

# A Clinical Outcomes Data Archive for a Comprehensive Fetal Diagnosis and Treatment Center

Thomas A. Reynolds<sup>a</sup> Matthew A. Goldshore<sup>a,b</sup> Sabrina Flohr<sup>a</sup>  
Sierra Land<sup>a</sup> Leny Mathew<sup>a</sup> Juliana S. Gebb<sup>a</sup> Edward R. Oliver<sup>c</sup>  
Natalie E. Rintoul<sup>d</sup> Anne M. Ades<sup>d</sup> Elizabeth E. Foglia<sup>d</sup>  
Catherine M. Avitabile<sup>e</sup> Howard B. Panitch<sup>f</sup> Gregory G. Heuer<sup>g</sup>  
Lori J. Howell<sup>a</sup> N. Scott Adzick<sup>a,b</sup> Holly L. Hedrick<sup>a,b</sup>

<sup>a</sup>The Richard D. Wood Jr. Center for Fetal Diagnosis & Treatment, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>b</sup>Division of Pediatric General, Thoracic, and Fetal Surgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>c</sup>Department of Radiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>d</sup>Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>e</sup>Division of Cardiology, Children Hospital of Philadelphia, Philadelphia, PA, USA; <sup>f</sup>Division of Pulmonary Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>g</sup>Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA

## Keywords

Data quality · Data management · Perinatal outcome · Electronic health record

## Abstract

**Introduction:** Data on near- and long-term clinical outcomes are critical for the care of all maternal-fetal patients presenting to a fetal center. This is especially important since physiologic and neurodevelopmental attributes do not manifest until later childhood when multilevel (e.g., individual, family, policy) factors have a direct influence on health outcomes. Electronic health records (EHRs) create opportunity for efficient data collection. However, documentation structures are not designed for acquisition of key attributes, and changes over time and between-clinician differences can affect resultant output. Therefore, EHR de-

rived datasets have limited ability to accurately characterize the clinical presentation and care trajectory of patients with congenital anomalies. In addition, in most systems, the fetus lacks a digital identity and requires relinking fetal attributes documented in the maternal chart to those from the pediatric EHR. This conundrum amplifies in the setting of multiple gestation, returning maternal patients, and pregnancies with fetal demise. Moreover, current data capture systems result in incomplete abstraction of variables that may confound, mediate, or moderate critical associations. Our objective was to develop and implement a prospective data capture platform to transform EHR data into an analytic-grade database for multipurpose use. **Methods:** A unified platform for longitudinal follow-up of maternal-child dyads cared for at our fetal center, named the Clinical Outcomes Data Archive (CODA), was constructed. CODA was designed using a data dictionary based on multidisciplinary

and interprofessional expert input, a relational identity for each patient, fetus, and pregnancy, and a process by which EHR-sourced and chart-abstracted data are validated by a well-trained team. Descriptive analyses were performed for data acquired between July 2022 and July 2023, and a comparison of studies before and after implementation of CODA is presented. **Conclusion:** 5,394,106 data points were validated for 7,662 patients across 12 conditions. 2% of data points were found to be unreliable or undocumented. 91% of data points were sourced from the EHR. Eighty-five percent of condition-specific variables required manual chart abstraction. The study conducted with CODA was able to contribute to 18 other studies. CODA successfully merges EHR-sourced and manually abstracted documentation for longitudinal study of the maternal-child dyad.

© 2024 S. Karger AG, Basel

## Introduction

Clinical outcomes information derived from routine care play a vital role in informing research and clinical practice where randomized control trials (RCTs), or prospective research studies are not feasible [1, 2]. These data are especially critical in the care of maternal-fetal patients presenting to a fetal center, as accurate near- and long-term risk-adjusted outcomes inform comprehensive prenatal counseling, define targets for basic and translational research, and support optimizing perinatal treatment of children with rare conditions [3–6]. Here we define a unique method for structured longitudinal prospective and retrospective data collection for maternal-fetal patients, including a distinct fetal identity, data quality control, multi-conditional data dictionaries, and systematic triggers for data collection from all patient visits.

The Center for Fetal Diagnosis and Treatment (CFDT) at the Children's Hospital of Philadelphia (CHOP) is a high-volume quaternary referral center, which seeks to understand condition-specific risk-adjusted outcomes for all maternal-fetal patients. Overall, current data management approaches are inefficient, highly error prone, and do not account for longitudinal interdisciplinary care of patients with complex congenital anomalies [7]. We developed and implemented a prospective data capture platform, the Clinical Outcomes Data Archive (CODA), to transform electronic health record (EHR) data into an analytic-grade database for multipurpose use to address these issues. CODA established a unified database for the longitudinal study of CFDT patients, and foundation for a Learning Health System to advance diagnosis, therapy,

and outcomes for patients with congenital anomalies [8, 9]. Here we describe the challenges encountered, and our approach to achieving (1) efficiency, (2) validity, and (3) trajectory through the CODA infrastructure.

