

A Machine Learning Decision Support Tool Optimizes Whole Genome Sequencing Utilization in a Neonatal Intensive Care Unit

Edwin F. Juarez, Bennet Peterson, Erica Sanford Kobayashi, Sheldon Gilmer, Laura E. Tobin, Brandan Schultz, Jerica Lenberg, Jeanne Carroll, Shiyu Bai-Tong, Nathaly M. Sweeney, Curtis Beebe, Lawrence Stewart, Lauren Olsen, Julie Reinke, Elizabeth A. Kiernan, Rebecca Reimers, Kristen Wigby, Chris Tackaberry, Mark Yandell, Charlotte Hobbs, Matthew N. Bainbridge

The Mendelian Phenotype Search Engine (MPSE), a clinical decision support tool using Natural Language Processing and Machine Learning, helped neonatologists expedite decisions to whole genome sequencing (WGS) to diagnose patients in the Neonatal Intensive Care Unit. After the MPSE was introduced, utilization of WGS increased, time to ordering WGS decreased, and WGS diagnostic yield increased.

Main Body (Results & Discussion)

Genetic disorders are a leading cause of death and disability for infants admitted to the Neonatal Intensive Care Unit (NICU)¹. Rapidly diagnosing the underlying cause of critical illness and initiating targeted treatment are of paramount importance given the considerable morbidity and mortality associated with NICU admission¹⁻⁶. Rapid Precision Medicine utilizing Whole Genome Sequencing (WGS) can help identify patients with genetic disease and thus facilitate care tailored to the individual⁵⁻¹³. However, due to economic considerations and clinician familiarity with WGS, deciding which patients should receive WGS in the NICU can be challenging^{6,12-15}. We hypothesized that an automated clinical decision support tool utilizing machine learning to continually reassess the appropriateness of rapid WGS (rWGS) could assist neonatologists with patient prioritization for rWGS.

A single-group study was designed to compare findings before and after the implementation of a clinical decision support tool. The clinical support tool, Mendelian Phenotype Search Engine (MPSE), was designed to utilize Machine Learning (ML) to leverage the Human Phenotype Ontology (HPO) terms to calculate scores for prioritizing patients for WGS¹⁶. The HPO provides a hierarchical representation of the clinical abnormalities observed in human disease, and thereby facilitates computational analysis of patient phenotypes^{17,18}. Natural Language Processing (NLP) tools can identify HPO terms found in Electronic Medical Record (EMR) notes that describe patient phenotypes related to Mendelian disease, allowing for analysis via machine learning (ML)¹⁹⁻²².

We developed a software pipeline to automatically extract HPO terms from unstructured physician notes embedded within the EMR of patients recently admitted to the NICU. These HPO terms were used by the MPSE to compute a prioritization score that reflects the similarity of newly admitted NICU patients to observed phenotypes of patients within the NICU who previously received WGS¹⁶.

We performed this study in two phases. The objective of Phase 1 (the pre-implementation phase), was to collect baseline data on the number of babies nominated for WGS, the time to nomination, and the diagnostic yield of WGS. During this phase, MPSE scores for each patient were computed daily but were not provided to the clinical team. During Phase 2 (the implementation phase), the attending neonatologists were provided with a daily report containing MPSE scores for each NICU patient on the census. This MPSE report was presented to the neonatologists as an additional piece of information to be taken into consideration when deciding which patients should receive WGS.

Three primary outcomes were measured: 1) number and proportion of babies nominated for WGS; 2) time from admission to nomination for WGS, and; 3) diagnostic yield of WGS.

In total, 118 patients were nominated for rWGS; 27 in Phase 1 (14 weeks, 1.9 nominations/week) and 91 in Phase 2 (38 weeks, 2.4 nominations/week) (Mann-Whitney-Wilcoxon two-sided test, $p=0.35$); in both phases 13% of the eligible patients in the NICU were deemed by the attending physician to benefit from rWGS. Of the nominated patients, 98 patients (83%) were enrolled and underwent WGS (reasons for decline listed in Supplementary Table 1); 25 from Phase 1 (1.8 enrollments per week) and 73 from Phase 2 (1.92 enrollments per week). Enrollment rates were not significantly different between Phase 1 and Phase 2 (Mann-Whitney-Wilcoxon two-sided test, $p=0.63$).

