al*. Synthetic disialyl hexasaccharides protect neonatal rats from necrotizing enterocolitis. *Angew Chem Int Ed Engl* 2014;53:6687–91.

Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants

Chloe A Autran, Benjamin P Kellman, Jae H Kim, Elizabeth Asztalos, Arlin B Blood, Erin C Hamilton Spence, Aloka L Patel, Jiayi Hou, Nathan E Lewis and Lars Bode

Gut published online April 5, 2017

SUPPLEMENTARY APPENDIX

Human milk oligosaccharide composition predicts risk of necrotizing enterocolitis in preterm infants

Chloe A Autran, MSc[1*], Benjamin P Kellman, MSc[1,2*], Jae H Kim, MD[1], Elizabeth Asztalos, MD[3], Arlin B Blood, MD[4], Erin C Hamilton Spence, MD[5], Aloka L Patel, MD[6], Jiayi Hou, PhD[7], Nathan E Lewis, PhD[1,2,8], Lars Bode, PhD[1]

[1]Department of Pediatrics, University of California San Diego School of Medicine, San Diego, CA, USA

[2]Bioinformatics and Systems Biology Program, University of California, San Diego, CA, USA

[3] Department of Newborn & Developmental Pediatrics, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[4]Department of Pediatrics, Loma Linda University School of Medicine, Loma Linda CA, USA

[5]Cook Children's Health Care System, Forth Worth, TX, USA

[6]Rush University Medical Center, Chicago, IL, USA

[7]Clinical & Translational Research Institute, University of California San Diego, San Diego, CA, USA

[8]Novo Nordisk Foundation Center for Biosustainability at the University of California San Diego School of Medicine, San Diego, CA, USA

*Authors contributed equally

Corresponding Author: Lars Bode, PhD

Department of Pediatrics

University of California San Diego School of Medicine

9500 Gilman Dr., Mail code 0715

San Diego, CA 92093

USA

lbode@ucsd.edu phone: 858 246 1874

Table of Contents

1. Supplementary Figures

ID#	GA	BW	D	RE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
A003	29_2	1,350	C	A	N	N	N					**01**		**02**		**03**		**04**		N	N	N	N	N	N	N	**05**		**06**		**07**	**08**
A002	29_0	1,114	C	A	N	N	N	N	N	N						N			01		02		03		04		05		06		07	
A005	29_3	1,420	C	A	N	N		01		02					03	04		05		06		07		08			09		10		11	
A018	29_1	1,276	C	A	N	N	N	N		01		02		03		04	N	05		06		07		08		09		10		11		12
A034	28_0	944	C	W	N	N	N	N		01		02		03		04		05		06		07		08		09		10		11		12
A038	29_4	850	C	A	N	N			N	N				01		02		03		04		05		06		07		08	Transferred out			

Figure S1. Example of case-control matching and milk sample availability. Case A003 was matched with 5 controls from the same study site A based on gestational age (GA), birth weight (BW), delivery mode (D), then race/ethnicity (RE). Matching was also based on the similarity of collection days. Each number in green boxes shows that a milk sample was available for analysis for that specific day and subject. A003 was diagnosed with NEC on day-of-life 16 (until 22 days post partum; red boxes). N: *non per os* (infant did not receive oral feeds that day).