### *Efficiency*

Producing risk-adjusted long-term outcomes information across rare conditions of fetal development requires collecting great amounts clinical data. The widespread implementation of EHRs creates opportunity for efficient data collection on health outcome measures [10–12]. Large volumes of clinical data can be obtained from queries of warehoused EHR data such as inpatient medication administration, anthropometrics, respiratory support, and vital status. EHR queries allow for rapid consumption of granular information otherwise too burdensome to transcribe [13]. Other strengths of EHR-sourced data include event information such as inpatient and outpatient activity, surgeries, birth, and death. However, documentation structures are not designed for acquisition of key attributes, and changes over time and between-clinician differences can affect resultant output. EHR advancements and bioinformatics optimization efforts focused on improving the availability of discrete data elements used in clinical care have not closed this gap [14]. Implementation of documentation structures for condition-specific information originating from one subspecialist group to other specialties is challenging. Informatics improvements and data governance at system-wide EHR scale makes for difficult implementation [15]. Given documentation structures are not designed to acquire key attributes, datasets derived exclusively from the EHR have a limited ability to characterize the presentation and care trajectory of patients with congenital anomalies [16, 17].

Manual data abstraction is required to capture key attributes not found in warehoused EHR data. Electronic case report forms (eCRFs) are more efficient than paper forms to acquire and manage medical record data abstractions; however, manual chart review is associated with a high resource burden and data quality challenges [2, 18]. Each project devises a set of standard operating procedures that require significant training and knowledge to execute successfully. Unfortunately, eCRFs are typically designed for a single study, leading to decentralized redundant data collection across projects with overlapping patient populations [19, 20]. Additionally, due to static design and single-study focus, eCRF data capture systems often result in incomplete abstraction of variables that may confound, mediate, or moderate critical associations [21].

While few eCRF platforms have demonstrated success leveraging EHR data [22, 23], with CODA we aimed to combine efficiencies of data collection by EHR queries with manual chart abstraction in a bespoke eCRF platform that could be used across study and specialist interests.

### *Validity*

Internal validity and data quality are essential for clinical support and research, and both EHR derived data and manually abstracted data present quality concerns. Current best practices for data abstraction recommend eCRFs be designed in the order that data will appear in the EHR and have similar language as the data source. Each variable should be accompanied with a brief operational definition, synonyms, location of the information, and abstraction rules [20]. Additionally, specialized, and repetitive completion of the same abstraction task has been shown to improve proficiency and data quality [24]. Data validation, defined as the process by which the accuracy or “trueness” of data are assessed [25, 26], is critical for all downstream uses of the data. Validation standards for manual chart review employ data ranges, dual entry, and chart review for a randomly selected percentage of records [26]. Validation has been shown to improve consistency and accuracy; however, validation efforts are rarely reported in clinical research methods [27].

In addition, the generalizability of observational clinical studies was greatly improved following the 2018 Common Rule which through broad waiver of informed consent reduced enrollment bias associated with patient recruitment [28].

### *Trajectory*

CODA was conceived as a longitudinal clinical registry with no cap on cumulative ascertainment. Each dyad, regardless of condition presenting to the CFDT would be followed prior to knowledge of diagnosis or treatment plan for the duration of their life-long interaction with CHOP. Longitudinal follow-up data are especially important in the study of patients with congenital anomalies since many physiologic and neurodevelopmental outcomes do not present until later childhood when multilevel (e.g., individual, family, policy) factors have a direct influence on health outcomes.

Given the limitations of existing eCRF platforms mentioned previously, this was not feasible. Traditionally, to capture longitudinal data, the data abstractor must manually monitor patients for reevaluation, and, upon identifying a clinical event, perform rigorous chart review

to ascertain if any clinical events of interest have transpired since last review. This process creates a situation where the eCRF remains in an incomplete state until all care is completed. Few eCRF platforms have demonstrated success leveraging the EHR to support continuous surveillance of patient activity [22, 23].

Moreover, maternal-fetal patients pose a unique challenge as the fetus often lacks a digital identity in the EHR. Consequently, fetal attributes are documented in the maternal chart and are not connected to the pediatric EHR postdelivery. This conundrum is amplified in the setting of multiple gestation pregnancies, patients returning with subsequent pregnancies, and pregnancies with fetal demise [29].