Of the 99 sequenced patients, 29 received a molecular diagnosis (Supplementary Table 1), with 6 diagnoses in Phase 1 (24% diagnostic yield) and 23 in Phase 2 (32% diagnostic yield) (Fisher's Exact test, $p=0.61$). Each of the diagnosed patients had at least one genetic variant consistent with their phenotype classified as pathogenic or likely pathogenic according to The American College of Medical Genetics and Genomics (ACMG) guidelines²³.

The median time from admission to nomination decreased from 48.0 hours in phase 1 to 39.1 hours in phase 2 (18.5% reduction; Mann-Whitney-Wilcoxon two-sided test, $p=0.10$). This is particularly noticeable at 72 hours post admission where, in phase 2, 82% of all nominations had taken place, vs only 53% in phase 1 (Cox's proportional hazard regression, $p=0.10$).

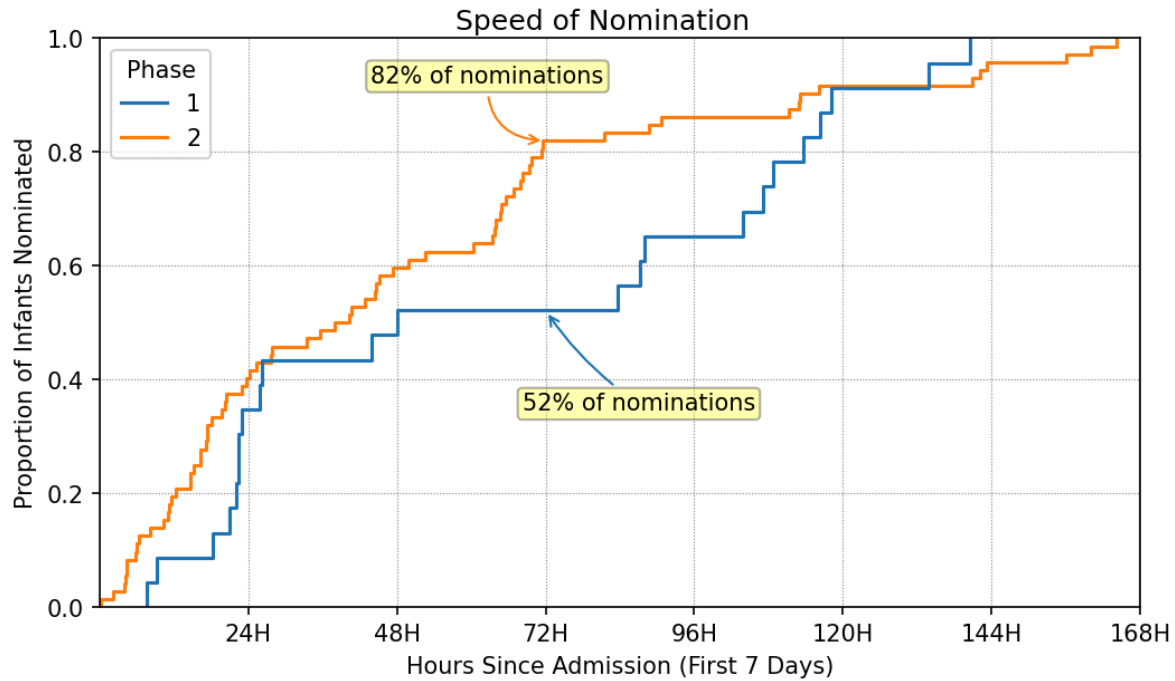


Figure 1: Speed of nomination curves showing the time-to-nomination for phase 1 (blue curve) and phase 2 (orange curve) patients nominated within the first 7 days of their NICU stay.