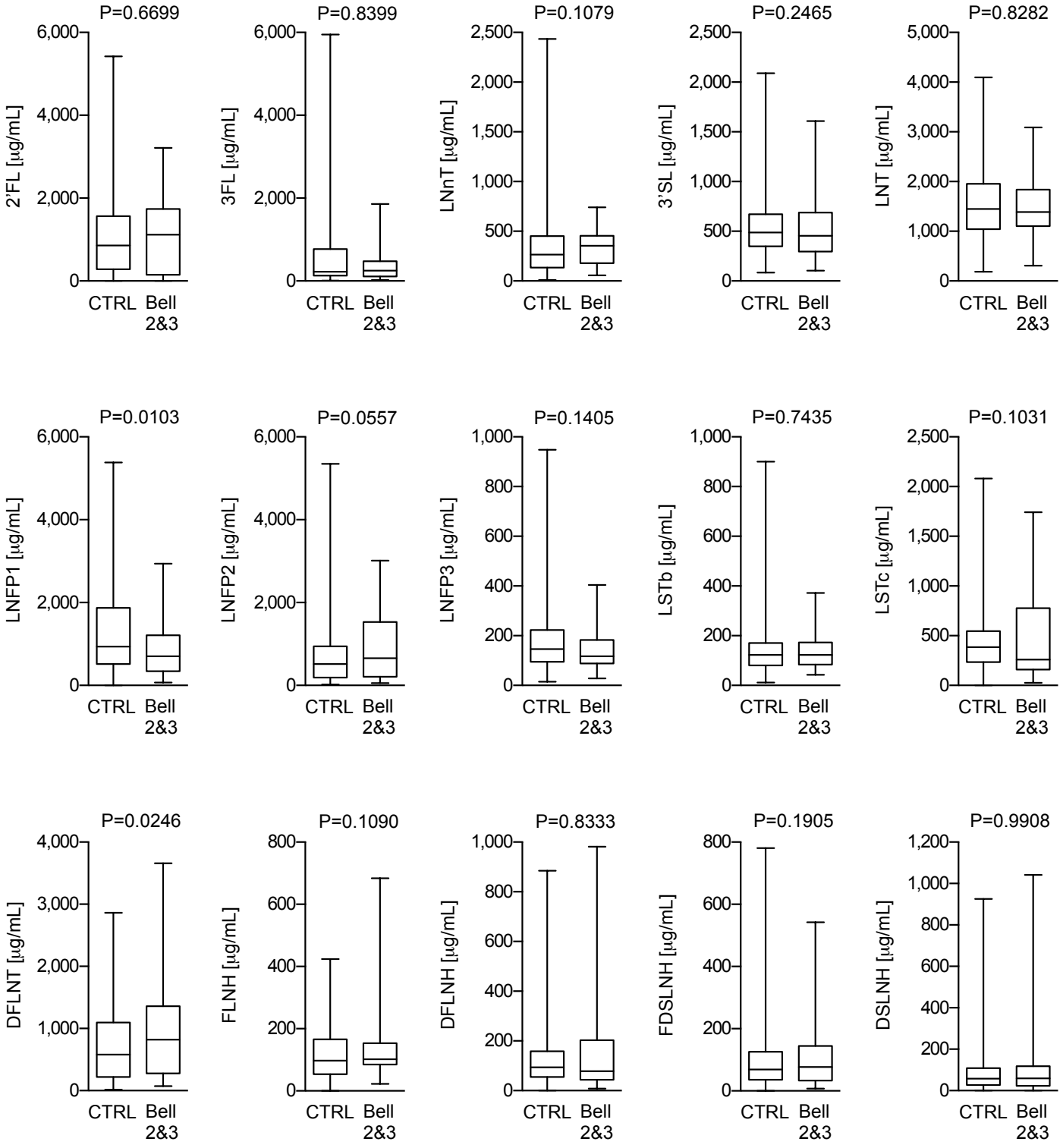


Figure S2. HMO comparisons for all NEC cases and all controls combined. Concentrations of all measured HMO were compared between NEC (Bell stages 2 & 3) cases and non-NEC controls. All available samples were included for each subject. Only DSLNT (**Figure 1**), LNFP1, and DFLNT exhibited significant changes in concentration.