## **Methods/Design**

### *Platform Paradigm*

CODA, an innovative software platform for longitudinal follow-up of all maternal-child dyads cared for at the CFDT, was developed. In this tool, the data collection for all patients is performed in a shared setting rather than confined to single-condition registries. eCRF design was modeled by three principles: (1) to include only those variables that are documented during routine care pertaining to a single clinical event that are reliable and valid enough to be reported in a research study as defined by the subspecialist who documents the event, (2) to separate variables that are common, i.e., applicable to any patient with the same experience, from those applicable to a given condition (e.g., observed to expected lung area to head circumference ratio for congenital diaphragmatic hernia, or congenital cystic adenomatoid malformation volume ratio [30, 31]), and (3) to assure that each eCRF contains only variables requiring the same training level, e.g., cardiac sonographer, pediatric neurosurgeon, trained research assistant. These limited scope eCRFs (hereby after referred to as “cards”) are created for all variables across common and condition-specific interests.

### *Roles*

All CODA team members are assigned a role which determines their responsibilities, how they interact with data, and which eCRFs are available to view within CODA.

### *Abstractionist*

The abstractionist is responsible for data collection and validation of clinical documentation from the EHR to CODA. This role is composed of trained research

assistants, medical students, clinical staff, and physicians. The activities of this role are customized to match the level of training of the individual by card design. Abstractionist training includes shadowing clinical visits and surgical operations, as well as completing literature reviews pertaining to the clinical diagnosis or event specified in the card design. Abstractionists also serve as validators, which supports improving data definitions, escalating ambiguous situations for clinical review, and providing peer feedback, thereby enriching abstractionist education and understanding.

#### Data Analyst/Engineer

The data analyst/engineer creates source tables for ingestion of warehoused EHR data and datasets from validated data for consumption by research, completes internal data validity review, and performs interactive data visualization to inform clinical practice and translational research.

#### Biostatistician/Epidemiologist

The biostatistician/epidemiologist provides continuous methodology oversight and statistical analysis. This role ensures that data dictionary development is consistent with analytic requirements (i.e., variables are constructed in a way that are useful for analysis and require minimal primary transformation). The biostatistician/epidemiologist designs and develops analytic datasets specific to the research question.

#### Data Dictionary

The CODA data dictionary is built from multidisciplinary and interprofessional expert input. To date, expertise from fetal and pediatric clinical specialties including anesthesia, cardiology, critical care, gastroenterology, infectious disease, maternal-fetal medicine, medical genetics, neonatology, neurodevelopmental psychology, neurosurgery, nutrition, pulmonology, radiology, and surgery have participated in variable and card design. With each additional inquiry of a given population or care event, new variables and cards are designed and added to the growing data dictionary. The data dictionary definitions are embedded within the card design and displayed to CODA users abstracting and validating data. Definitions include the type, location within the medical record, reference date and time for the specific clinical care event, thereby ensuring that the data abstracted will be consistent among abstractionists.

Variables are designed such that none require computation or clinical interpretation by the abstractionist. Branching logic is precluded, thereby eliminating myriad

empty cells. All variables are answered with a value or are marked as missing or unreliably documented, causing no values to be stored as free-text inputs. For cards with variables focused on narrative documentation such as for surgeries, the complete dictated operative report is displayed within the card instance. This acquired text ensures the same record is referenced for data abstraction and secondary accuracy review.

#### Warehoused EHR Data Integration

The records of clinical events from EHR data housed in the CHOP Data Warehouse (CDW) are queried nightly (ingestion) to deploy cards, automatically abstract select data and create patient records in CODA. These queries are created through configured conditional statements in-application (hereby referred to as triggers) by users. Triggers prospectively and retrospectively assess for all patients meeting the trigger conditions. Triggers can combine multiple EHR data concepts (such as the 24 procedure names for congenital diaphragmatic hernia repair or 54 procedure names for gastrostomy tube placement) to deploy one or more cards from the same event. Variables can be configured to source data from the CDW to fill or partially fill cards. Examples include surgical date, ventilation mode and settings, and vital status. New patient records are created by triggers based on specific clinical activity, including being born in the Garbose Family Special Delivery Unit [29] or surgical event such as a postnatal spina bifida repair. The trigger feature is adaptable to inevitable changes in the EHR such as change in clinical documentation method over time and new procedure naming conventions.

