An Algorithm to Identify Less Invasive Surfactant Administration Using a Real-World Database of Preterm Infants

RWD25

Xuezheng Sun[1], Annie N. Simpson[2], Aditi Lahiri[3], Sanjida Mowla[1], Dana Edelman[3], Shelby Corman[2], Daniel Fuentes[1], Dalibor Kurepa[4], Michael W. Kuzniewicz[3,5,6]

[1]Chiesi USA Inc, Cary, North Carolina; [2]Precision AQ, Bethesda, Maryland; [3]Division of Research, Kaiser Permanente Northern California, Pleasanton, California; [4]Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, Ohio; [5]Department of Pediatrics, University of California, San Francisco, California; [6]The Permanente Medical Group, Oakland, California

BACKGROUND

- Surfactant replacement therapy is central to the management of respiratory distress syndrome (RDS) in preterm infants
- Less invasive surfactant administration (LISA) has been increasingly adopted due to improved neonatal outcomes
- However, there are no standardized methods to identify LISA in large real-world data (RWD)
- This study aimed to develop an algorithm to identify LISA procedures using administrative data

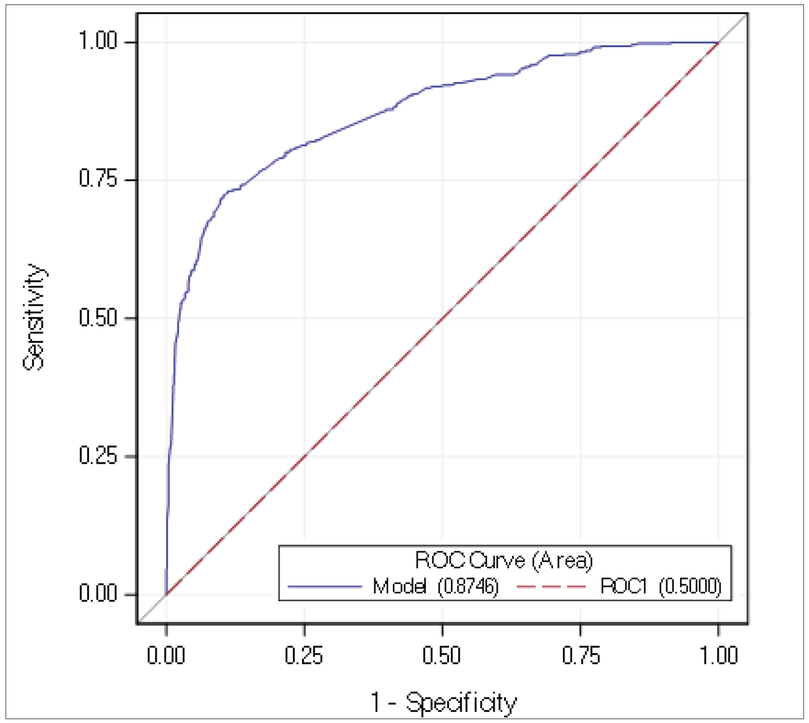
METHODS

- **Design and Data Source:** Retrospective cohort study using electronic health record data from Kaiser Permanente Northern California, 2019–2023
- **Population:** liveborn preterm infants (<37 weeks) who received surfactant during birth hospitalization
- **Algorithm Development and Validation**
  - Chart review served as gold standard
  - 82 candidate variables selected from administrative data between birth and date of first surfactant administration
  - Infants randomly split into a training set (70%; n=884) and testing set (30%; n=379)
  - A least absolute shrinkage and selection operator (LASSO) regression was used for variable selection and model fitting
  - Model discrimination was evaluated using area under the receiver operating characteristic (AUROC)
  - Algorithm performance was validated using a combined sample of the testing set and a 2024 birth cohort (n=622) overall and by gestational age (GA)
  - We evaluated sensitivity (Sn), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV)