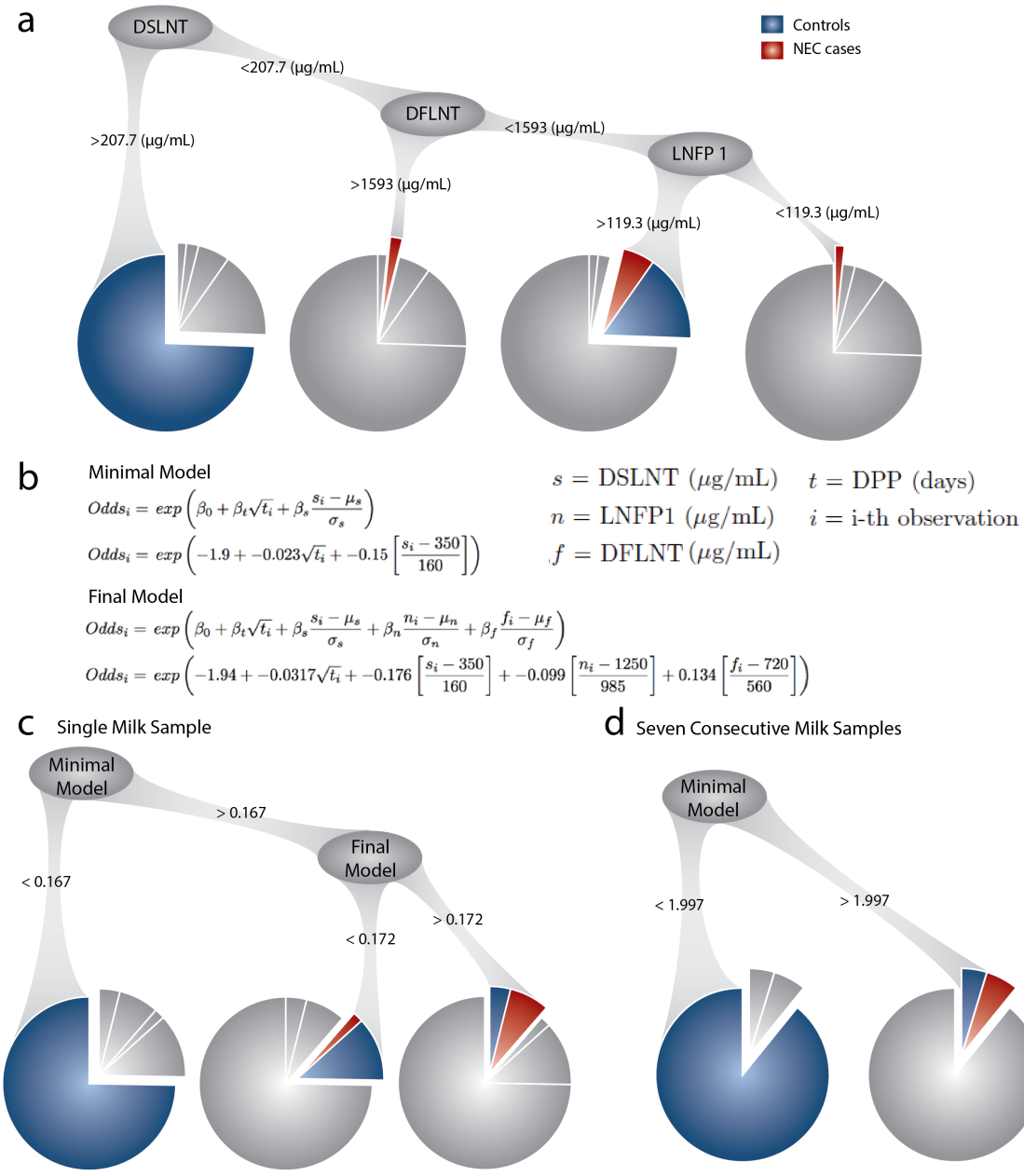


Figure S3. Identification of high NEC risk milk samples based on HMO analysis. (A) In this study, we identified three HMO that may aid in identifying milk samples of concern, and these were used to construct a decision tree for sample classification. Samples with greater than 207.7 µg/mL DSLNT are considered low risk by this tree. Samples with lower DSLNT concentrations are higher risk and can be subsequently tested for DFLNT and LNFP1 concentration to further identify high risk samples. The concentrations reported here are estimates based on the size of the study. More quantitatively robust thresholds can be achieved in future studies that address the longitudinal sampling in larger cohorts. **(B)** Equations for the odds that a given milk sample is associated with an infant that will develop NEC can be calculated based on DSLNT (minimal model) or DSLNT, LNFP1 and DFLNT (Final Model). **(C)** The final model based odds calculations can be used to partition samples that are associated with infants who will develop NEC. However, LNFP1 and DFLNT provide less information than DSLNT towards classification of milk samples. **(D)** The identification of milk associated with NEC cases is enhanced by the combined assessment of consecutive milk samples from a given infant, demonstrating that NEC is associated with consistently and continually altered levels of HMOs, specifically DSLNT.

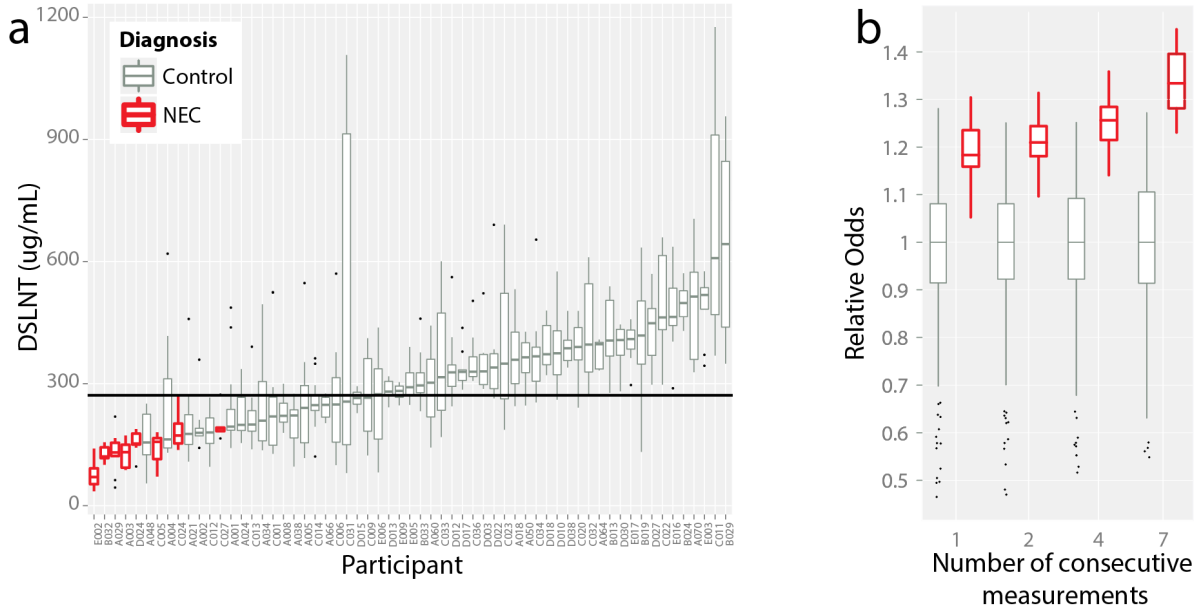


Figure S4. Analysis of DSLNT concentration for multiple days enhances the identification of high-risk infants. (A) DSLNT concentration is low in individual milk samples consumed by preterm infants, but many infants are exposed to occasional feedings with low DSLNT without developing NEC. **(B)** The infants who will develop NEC are more readily identifiable when DSLNT concentrations from multiple consecutive milk samples for each subject are cumulatively analyzed. When we combine the computed odds for 2, 4, and 7 consecutive milk samples from the same individual, separation of cases and controls increased. To compute relative odds, the odds for all NEC cases and controls were normalized by dividing their values by the median odd value for non-NEC controls for the corresponding day.