#### Data Review

Once data is abstracted, all variables (100%) are reviewed for accuracy prior to these data being available for analysis, referred to as validation. The validation of manually abstracted data occurs by a second abstractionist, trained in the same card design, re-reviewing the data source, and certifying validity of each data point. Questions are escalated to clinical specialists and are discussed in the weekly interdisciplinary CODA conference. EHR-integrated data are validated by an abstractionist assisted by CODA software. In-application defined ranges support more vigorous review of abnormal values, and longitudinal data are displayed graphically to support error detection. An automated dashboard performs additional tests for abnormal values, card missingness, and data completeness. When errors are detected, a validator can correct or account for discrepancies.

**Table 1.** Summary of validated data

	Validated data	Undocumented/ unreliable	Undocumented/ unreliable, %
Condition-specific variables	111,290	15,593	14
EMR sourced	16,321	103	1
Chart abstraction	94,969	15,490	16
Common variables	5,282,816	75,265	1
EMR sourced	4,909,893	43,643	1
Chart abstraction	372,923	31,622	8
Total variables	5,394,106	90,858	2

### *Fetal and Pregnancy Identities*

The platform maintains a relational identity for each patient, fetus, and pregnancy, which allows direct linking of fetal attributes documented in the maternal chart to the child once born, and *vice versa*.

### *Data Storage*

All validated data are stored in a separate inalterable (read-only) database for analysis. Values are stored in a table as variable-value pairs, requiring the creation of analytic datasets by the data analyst/biostatistician. The database resides on a HIPAA compliant cloud server within the institution's firewall. All CODA users must log in to CODA using their CHOP username and password. Users must have a profile configured by a CODA administrator. Therefore, only those with CHOP network access and authorization may use the application.

### *Regulatory*

The CODA Registry (IRB 21-018553) was deemed exempt and granted a waiver of consent by the CHOP Institutional Review Board for following longitudinal outcomes of patients seen within the fetal center or surgical subspecialties at CHOP. CODA data used for studies beyond the scope of this exempt registry require separate IRB approval.

## **Discussion**

With CODA, data are collected prospectively, prior to knowing the course of care or ultimate outcome for each new patient, as well as retrospectively for past patients in a single registry to adapt clinical documentation into an analytic-grade dataset for multipurpose use. Two evaluations of our methods are shown below:

### *Method Evaluation 1: Descriptive Analysis of Data Volume by Type and Reliability*

Descriptive analyses were performed for validated data created between July 2022 and July 2023 and are shown in Table 1. Overall, 4,402 children and 4,484 maternal patients are enrolled in the clinical registry. Among maternal patients, 4,645 unique pregnancies and 4,861 fetuses were followed. Maternal-fetal or child patients with data created in this timeframe numbered 7,662. Patients had condition-specific cards for 12 conditions. Following abstraction and validation, 5,394,106 data-points were created. Most common variables (91%) are sourced from EHR data and were evaluated for accuracy. Across all variables, 2% of data points were found to be unreliable or undocumented during the validation process. For condition-specific variables, 85% required manual chart abstraction. Data abstraction and validation for the timeframe evaluated was performed by approximately 5.7 full-time equivalent staff.

### *Method Evaluation 2: Comparative Example*

Measures of data collected between two studies were compared and are shown in Table 2. Study 1 was conducted solely using manual chart abstraction prior to the creation of CODA [32]. Study 2 was conducted after implementing CODA [33]. Comparing study 2 (2023) to study 1 (2018), the source population was 2.4 times larger, and the number of subjects analyzed was 3 times greater in study 2. In study 2, all observations of critical laboratory values and pharmacologic administration from the EHR were collected, compared to only two and four endpoints manually collected previously. For study 2, incremental data collection and validation from maternal, pregnancy, fetal, and neonatal care were accomplished in less than 1 month. Data collected for study 1 was limited to the scope of the study and was not integrated into other datasets. The study 2 population

**Table 2.** Comparison of studies before and after the implementation of CODA

	Study 1 [32]	Study 2 [33]
Year of publication	2018	2023
Study subjects	17 patients	51 patients
Population	147 patients	351 patients
Study timeframe	69 months	103 months
Enrollment per month	2.1 patients	3.4 patients
Primary outcome (laboratory)	4 timepoints	All observations (thousands)
Primary exposure (medical)	2 timepoints	All administrations (thousands)
Project contribution	1 project	18 projects

and information from all subsequent patients were continuously added to CODA, contributed to 18 additional projects, and supported real-time risk-adjusted clinical outcomes information available for patient counseling.

#### *Implication*

The study comparison illustrates how CODA allows for more efficient production of an analytic database that continues to be updated and benefits future studies. Since data are abstracted and validated on an ongoing basis, the validated database supports real-time reporting of risk-adjusted patient outcomes reducing the barriers to answering important clinical questions. These examples serve to illustrate how the CODA platform was designed to solve for three key attributes: efficiency, validity, and trajectory.