RESULTS

- Among 1,263 preterm infants who received surfactant, 462 (36.6%) received surfactant via LISA and 801 (63.4%) received surfactant via invasive modalities (ETT or INSURE)
- Within the training cohort (n=884), infants who received LISA had higher gestational age and greater use of non-invasive respiratory support than those treated with non-LISA methods (**Table 1**) and similar patterns were observed in the testing set
- The LASSO-based model selected 21 variables predictive of LISA methods based on the training set, which included indicators of gestational age and birth weight, respiratory support, clinical factors like medication use and procedure timing
- The model demonstrated strong discrimination for identifying LISA (AUROC=0.87; **Figure 1**)
- Using the maximum specificity cut-point (predicted probability ≥0.79), the model achieved Sn=43.9%, Sp=96.8%, and PPV=90.0% when evaluated in the combined testing set and 2024 birth cohort (**Table 2**)
- Algorithm performance was consistent across GA subgroups

Figure 1. Area Under the Receiver Operating Characteristic



CONCLUSION

- We developed and validated a machine-learning algorithm that accurately identifies surfactant administered via LISA among preterm infants using administrative data
- This algorithm supports large-scale studies of effectiveness, safety and resource use across surfactant methods
- Future research should validate the algorithm across diverse healthcare systems

Table 1. Demographic characteristics for training cohort

Characteristic	Training Cohort (N = 884) Total (n, %) N=884	Non-LISA (n, %) N = 563	LISA (n, %) N = 321	p-value
Sex				>0.9
Female	342 (38.7)	217 (38.5)	125 (38.9)	
Male	542 (61.3)	346 (61.5)	196 (61.1)	
Type of delivery				0.2
Cesarean	662 (74.9)	429 (76.2)	233 (72.6)	
Vaginal	222 (25.1)	134 (23.8)	88 (27.4)	
Gestational age (weeks)				<0.001
23-27	309 (35.0)	249 (44.2)	60 (18.7)	
28-31	295 (33.4)	158 (28.1)	137 (42.7)	
32-36	280 (31.7)	156 (27.7)	124 (38.6)	
Final disposition				<0.001
Alive/Home	785 (88.8)	475 (84.4)	310 (96.6)	
Death	99 (11.2)	88 (15.6)	11 (3.4)	
Apgar 5 score < 7	241 (27.3)	204 (36.2)	37 (11.5)	<0.001
Highest FiO2 pre-surfactant				
Median (IQR)	-	51 (40, 100)	44 (35, 60)	<0.001
Respiratory support pre-surfactant				<0.001
CVENT	172 (19.5)	172 (30.6)	0 (0.0)	<0.001
HFNC	2 (0.2)	0 (0.0)	2 (0.6)	
HFV	10 (1.1)	10 (1.8)	0 (0.0)	
NCPAP	563 (63.7)	263 (46.7)	300 (93.5)	
NIPPV	22 (2.5)	7 (1.2)	15 (4.7)	

Table 2. Algorithm validation in testing + 2024 birth cohort overall and by gestational age using maximized specificity cut point

Statistic	Overall	GA <34 weeks	GA ≥34 weeks
Number of infants	622	494	128
Sensitivity	43.9 (37.6–50.4)	39.8 (32.8–47.1)	58.2 (44.1–71.4)
Specificity	96.8 (94.5–98.3)	97.0 (94.4–98.6)	95.9 (88.5–99.1)
Positive predictive value	90.0 (83.5–94.1)	89.4 (81.3–94.3)	91.4 (77.5–97.1)
Negative predictive value	72.5 (70.2–74.7)	71.9 (69.5–74.2)	75.3 (68.6–80.7)
Accuracy	75.9 (72.3–79.2)	74.9 (70.8–78.7)	79.7 (71.7–86.3)
Positive likelihood ratio	13.8	13.4	14.2
Negative likelihood ratio	0.58	0.62	0.44
Estimated disease prevalence	39.6 (35.7–43.5)	38.7 (34.4–43.1)	42.9 (34.3–52.0)