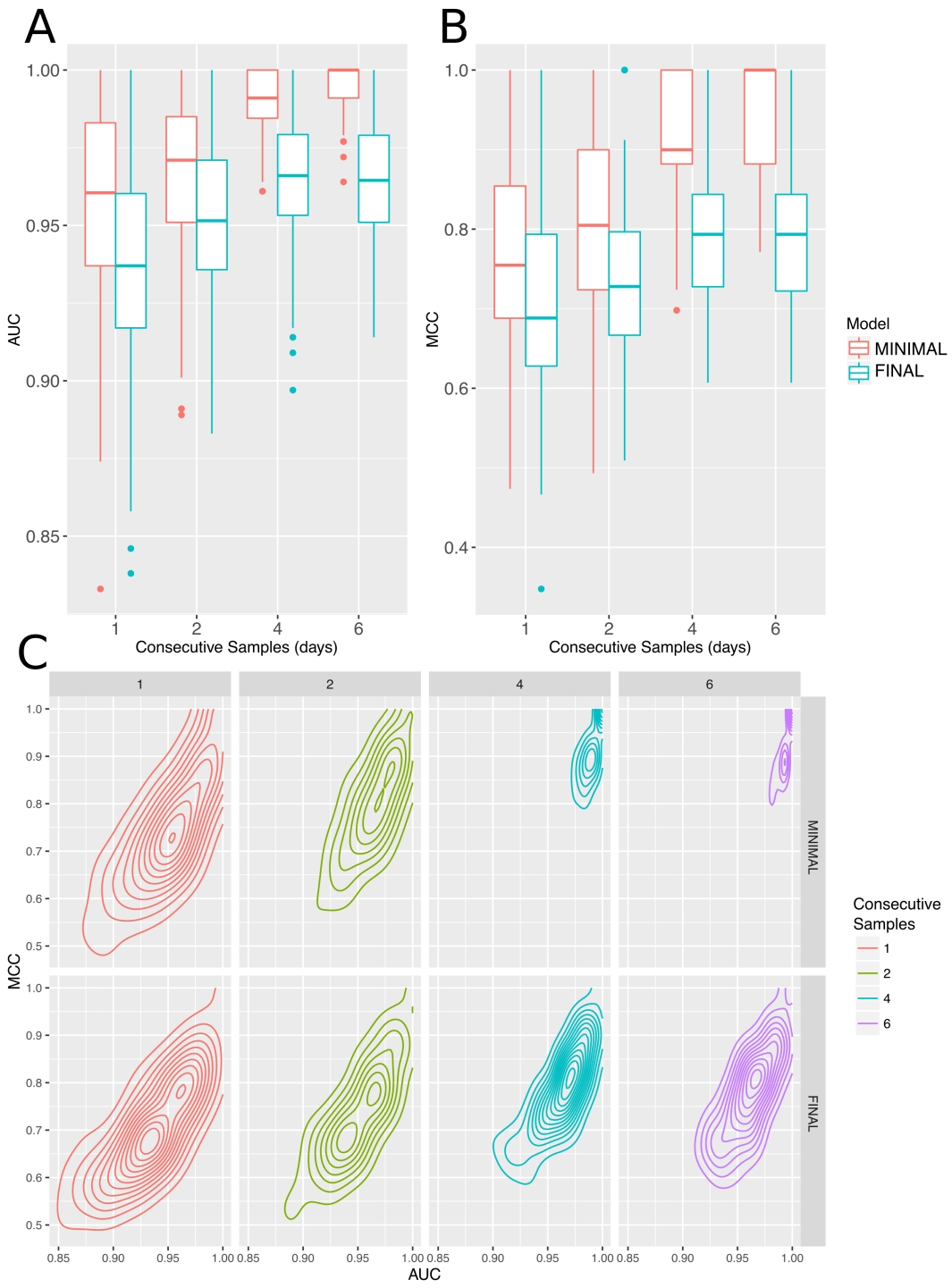


Figure S5. Early Estimate of Milk Classification Performance. Resampling-Bias Conscious Estimation of High/Low Risk Milk Classification Performance. **A.** Area under the (ROC) Curve (AUC) was used to evaluate the performance of our milk risk odds estimators. **B.** Mathews Correlation Coefficient (MCC) was also used to assess performance accounting for bias due to the class imbalance (the ratio of case:control subjects is approximately 3:20 while ratio of case:control samples is approximately 1:10). The odds estimating models assesed here are the "full" model (blue) which considers time, DSLNT, LNFP1 and DFLNT as well as the "minimal" model (red) which considers only time and DSLNT. **C.** AUC vs MCC aggregated over different numbers of days (top border and color) and different models (right border). The optimal cutoff in these models was approximatley odds=0.16.

2. Supplemental Tables

Minimal Model	OR	95% CI	Pr(W)
(Intercept)	0.15	(0.04 - 0.259)	P<0.001
$\sqrt{DPP}$	0.98	(0.963 - 0.991)	0.0013
standardized(DSLNT)	0.86	(0.828 - 0.893)	P<0.001

Minimal Model + LNFP1	OR	95% CI	Pr(W)
(Intercept)	0.16	(0.0414 - 0.272)	P<0.001
$\sqrt{DPP}$	0.96	(0.945 - 0.982)	P<0.001
standardized(DSLNT)	0.86	(0.831 - 0.894)	P<0.001
standardized(LNFP1)	0.94	(0.87 - 1)	0.063

Minimal Model + DFLNT	OR	95% CI	Pr(W)
(Intercept)	0.14	(0.0362 - 0.237)	P<0.001
$\sqrt{DPP}$	0.99	(0.969 - 1.01)	0.17
standardized(DSLNT)	0.84	(0.795 - 0.883)	P<0.001
standardized(DFLNT)	1.12	(0.986 - 1.26)	0.062

Final Model	OR	95% CI	Pr(W)
(Intercept)	0.14	(0.0374 - 0.25)	P<0.001
$\sqrt{DPP}$	0.97	(0.946 - 0.992)	0.0089
standardized(DSLNT)	0.84	(0.793 - 0.883)	P<0.001
standardized(LNFP1)	0.91	(0.842 - 0.97)	0.006
standardized(DFLNT)	1.14	(1.01 - 1.28)	0.022

Complete Model	OR	95% CI	Pr(W)
(Intercept)	0.14	(0.0381 - 0.252)	P<0.001
$\sqrt{DPP}$	0.97	(0.943 - 0.991)	0.0071
standardized(DSLNT)	0.84	(0.794 - 0.885)	P<0.001
standardized(LNFP1)	0.91	(0.839 - 0.97)	0.0063
standardized(DFLNT)	1.15	(1.01 - 1.28)	0.024
standardized(LNFP3)	0.99	(0.956 - 1.02)	0.55

Table S1. Multivariate models considered in the analysis. DSLNT, LNFP1 and DFLNT show significant (p<0.05) contributions to NEC onset when combined in a final GEE model clustered by subject while considering days postpartum (DPP). The odds ratio (OR) is the exponentiated coefficient and the 95% CI describes the range of possible OR. The p-value was calculated from the Wald statistic of each coefficient.