#### *Efficiency*

The descriptive analysis of datapoints created in 1 year illustrates the impact of EHR data to increase scale. By leveraging the EHR triggers created through simple conditional statements configured in the application, the onus of EHR database queries shifts from the data analyst to the abstractionist. Since longitudinal care events are systematically identified, the time and effort required for chart abstraction are substantially reduced, eliminating the manual task of monitoring patient activity from the abstractionist. Additionally, the development of data queries is shifted from the data analyst's portfolio, thus freeing time for specialized tasks.

With our finding that 85% of condition-specific information requires manual abstraction, the role of a well-trained abstractionist remains vital. For example, at our institution, fundamental elements for risk-

adjustment for CDH including CDH side, liver position, and observed to expected lung area to head circumference ratio cannot be ascertained from warehoused EHR data, while length of stay, respiratory support, and vitals status can be extracted. Most cards, based on their design, are amenable to abstraction and validation by research assistants, and select few cards were designed for highly trained clinical staff and physicians. Because data abstraction focuses on individual clinical events that could be repeatedly performed, abstractionist training, proficiency, and productivity are enhanced.

The multi-conditional nature of CODA reduces redundancy. Centralizing patient data records reduces burden of data abstraction for identical or overlapping cohorts across research projects and ensures data definitions are equal. With common variables in place, the incremental effort required to add new condition-specific cohorts is low. These efforts allow for more complete patient data profiles that reduce the number and frequency of missing variables that often confound, mediate, or moderate outcomes.

#### *Validity*

While CODA has significantly expanded the number of datapoints able to be captured within a clinical registry, the thoughtful design process is key to its success. Each card, variable, and trigger is carefully crafted with a clinical subject matter expert, biostatistician, epidemiologist, and data analyst to ensure that the data answer questions accurately. Clinical specialists create variables within their area of expertise which are then used by all investigators interested in those metrics to answer specific clinical or research inquiries, thus improving the quality of data definitions. For example, a surgeon is able to use obstetric variables designed by expert maternal-fetal

medicine specialists, while the surgeon's expertise is used to define surgical variables which are then made available to others. CODA's multidisciplinary design replicates the team-based approach to patient care used in our fetal center.

The descriptive analysis of datapoints created in 1 year illustrates the importance of manual chart review to validate manually abstracted and EHR-sourced variables. For the latter EHR data, careful review for artifacts of clinical documentation is warranted. While not pervasive, these artifacts greatly affect analysis and are not amenable to filtering based on parameters alone. Triggers allow for adaptation of EHR database queries to combine multiple data concepts. This is important to allow for change in the EHR documentation to be mapped to the same data concept in CODA.

Trigger-based patient enrollment captures each patient systematically. This semi-automated process allows CODA to accumulate a large volume of patients and variables in the archive and makes it amenable to epidemiological study designs (case-control) suitable for rare events. Consistent patient enrollment has eliminated nonresponse bias within our patient populations, thereby improving the epidemiologic integrity of research studies (demonstrated by the comparison of studies).

#### Trajectory

The centralized model shifts the focus of clinical outcomes information from project to patient. Patient activity, not project specific goals, drives data collection and each component of care is followed analogous to a specialist. Centralized clinical outcomes research shares resources and knowledge to capture all data encompassing the multifaceted care experienced by the patient over time. The prospective longitudinal data capture system allows investigators to develop traditional cohort studies with adequate sample size and accurate outcomes as well as studies to evaluate change in metrics over time. Auto-enrollment of patients allows patients to be followed prospectively from the moment of the inclusion event (prenatal evaluation, birth, surgery, etc.) until their last interaction with our institution. CODA's unique fetal identity creates the ability to connect fetal attributes directly to the child which allows prenatal prognosticators to be evaluated as factors impacting long-term outcomes. The continual evaluation of outcomes that arise over time allows discoveries to grow alongside the patient and across the specialties they interact with at CHOP.

#### Limitations

Although CODA addresses these three thematic issues with EHR-based study, there remain limitations to consider. While CODA's EHR-integrated design reduces abstraction burden, manual data abstraction is required, including ample allocation of staff time and training. Linking fetal attributes from maternal EHR to the child's record postdelivery also requires manual effort as the EHR currently lacks a fetal identity. The breadth of data collection within CODA helps inform clinical practice and research, but it should be noted that the unstructured nature of clinical care documentation is not matched to the rigor of a RCT or a purpose-designed prospective cohort study. Although RCTs are the gold standard in clinical research, well-designed prospective cohort studies that control for confounders may be the only option for situations where RCTs are not feasible due to ethical or financial reasons. We also note that the EHR integration is limited to our single center, which reduces generalizability of follow-up data as only encounters documented within the CHOP EHR are captured. Additionally, our study and method was not designed to evaluate the validity of data within or derived from the EHR.