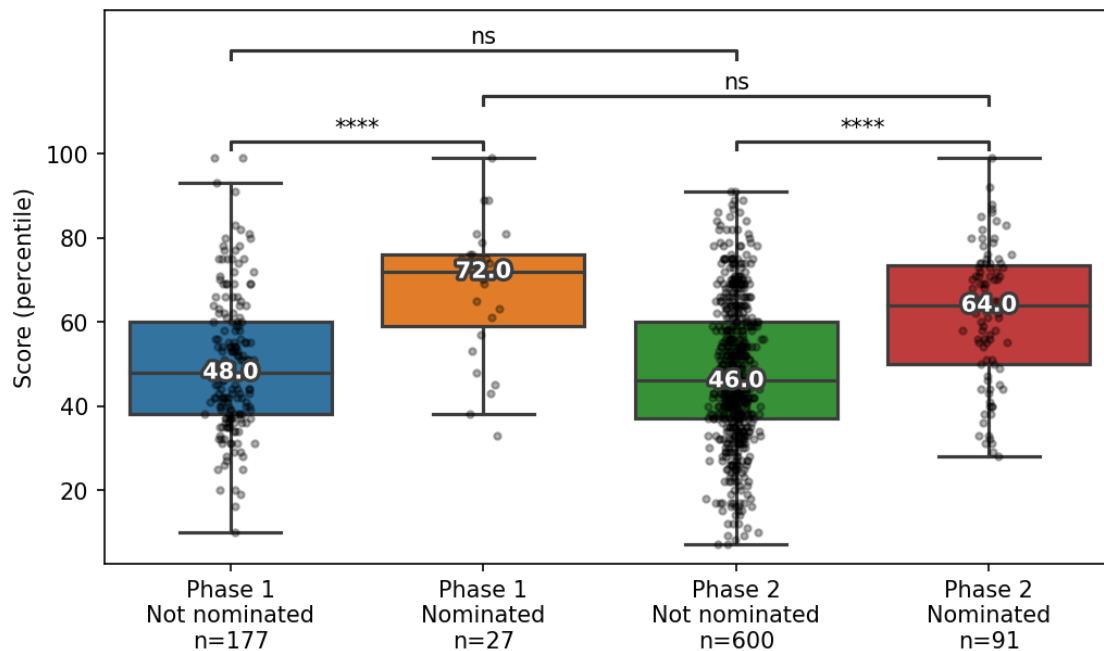


Figure 2. MPSE score percentiles for every patient admitted to the NICU during the duration of the study. For nominated patients, the MPSE score at the time of nomination is shown. For patients who were not nominated, the maximum MPSE score within the first seven days of their NICU admission is shown.

In both phases the MPSE scores for the nominated patients were significantly higher than the scores of the patients who were not nominated (Mann-Whitney-Wilcoxon two-sided test; Phase 1: $p=1.5 \times 10^{-4}$; Phase 2: $p=4.6 \times 10^{-17}$) and is consistent with our previous work that found that the MPSE scores of patients nominated for WGS were higher than those not nominated¹⁶.

Although the differences between the three primary outcomes were not statistically different between pre-implementation (Phase 1) and implementation (Phase 2) of MPSE, there was a trend towards improvement for all three primary outcomes (number and proportion of babies nominated for WGS, time from admission to nomination for WGS, and diagnostic yield of WGS) in Phase 2 when MPSE data were used to assist neonatologists' decisions to use rWGS. Specifically, observed promising results were found regarding nomination frequency, nomination speed, and diagnostic yield after utilizing MPSE for clinical decision support. After implementing MPSE, we saw a modest but important increase in both the speed of nomination (how soon after admission does nomination occur) and weekly rate of nominations. Importantly, increased nomination frequency and decreased elapsed time between admission and WGS nomination after introduction of MPSE were observed together with a modest increase in diagnostic yield (24% to 32%), suggesting that the increased frequency and speed of nomination did not degrade the yield of rWGS and may have improved it.

Limitations in this study include a small sample size, especially during pre-implementation and lack of long-term outcome data. These challenges should be addressed with future studies. Additionally, the established familiarity of the study site's NICU physicians with rWGS suggests MPSE might hold greater influence in settings where Rapid Precision Medicine has not been established. Further research is needed to confirm these preliminary findings and to assess generalizability between NICUs and clinical teams.

Although statistically significant differences were not observed, likely due to limitations in sample size, these findings hold promise for future research. This study contributes to the ongoing effort to inform the design and implementation of ML tools within healthcare environments. This study demonstrates MPSE's capability for integration into existing clinical workflows and indicates MPSE could be similarly employed at other healthcare systems.