A

Model	qLik	QIC	Δ_{IC}	ω
$t + S + N_1 + F + N_3$	-187.34	186.59	0.00	0.56
$t + S + N_1 + F$	-187.48	187.49	0.90	0.36
$t + S + F + N_3$	-190.53	190.56	3.97	0.08
$t + S + N_1 + N_3$	-194.70	234.95	48.36	0.00
$t + N_1 + F + N_3$	-201.48	365.69	179.10	0.00

B

Model	qLik	QIC	Δ_{IC}	ω
$t + S + N_1 + F$	-187.48	187.49	0.00	0.86
$t + S + F$	-190.59	191.07	3.58	0.14
$t + S + N_1$	-194.71	238.28	50.79	0.00
$t + N_1 + F$	-202.28	384.02	196.53	0.00

C

Model	qLik	QIC	Δ_{IC}	ω
$t + S + F$	-190.59	191.07	0.00	1.00
$t + S$	-196.31	238.23	47.16	0.00
$t + F$	-205.48	403.93	212.86	0.00
t	-208.01	419.61	228.55	0.00
1	-208.34	427.20	236.13	0.00

Table S2. Overview of multivariate model selection by backward selection. Panels A, B and C enumerate the first, second and third iterations of model comparisons considered during the backwards elimination used to construct the final model: t+S+N1+F. Panel A details the quasi-Likelihood, Quasi-AIC, Change in Q-AIC (Δ_{IC}) and Cumulative Akaike Weight (ω) when comparing the complete model to three HMO models (excluding time, t); the best three HMO model is the final model. Panel B shows these statistics comparing the best three HMO model to candidate two HMO models. Panel C shows these comparisons between the best two HMO model, candidate single HMO models and the null models. HMOs are abbreviated: DSLNT (S), LNFP1 (N1), DFLNT (F), LNFP3 (N3).

3. Supplemental Methods

3.1 Case Descriptions

A003 (Bell stage 2)
This was a male infant born at 29+2 week twin diamniotic monochorionic gestation with a birth weight (BW) of 1,350 g from an *in vitro* fertilization pregnancy born to a 37 year-old Asian mother who received betamethasone. Maternal serology was unremarkable. Delivery was by C-section. Apgars were 5 and 8 at 1 and 5 minutes. Feeds were started on day of life (DOL) 4 with feeds advancing by protocol until on DOL 16, two days after fortifier was added, he had blood in his stool. He was otherwise well but kept *non per os* (NPO) for 8 days for NEC. The white blood count (WBC) was 12,000/mm3 and C-reactive protein (CRP) was <0.1 mg/dL. He was diagnosed with urinary tract infections growing *Enterococcus faecalis* and was treated with vancomycin and meropenem then switched to gentamicin based on sensitivities. Radiographs had RLQ and L flank pneumatosis that persisted on subsequent radiographs before resolution. Despite the pneumatosis, the abdominal signs were benign. He was diagnosed with Stage 2 NEC and had no NEC sequelae up to discharge.

A029 (Bell stage 2)
This was a male infant born at 27+3 weeks twin B diamniotic dichorionic gestation with a BW of 1,069 g born to a 30 year-old Caucasian mother. Pregnancy was complicated by preterm labor and premature rupture of membranes (PROM). Delivery was by C-section. Feeds were started on DOL 3 and reached full feeds on DOL 18. On DOL 22 he was made NPO with a presentation of abdominal distension, frank blood in stools and apnea. His WBC was 4,000/mm3 and CRP 17.7 mg/dL. He was treated with vancomycin and meropenem and then found to have a positive blood culture for *Staphylococcus epidermidis*. He was refed after 12 days. Radiographs showed pneumatosis that was short-lived. He had Stage 2 NEC. No further sequelae came after feedings resumed.

A066 (Bell stage 1)
This was a male infant born at 26+6 weeks with a BW of 961 g born to a 31 year-old Hispanic mother. Maternal serologies were unremarkable. Pregnancy was complicated by intrahepatic cholestasis and then PROM but mother was given betamethasone. Delivery was by C-section. Apgars were 3,6, and 7 at 1, 5, and 10 minutes. The infant was intubated at birth and given surfactant. He was extubated on DOL 3. Feeds were advancing up to 50% on DOL 19 before abdominal distension and blood in stools were noted. The WBC was 12,000/mm3 and CRP was 0.1 mg/dL. Blood cultures were negative. The subject was treated with vancomycin, cefotaxime and metronidazole for 10 days before feeds resumed. Radiographs were never positive for pneumatosis or free air. He had Stage 1 NEC and had no NEC sequelae up to time of discharge and at two-year follow-up.

B032 (Bell stage 3)
This was a female infant born at 31+0 weeks twin Di-Di gestation with a BW of 1,160 g to a 26 year-old G2P1 African American mother. Pregnancy was complicated by twin gestation, prolonged premature rupture of membranes (PPROM), and breech presentation. Antenatal betamethasone and ampicillin were given. Delivery was by vaginal route. She presented at birth limp and positive pressure ventilation (PPV) was initiated within 1 min of life but heart rate (HR) remained at 60 bpm. Infant was intubated and given 10 seconds of chest compressions before HR rose above 80. Apgars were 1 and 7 for 1 and 5 minutes. She received continuous positive airway pressure (CPAP) and nasal cannula ventilatory support until ten days after birth, and was then placed back on ventilatory support at 15 days of life due to apnea. Serial complete blood counts (CBCs), procalcitonin, and CRP were negative in the first few days of life. Total parental nutrition (TPN) was started at admission and continued for 13 days. Feeds were started DOL 2, liquid human milk fortifier added DOL 10 and then advanced to full feeds by DOL 13. On DOL 15 she developed feeding intolerance with a small emesis and gastric residuals. Radiographs showed pneumatosis in RUQ and LUQ with possible portal venous air. A laparotomy was performed resulting in the resection of half of the bowel with remaining half left with questionable viability. She continued to deteriorate after surgery and died on DOL 17.