#### Next Steps, Opportunities

Additional improvements to this design and future clinical databases include expanding EHR integration to automate integration of data captured by advanced patient-monitoring systems and working with clinicians to improve data input to the EHR for auto-integration of clinical notes. Centralization and collaboration of clinical outcomes research should not be limited within a single center in a single hospital. Multicenter and multi-institution EHR-integrated registries are the future to optimizing clinical outcomes research across conditions. Additionally, with a large volume of manually abstracted data, there is an opportunity to create a machine learning feature to automate the abstraction of unstructured clinical documentation and present this information to a validator.

#### Conclusion

With the implementation of CODA, our fetal center has successfully mitigated complexities of EHR documentation for longitudinal clinical outcomes study of thousands of datapoints for thousands of maternal-child dyads. This achievement was due to an innovative

process, a well-trained team, consistent data definitions, and interdisciplinary collaboration. A unified data platform provides opportunities for interdepartmental collaboration to build projects that account for all aspects of the child's life with high quality data definitions, and for which underlying data can be used in future studies. The unified platform also avoids contradictory conclusions from incomplete datasets when specialists study overlapping patient cohorts. This overview of our methods is intended to serve as a reference for future research supported by CODA and to encourage innovation in EHR to eCRF development.

## Acknowledgments

A culture of innovation and collaboration resulted in the progress outlined in this report. Many individuals contributed to this success, including Deepak Nandagopal, MS (Data Engineer); Jake Liguori (Data Analyst); Anna Bostwick, MPH (Clinical Research Coordinator); Meghan Yizzi, Kiersten Barr, Maddie Dombroski, Jamie Fromm, Colby Gallagher, Cara Katterman, and Maria Ordonez (Clinical Research Assistants) from the Center for Fetal Diagnosis and Treatment team. CODA technology is continuously supported by an exceptional CHOP Information Systems team led by Ryan Treftz, MS (IS Developer). Important contributions were made by Suresh Anjetty, BS; Phillip Duba, MS (Systems Manager); Frank Lamacchia, MBA (IS Business Analyst Principal); Amrita Srivastava, MS (IS Developer); Maria Tickner, BS (Validation Analyst); and Jeffrey Bachovchin, MSE (Senior Director of Web Technology and Application Development). CHOP attending physician contributors also include Drs. Thane Blinman, Beverly Coleman, Tracy Flanders, Dustin Flannery, Olivia Nelson, Christina Pidas-Teefey, Emily Partridge, Amanda Shillingford, and Taylor Wild along with psychologists Drs. Casey Hoffman and Jane Schreiber. CHOP physician trainee contributors include Drs. Felix De Bie, Anthony Di Natale, Alicia Eubanks, Sandy Johng, and Uchenna Kennedy. Student contributors pursuing medical studies included Sarah Hamimi, Danielle Ireland, Jenny Engelman, Alana Platukus, Agustina Quiroga Carrizo, Elizabeth Eppley, and Isabel Ryan. Clinical context and content development support was provided by Jim Connelly, RRT-NPS (ECMO Center Manager); Suzanne DeBari, RDMS, RVT, RT (Fetal Ultrasound Team Lead); Audrey Foster, MS, RD, LDN (Clinical Dietician); Lisa Herkert, MSN, CRNP (Advanced Practice Nurse); Joy Macdonald, RN (Fetal Therapy Coordinator); Katie Schmidt, MSN, CRNP (Surgical Nurse Practitioner); and Susan Spinner, MSN, RN (Clinical Program Operations Director).

We wish to specially thank Lori Howell DNP, MS, RN, a key author of this work, who was a visionary in the field of patient-centered care in fetal centers. Her unwavering commitment to pursuing the impossible for patients not only established an international model for patient-centered care but also served as a driving force behind the development of this platform. We honor her memory by recognizing her pivotal role in setting the goal of

establishing real-time, risk-adjusted outcomes data for patient care. This paper is dedicated to her memory, as a tribute to her lasting impact on the field. The authors would also like to thank the International Fetal Medicine and Surgery Society for the opportunity to present this work at the 40th Annual International Fetal Medicine and Surgery Society Conference in Visby, Stockholm.

## Statement of Ethics

This study was approved by the Institutional Review Board at the Children's Hospital of Philadelphia (IRB 21-018553). The CHOP Institutional Review Board approved this observational study with a waiver of informed parental consent.