These findings underscore the immediate impact that carefully applied clinical decision support tools harnessing NLP and machine learning can potentially have for clinicians in the intensive care unit with regards to efficiently and appropriately selecting patients for genomic sequencing.

Methods

Patient enrollment

This clinical prospective study was conducted in the Level IV NICU of Rady Children's Hospital in San Diego (RCHSD). Our study was implemented in 2 phases. In each phase, attending neonatologists nominated patients for WGS following broad inclusion criteria, which included any NICU patient within the first seven days of life who was suspected of having a genetic

disease or a patient with an abnormal response to therapy after the first seven days of admission to the NICU. The patient's family provided written informed consent, and whenever possible, parent samples were also collected.

Phase 1 lasted 14 weeks (July to October 2022) and Phase 2 lasted 38 weeks (October 2022 to July 2023).

MPSE Score Computation

The Mendelian phenotype Search Engine (MPSE) employs Human Phenotype Ontology (HPO) terms to determine the likelihood that a Mendelian condition underlies a patient's phenotype. MPSE employs a simple, well-established approach: a Naïve Bayes (NB) classifier that has previously been published in detail by our group¹⁶. Briefly, MPSE uses the differences in HPO term frequencies between a collection of cases and controls to score each patient by calculating NB the log-odds ratio.

HPO-based phenotype descriptions were generated for all patients in Phases 1 and 2 by NLP analysis of clinical notes using CLiX ENRICH (Clinithink, Alpharetta, GA). A pre-trained MPSE model was then used to calculate MPSE scores for each patient¹⁶. In this report, MPSE score percentiles are reported to simplify interpretation.

MPSE scores and percentiles for each patient in the NICU were computed automatically every three hours during the study period. Each score's percentile represents the position that a given score would have taken in the training cohort, thus percentiles can be compared to each other without the need to recalculate them with each new score added to the distribution.

Statistics

Statistics were computed in Python version 3.10.2 with SciPy version 1.8.0, statannotations version 0.5.0, and lifelines version 0.27.8.

Group statistics utilize the Mann-Whitney-Wilcoxon two-sided test, the Fisher's Exact test to compare proportions, and Cox's proportional hazard model to compare time to nomination. For all testing, $p < 0.05$ was considered statistically significant.

Data availability

De-identified data utilized in this paper is attached as supplementary material; including time from admission to nomination, MPSE score, WGS results for enrolled patients, and reasons for decline for patients who did not enroll.

Acknowledgments

This study was funded by the Conrad Prebys Foundation, R.R. is supported by NCATS grant number K12TR004410. We are grateful to the families who participated in this study. We are also thankful to the RCHSD NICU team for their collaboration and contributions to this study.

We acknowledge the assistance that Ricky (Hung) Nguyen provided to the study by delivering the daily reports and Shauna Briscoe for serving as a project manager.

Author contributions

C. H., M. Y., C. T., J. R. conceptualized this study design and supervised the implementation; E. J. coordinated the implementation of the software and report delivery, performed the data analyses, and wrote the first draft of the manuscript; M. B., M. Y., C. H., L. T., and C. T. revised the figures and data analyses; M. B., E. S. K., B. P., C. H., and L. T. revised earlier versions of the manuscript; E. S. K., R. R., K. W., J. C., S. B. T., and C. H. provided clinical expertise throughout the study and during manuscript revisions; M. B., C. T. and M. Y. provided supervision for data analysis and interpretations; E. J., S. G., C. B., D. D., C. T., M. Y., and C. H., designed and implemented the score report. J. L., B. S., E. K., L. O., E. S. K., R. R., K. W., and C. H. contributed to study enrollment; and J. L., B. S. returned results to patients' families. All authors reviewed and approved the final versions of this manuscript.

Competing Interests

C Tackaberry is a director and employee of Clinithink and also a shareholder. No other competing interests from any author to disclose.