C005 (Bell stage 3)
This was a male infant born at 26+3 weeks with a BW of 1,040 g to a 34 year-old G6P0 African American woman. Maternal history included polycystic ovarian syndrome (PCOS), insulin dependent gestational diabetes mellitus (GDM), and PPROM. Mother received antibiotics for latency and prenatal steroids. Delivery was by emergent C-section for maternal chorioamnionitis. Infant required positive pressure ventilation and intubation in the delivery room. Apgar scores were 5 and 8 at 1 and 5 minutes. Infant remained intubated and received 3 doses of surfactant. He was treated with indomethacin for a patent ductus arteriosus (PDA) that was functionally closed by DOL 9. He received a total of 6 days of empiric IV antibiotics before the development of NEC. Trophic enteral feedings were started on DOL 3 with maternal breast milk and were gradually advanced to 136 ml/kg/d with breast milk, until DOL 25 when feeds were changed to 20 kcal/oz preterm formula due to low milk supply. On DOL 26, there were recurrent residuals of digested formula. Abdominal radiographs showed dilated loops of bowel, pneumatosis and portal venous gas. Infant's abdomen was distended, firm, erythematous in the RLQ, with hypoactive bowel sounds. Infant was placed on bowel rest and treated with 21 days of IV antibiotic for a positive blood culture for E. coli and CSF pleocytosis. Infant improved his clinical condition with resolution of pneumatosis and bowel dilation within 3 days of starting therapy. He did not tolerate resumption of enteral feedings after 14 days of bowel rest with recurrent bilious residuals and emesis. Persistent feeding intolerance prompted two UGIs with the second one showing a LUQ dilated loop of bowel. He underwent exploratory laparotomy that revealed a large, dense, inflammatory mass in proximal small bowel causing partial small bowel obstruction, with pathologic examination of resected small bowel demonstrating adventitial hemorrhage, serositis, and abscess formation with necrosis. After bowel resection and jejuno-jejunostomy, he tolerated full enteral feedings with spontaneous bowel movements and was discharged home at 43+2 weeks.

C024 (Bell stage 2)
This was a female infant born at 27+1 weeks with a BW of 830 g to a 25 year-old G2P0 African American woman. Maternal history was complicated by severe preeclampsia, HELLP (hemolysis, elevated liver enzymes low platelet) syndrome and Group-B-streptococcus (GBS) positivity. Mother received antenatal steroids, antihypertensive medications, magnesium sulfate and clindamycin. Delivery was by C-section. Infant cried spontaneously and was given CPAP. Apgars were 8 and 9 at 1 and 5 minutes of life. She remained on CPAP until she was intubated for respiratory failure associated with NEC. She was treated for a PDA with indomethacin that closed by DOL 17. She received a total of 5 days of empiric intravenous antibiotics before development of NEC. Trophic enteral feedings were started on DOL 3 with mom's own milk, however, preterm formula was introduced on DOL 14 when mother's milk supply was low. On DOL 24, she received a blood transfusion for symptomatic anemia. Then on DOL 26, she presented with abdominal distension, hypoactive bowel sounds, apnea, lethargy and hypothermia. Abdominal radiographs showed mildly dilated loops of bowel and pneumatosis. She was managed with bowel rest and IV antibiotics for 14 days. Blood culture was negative but the CRP was elevated. She tolerated resumption of enteral feedings of preterm formula and was discharged home at 36+5 weeks.

C027 (Bell stage 2)
This was a female infant born at 29+5 weeks with a BW of 1,370 g to a 19 year-old African American G1P0 woman by C-section with prolonged preterm premature rupture of membranes. Mother received prenatal steroids, antibiotics, and magnesium sulfate. Apgars of 6 and 9 at 1 and 5 minutes of life. She transitioned from CPAP to nasal cannula oxygen by DOL 2. Trophic enteral feedings were started on DOL 2 with mother's own milk and preterm formula was introduced on DOL 12 when milk supply was low. She received a total of 4 days of empiric intravenous antibiotics prior to the development of NEC. She tolerated full enteral feeds until DOL 17 when she presented with residuals, emesis, bradycardia, abdominal distension and bloody stools. Abdominal radiographs showed dilated loops of bowel and pneumatosis. Her abdomen was distended with bowel sounds present. She was placed on bowel rest and intravenous antibiotics for 9 days. Blood culture was negative with

an elevated CRP. She tolerated resumption of enteral feedings with preterm formula and was discharged home at 36+5 weeks postmenstrual age (PMA).

D015 (Bell stage 1)
This was a male infant born at 26+0 weeks to a 34 year-old G2P1 mother by spontaneous labor and spontaneous vaginal delivery (SVD). Maternal serologies were unremarkable aside from GBS positivity. Pregnancy was complicated by early vaginal bleeding from 16 weeks with concern for chronic abruption. Maternal history included hypothyroidism on thyroid replacement and a history of asthma. PROM occurred at 25+3 weeks. Decreased amniotic fluid was noted at 25 6/7 weeks along with poor biophysical profiles of of 4/8 and then 6/8 at 26 weeks. Antenatal steroids were given as well as peripartum antibiotics. Apgars were 5, 6, 8 at 1, 5, 10 minutes. He was intubated at 20 minutes of life, given surfactant and placed on high frequency oscillation ventilation (HFOV). On DOL 6, he developed abdominal distention, feeding intolerance and bowel wall thickening on abdominal x-ray. He was treated medically with bowel rest and antibiotics for 10 days and went home after a prolonged neonatal course at 43+3 weeks.