## Conflict of Interest Statement

The authors have no conflicts of interest to declare.

## Funding Sources

These efforts are possible because of the philanthropy of grateful CFDT families and were amplified by internal funding through the CHOP Frontier Programs and CHOP Information Services.

## Author Contributions

T.A.R. devised the method, led implementation, conducted literature review, and drafted the manuscript. M.A.G. cocreated the method, assisted in program implementation, and supplied manuscript revisions. S.F. and S.L. co-implemented the method and assisted in drafting the manuscript and conducting literature review. L.M. provided critical oversight in method implementation, assisted in methods development, literature review, and manuscript revisions. A.M.A., C.M.A., E.E.F., E.R.O., G.G.H., H.B.P., J.S.G., and N.E.R. contributed clinical expertise essential to method implementation. L.J.H. and N.S.A. established the goal of patient-centered real-time risk-adjusted clinical outcomes for all CFDT patients, allocated funding, support, and critical revisions of manuscript. H.L.H. contributed clinical expertise, support, project development, mentorship, and revised the manuscript. All authors approved the final version of the manuscript prior to publication.

## Data Availability Statement

The patient health information data that support the findings of this study are not publicly available due to health data privacy protections. The metadata used to illustrate the system are not publicly available as they are derived from and stored within the proprietary database. Further data inquiries can be directed to the corresponding author upon reasonable request.