References

1. Michel, M. C., Colaizy, T. T., Klein, J. M., Segar, J. L. & Bell, E. F. Causes and Circumstances of Death in a Neonatal Unit over 20 Years. *Pediatr. Res.* **83**, 829–833 (2018).
2. Trowbridge, A., Walter, J. K., McConathey, E., Morrison, W. & Feudtner, C. Modes of Death Within a Children's Hospital. *Pediatrics* **142**, e20174182 (2018).
3. Chow, S. *et al.* A Selected Review of the Mortality Rates of Neonatal Intensive Care Units. *Front. Public Health* **3**, (2015).
4. Burns, J. P., Sellers, D. E., Meyer, E. C., Lewis-Newby, M. & Truog, R. D. Epidemiology of Death in the Pediatric Intensive Care Unit at Five U.S. Teaching Hospitals. *Crit. Care Med.* **42**, 2101–2108 (2014).
5. NICUSeq Study Group *et al.* Effect of Whole-Genome Sequencing on the Clinical Management of Acutely Ill Infants With Suspected Genetic Disease: A Randomized Clinical Trial. *JAMA Pediatr.* **175**, 1218–1226 (2021).
6. Dimmock, D. P. *et al.* An RCT of Rapid Genomic Sequencing among Seriously Ill Infants Results in High Clinical Utility, Changes in Management, and Low Perceived Harm. *Am. J. Hum. Genet.* **107**, 942–952 (2020).
7. Clark, M. M. *et al.* Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* **11**, eaat6177 (2019).
8. Kingsmore, S. F. *et al.* A Randomized, Controlled Trial of the Analytic and Diagnostic Performance of Singleton and Trio, Rapid Genome and Exome Sequencing in Ill Infants. *Am. J. Hum. Genet.* **105**, 719–733 (2019).
9. Franck, L. S., Dimmock, D., Hobbs, C. & Kingsmore, S. F. Rapid whole-genome sequencing in critically ill children: shifting from unease to evidence, education, and equitable

- implementation. *J. Pediatr.* **238**, 343 (2021).
10. Franck, L. S. *et al.* Implementing Rapid Whole-Genome Sequencing in Critical Care: A Qualitative Study of Facilitators and Barriers to New Technology Adoption. *J. Pediatr.* **237**, 237-243.e2 (2021).
 11. Kingsmore, S. F. *et al.* Mortality in a neonate with molybdenum cofactor deficiency illustrates the need for a comprehensive rapid precision medicine system. *Mol. Case Stud.* **6**, a004705 (2020).
 12. James, K. N. *et al.* Partially automated whole-genome sequencing reanalysis of previously undiagnosed pediatric patients can efficiently yield new diagnoses. *Npj Genomic Med.* **5**, 1–8 (2020).
 13. Dimmock, D. *et al.* Project Baby Bear: Rapid precision care incorporating rWGS in 5 California children’s hospitals demonstrates improved clinical outcomes and reduced costs of care. *Am. J. Hum. Genet.* **108**, 1231–1238 (2021).
 14. Clark, M. M. *et al.* Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* **11**, eaat6177 (2019).
 15. Franck, L. S. *et al.* Healthcare Professionals’ Attitudes toward Rapid Whole Genome Sequencing in Pediatric Acute Care. *Children* **9**, 357 (2022).
 16. Peterson, B. *et al.* Automated prioritization of sick newborns for whole genome sequencing using clinical natural language processing and machine learning. *Genome Med.* **15**, 18 (2023).
 17. Gargano, M. A. *et al.* The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res.* **52**, D1333–D1346 (2024).
 18. Daniali, M. *et al.* Enriching representation learning using 53 million patient notes through human phenotype ontology embedding. *Artif. Intell. Med.* **139**, 102523 (2023).
 19. Havrilla, J. M. *et al.* PheNominal: an EHR-integrated web application for structured deep phenotyping at the point of care. *BMC Med. Inform. Decis. Mak.* **22**, 198 (2022).
 20. Bastarache, L. *et al.* Improving the phenotype risk score as a scalable approach to identifying patients with Mendelian disease. *J. Am. Med. Inform. Assoc. JAMIA* **26**, 1437–1447 (2019).
 21. Morley, T. J. *et al.* Phenotypic signatures in clinical data enable systematic identification of patients for genetic testing. *Nat. Med.* **27**, 1097–1104 (2021).
 22. Keles, E. & Bagci, U. The past, current, and future of neonatal intensive care units with artificial intelligence: a systematic review. *Npj Digit. Med.* **6**, 1–36 (2023).
 23. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).