D024 (Bell stage 2)
This was a female infant born at 26+2 weeks to a 32 year-old G2P1 mother. Mother had unremarkable serology. This was a twin pregnancy with a vanishing twin. Maternal history included a two-year history of infertility and PCOS with this pregnancy assisted by intracytoplasmic sperm injection (ICSI) and intraventricular hemorrhage (IVH). PPROM occurred at 25+3 weeks and mother was given penicillin and then erythromycin for GBS positivity. She received betamethasone 4-5 days prior to delivery. Emergency C-section was done for possible cord prolapse with breech presentation but was found to have legs protruding. Apgars were 5, 7, 9 at 1, 5, 10 minutes. The infant was given positive pressure and placed on CPAP and room air. She remained on CPAP until 2 weeks of age. She had a PDA that required two courses of indomethacin to close. She was started on breast milk on day 3 of life and reached full feeds by day 13. Feeding intolerance occurred on day 14 concurrent with fortifier introduction, presence of the PDA and possible coagulase negative Staphylococcus (CONS) sepsis. Small trophic feeds were started after two days followed by advancement but increasing abdominal girth and free air on abdominal x-ray were noted. She was started on antibiotics and transferred to Children's Hospital where she underwent surgical management.

E002 (Bell stage 3)
This was a male infant born at 25+0 weeks to a 17 year-old G1P0 mother by vaginal delivery. Maternal history included marijuana use early in the first trimester, but no other drugs of abuse. She did not receive antenatal steroids. He received prophylactic indomethacin for IVH prevention. In his first week of life, he required extensive delivery resuscitation, pressors, and extreme volumes of blood products. On DOL 23 he developed increased abdominal distention and an ileus. Abdominal US revealed contaminated ascites consistent with a perforation. His first laparotomy revealed an intestinal perforation at the site of a Meckel's diverticulum with subsequent ileostomy. At 2 months of age, he developed symptoms of Bell Stage 3 NEC with bloody stool, pneumatosis, and feeding intolerance. This diagnosis was confirmed on laparotomy, along with stricture from early NEC that was not obvious during his first laparotomy. He stayed in the hospital for 148 days. He demonstrated severe functional short bowel syndrome, despite adequate small bowel length. Since his NICU discharge, he has required 3 subsequent hospitalizations and 3 ER visits.

3.2 Statistical Analysis and Classification

3.2.1 Multivariate Model Selection by Backward Elimination

Multivariate models were built to include factors that marginally contributed to the onset of NEC based on covariate and HMO pre-screening ($Pr(W) < 0.2$). Backward elimination (BE) was used to select a final multivariate model. BE removes one variable at a time starting from a model containing all variables passing the pre-screening. In each iteration of BE, the variable with the smallest contribution to the likelihood of the model is removed; contribution to likelihood is assessed by removing the variable an comparing the likelihoods of the original and new model. BE terminates when no move variables can be removed from the model without decreasing the likelihood. The performance of notable models examined in the backward elimination is shown in **Table S1**. The iterations of the backward propagation are detailed in **Table S2** and the accompanying text.

3.2.2 Odds ratio calculation from logistic regression models

All GEE models and regressions were logistic therefore analysis involved the calculation of odds, odds ratios, confidence intervals and significance. The odds ratio (OR) are the exponentiated coefficients (β) of each variable, $OR = e^{\beta}$. Confidence intervals (CI) for the odds ratios were calculated using the delta method, a function of standard error (SE) of the coefficient of each variable, $CI = e^{\beta} +/- 1.96\ e^{\beta}\ SE$. The Wald statistic tests if a coefficient diverges from zero along a normal distribution to describe the probability that the coefficient is non-zero.

3.2.3 Constructing the decision tree

Decision trees were constructed using J48, a Java implementation in Weka [Witten & Frank, 2005] of the C4.5 decision tree-generating algorithm. C4.5 selects recursive partitions to the data by optimizing normalized information gain [Quinlan, 1993]. We used RWeka package in R, a wrapper package for the Weka library [Hornik, 2009]. While training our decision trees, the minimum number of observations per leaf node was set to 20 to avoid over-fitting. J48 decision trees were learned from the odds produced by our multivariate models to exemplify their prescriptive information. These decision trees exemplify the prescriptive capabilities of our observations. Given the smaller cohort in this study, the decision trees are provided as qualitative analysis. Future work will be able to evaluate and provide a more quantitative classifier that addresses longitudinal sampling in a larger cohort.

3.2.4 Consecutive Day Measurements: Cumulative odds of NEC over multiple days

We also applied the cumulative odds across multiple consecutive samples in decision tree training. Cumulative odds (CO) were calculated as the one minus the joint probability that several consecutive samples (s_i) would not be associated with NEC (N). The cumulative odds demonstrated enhanced separation between milk associated with NEC cases vs. controls (**Figure S3D** and **Figure S4B**). The cumulative odds are used to aggregate over multiple consecutive observations and thereby improve classification.

$$CO_{n:m} = 1 - \prod_{i=n}^{m} \Pr\left(s_i \notin N\right)$$

3.2.5 Odds-Based Classification: Resampling consious performance assessment

Bootstrapping was used to estimate of the performance of our models at the task of high/low risk milk classification. Bootstrapping estimates tend to be conservative and provide a lower variance estimate, which was preferred here due to the small number of cases currently available. To avoid subject re-sampling bias we chose at most 1 sample per subject in each performance-assessing iteration. In each of 100 iterations, 2/3 of subjects were randomly selected. From each subject selected, only one sample was used per iteration. Sampling 2/3 of the available subjects excludes multiple subjects while still maintaining a representative sample (n=40 samples/iteration). This sampling proportion allowed for good assessment of the performance variance. Limiting each subject to one sample per iteration ensures no subject was over-represented in any of the performance assessments.

Code Repository: Code can be found at https://github.com/bkellman/NEC_HMO

4. Supplemental Results

4.1 Multivariate Model Selection by Backward Elimination

Multivariate models were selected using Backwards Elimination (BE) to search combinations of the four HMO (DSLNT, LNFP1, DFLNT, LNFP3), that passed (Pr(W) < 0.2) the univariate screening. We sought to minimize QIC and maximize parsimony. The first BE iteration (**Table S3A**) showed a small increase in QIC from the final model (t+S+N1+F+N3) to a smaller, more parsimonious model (t+S+N1+F). The next BE iteration (**Table S3B**) found a larger increase (Δ_{IC} > 3.58) in QIC in all candidate bivariate models. The failure to find a more parsimonious model with lower QIC concluded Backward Elimination and suggests that t + S + N1 + F (the final model) is the most parsimonious and descriptive model. We note that Δ_{IC}(t+S; t+S+N1+F) = 50.74 (**Table S3A and S3C**). This is not a negligible Δ_{IC}, but due to the substantial increase in parsimony offered by the univariate model over the trivariate model we consider t + S (the minimal model) to be a notable model as well.

The formulas below compute the odds that a child will develop NEC based on HMO concentrations in milk consumed on the i^{th} day of life. The odds that a child will develop NEC can be computed based on HMO concentration. This is done by using the exponentiation of the coefficients parameterized by the time (t) and the z-statistic standardized levels of DSLNT (s), LNFP1 (n) and DFLNT (f) at observation *i*. The first equation describes the Minimal Model including only time and DSLNT and the second equation describes the Final Multivariate Model including time, DSLNT, LNFP1 and DFLNT.

Minimal Model

$$Odds_i = exp\left(\beta_0 + \beta_t\sqrt{t_i} + \beta_s\frac{s_i - \mu_s}{\sigma_s}\right)$$

$$Odds_i = exp\left(-1.9 + -0.023\sqrt{t_i} + -0.15\left[\frac{s_i - 350}{160}\right]\right)$$

s = DSLNT (μg/mL) $\quad t$ = DPP (days)
n = LNFP1 (μg/mL) $\quad i$ = i-th observation
f = DFLNT (μg/mL)

Final Model

$$Odds_i = exp\left(\beta_0 + \beta_t\sqrt{t_i} + \beta_s\frac{s_i - \mu_s}{\sigma_s} + \beta_n\frac{n_i - \mu_n}{\sigma_n} + \beta_f\frac{f_i - \mu_f}{\sigma_f}\right)$$

$$Odds_i = exp\left(-1.94 + -0.0317\sqrt{t_i} + -0.176\left[\frac{s_i - 350}{160}\right] + -0.099\left[\frac{n_i - 1250}{985}\right] + 0.134\left[\frac{f_i - 720}{560}\right]\right)$$

4.2 Odds-Based Classification: Classification performance is exceptional when aggregating multiple samples and controling for resampling bias.

The decision trees in **Figure S3** provide a nice visualization of the potential discriminative power of the odds produced by the models above. With both single and multiple day considerations of odds, the apparent classification is excellent. For a rigorous assessment of the discriminative power of the minimal and final GEE model generated odds, we directly examined the performance with the rolling threshold of a Receiver Operating Characteristic (ROC) curve.

By evaluating the classification performance of the GEE generated odds we can focus on mitigating the resampling bias only in the performance evaluation since the learning is already done. As discussed in the supplemental methods, resampling bias was addressed using a bootstrapping approach to include no more that 1 sample per subject for each bootstrap iteration. Performance of the final and minimal model calculated odds were found to be exceptional, as quantified by the median AUC and median MCC.

In all assessments, the minimal model outperformed the final model. Considering that the final model utilizes more information, the underperformance of the final model may be due to over-fitting while the minimal model is more generalizable. The median AUC is typically above .95 for all models. The minimal model AUC is close to 1 considering only 2 consecutive milk samples while the final model converged at approximately AUC=.96 considering 4 consecutive milk samples. As expected, the AUC is inflated, relative to the MCC, by the class imbalanced of the data. The MCC follows a similar pattern to the AUC. The minimal model converges at

MCC=1 considering 6 consecutive milk samples. The minimal model outperforms the final model which converges at MCC=.8 considering 4 consecutive milk samples. As expected AUC and MCC show correlation. Additionally, these metrics converge and improve as more consecutive samples are included. From this examination, the minimal model considering 2 consecutive samples appears to be the most economical model, with AUC=.97 and MCC=.8. The minimal model considering 4 consecutive samples appears to have the highest performance before effective convergence, AUC=.99 and MCC=.9. In most assessments, the optimal cutoff was approximately odds=.16. However, due to the limited size of this study, this cutoff is provided as a proof of principle and is not yet appropriate for broader use. Larger follow up studies will be required to obtain more robust estimates and further mitigate potential contributions from over-fitting. Furthermore, we believe that the minimal model is the most generalizable instrument produced by this investigation. This is supported by the consistent higher median AUC and MCC of the minimal model generated odds over the final model generated odds in this bootstrapping validation. One explanation for the lower performance of the odds generated by the final model is that it is too specific for generalization. This means the final model is more likely to contain information pertinent to these specific instances of NEC while the minimal model is more likely to describe the broader population of NEC cases. More specifically, DFLNT and LNFP1 are more likely to be associated with specific instances of NEC induction while the depletion of DSLNT is a global feature of the disease.

5. Supplementary Reference

Witten IH and Frank E. Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco. 2005

Hornik K, Buchta C, Zeileis A. Open-Source Machine Learning: R Meets Weka. Computational Statistics 2009;24(2):225-232.

Quinlan JR. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.