## References

- 1 Gerritsen MG, Sartorius OE, vd Veen FM, Meester GT. Data management in multicenter clinical trials and the role of a nation-wide computer network. A 5 year evaluation. *Proc Annu Symp Comput Appl Med Care*. 1993;659–62.
- 2 Zozus MN, Choi BY, Garza MY, Facile R, Lanham HJ, Wang Z, et al. Collaborative program to evaluate real world data for use in clinical studies and regulatory decision making. *AMIA Jt Summits Transl Sci Proc*. 2023;2023:632–41.
- 3 Howell LJ, Adzick NS. The essentials of a fetal therapy center. *Semin Perinatol*. 1999;23(6): 535–40. [https://doi.org/10.1016/s0146-0005\(99\)80032-9](https://doi.org/10.1016/s0146-0005(99)80032-9)
- 4 Evans MI, Harrison MR, Flake AW, Johnson MP. Fetal therapy. *Best Pract Res Clin Obstet Gynaecol*. 2002;16(5):671–83. <https://doi.org/10.1053/beog.2002.0331>
- 5 Baschat AA, Blackwell SB, Chatterjee D, Cummings JJ, Emery SP, Hirose S, et al. Care levels for fetal therapy centers. *Obstet Gynecol*. 2022;139(6):1027–42. <https://doi.org/10.1097/AOG.0000000000004793>
- 6 Deprest JA, Flake AW, Gratacos E, Ville Y, Hecher K, Nicolaides K, et al. The making of fetal surgery. *Prenat Diagn*. 2010;30(7): 653–67. <https://doi.org/10.1002/pd.2571>
- 7 Ruth CJ, Huey SL, Krisher JT, Fothergill A, Gannon BM, Jones CE, et al. An electronic data capture framework (ConnEDCt) for global and public health research: design and implementation. *J Med Internet Res*. 2020; 22(8):e18580. <https://doi.org/10.2196/18580>
- 8 Buchan I, Ainsworth J. Combining health data uses to ignite health system learning. *Methods Inf Med*. 2015;54(06):479–87. <https://doi.org/10.3414/ME15-01-0064>
- 9 Denaxas S, Friedman CP, Geissbuhler A, Hemingway H, Kalra D, Kimura M, et al. Discussion of “combining health data uses to ignite health system learning. *Methods Inf Med*. 2015;54(06):488–99. Erratum in: *Methods Inf Med*. 2016;55(2):201. <https://doi.org/10.3414/ME15-12-0004>
- 10 Rosenbloom ST, Smith JRL, Bowen R, Burns J, Riplinger L, Payne TH. Updating HIPAA for the electronic medical record era. *J Am Med Inform Assoc*. 2019;26(10):1115–9. <https://doi.org/10.1093/jamia/ocz090>
- 11 Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;309(13):1351–2. <https://doi.org/10.1001/jama.2013.393>
- 12 Psaty BM, Breckenridge AM. Mini-Sentinel and regulatory science—big data rendered fit and functional. *N Engl J Med*. 2014;370(23):2165–7. <https://doi.org/10.1056/NEJMp1401664>
- 13 Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood S, et al. Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform*. 2014;52:28–35. <https://doi.org/10.1016/j.jbi.2014.02.003>
- 14 Shahmoradi L, Darrudi A, Arji G, Farzaneh Nejad A. Electronic health record implementation: a SWOT analysis. *Acta Med Iran*. 2017;55(10):642–9.
- 15 Tasian GE, Ellison JS; Pediatric KIDney Stone (PKIDS) Care Improvement Network. The surgical improvement cycle: improving surgical outcomes through partnerships and rigor. *J Urol*. 2021;205(6):1554–6. <https://doi.org/10.1097/JU.0000000000001626>
- 16 Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clin Epidemiol*. 2017;9: 245–50. <https://doi.org/10.2147/CLEP.S129779>
- 17 Cohen B, Vawdrey DK, Liu J, Caplan D, Furuya EY, Mis FW, et al. Challenges associated with using large data sets for quality assessment and research in clinical settings. *Policy Polit Nurs Pract*. 2015; 16(3–4):117–24. <https://doi.org/10.1177/1527154415603358>
- 18 Zozus MN, Pieper C, Johnson CM, Johnson TR, Franklin A, Smith J, et al. Factors affecting accuracy of data abstracted from medical records. *PloS one*. 2015;10(10): e0138649. <https://doi.org/10.1371/journal.pone.0138649>
- 19 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2): 377–81. <https://doi.org/10.1016/j.jbi.2008.08.010>
- 20 Engel L, Henderson C, Fergenbaum J, Colantonio A. Medical record review conduction model for improving interrater reliability of abstracting medical-related information. *Eval Health Prof*. 2009;32(3):281–98. <https://doi.org/10.1177/0163278709338561>
- 21 Allison JJ, Wall TC, Spettell CM, Calhoun J, Fargason CA Jr, Kobylinski RW, et al. The art and science of chart review. *Jt Comm J Qual Improv*. 2000;26(3):115–36. [https://doi.org/10.1016/s1070-3241\(00\)26009-4](https://doi.org/10.1016/s1070-3241(00)26009-4)
- 22 Murphy EC, Ferris FL 3rd, O'Donnell WR. An electronic medical records system for clinical research and the EMR EDC interface. *Invest Ophthalmol Vis Sci*. 2007;48(10): 4383–9. <https://doi.org/10.1167/iovs.07-0345>
- 23 Ethier JF, Curcin V, McGilchrist MM, Choi Keung SNL, Zhao L, Andreasson A, et al. eSource for clinical trials: implementation and evaluation of a standards-based approach in a real world trial. *Int J Med Inform*. 2017;106:17–24. <https://doi.org/10.1016/j.ijmedinf.2017.06.006>
- 24 Eder C, Fullerton J, Benroth R, Lindsay SP. Pragmatic strategies that enhance the reliability of data abstracted from medical records. *Appl Nurs Res*. 2005;18(1):50–4. <https://doi.org/10.1016/j.apnr.2004.04.005>
- 25 Scheid LM, Brown LS, Clark C, Rosenfeld CR. Data electronically extracted from the electronic health record require validation. *J Perinatol*. 2019;39(3):468–74. <https://doi.org/10.1038/s41372-018-0311-8>
- 26 Newgard CD, Zive D, Jui J, Weathers C, Daya M. Electronic versus manual data processing: evaluating the use of electronic health records in out-of-hospital clinical research. *Acad Emerg Med*. 2012;19(2):217–27. <https://doi.org/10.1111/j.1553-2712.2011.01275.x>
- 27 Zbrozek A, Hebert J, Gogates G, Thorell R, Dell C, Molsen E, et al. Validation of electronic systems to collect patient-reported outcome (PRO) data: recommendations for clinical trial teams: report of the ISPOR ePRO Systems Validation Good Research Practices Task Force. *Value in Health*. 2013;16(4): 480–9. <https://doi.org/10.1016/j.jval.2013.04.002>
- 28 Department of Health and Human Services. 45 cfr 46.104. <https://www.ecfr.gov/current/title-45/part-46/section-46.104>
- 29 Harron K, Gilbert R, Cromwell D, van der Meulen J. Linking data for mothers and babies in de-identified electronic health data. *PLoS One*. 2016;11(10):e0164667. <https://doi.org/10.1371/journal.pone.0164667>
- 30 Oluyomi-Obi T, Kuret V, Puligandla P, Lodha A, Lee-Robertson H, Lee K, et al. Antenatal predictors of outcome in prenatally diagnosed Congenital Diaphragmatic Hernia (CDH). *J Pediatr Surg*. 2017;52(5):881–8. <https://doi.org/10.1016/j.jpedsurg.2016.12.008>
- 31 Crombleholme TM, Coleman B, Hedrick H, Liechty K, Howell L, Flake AW, et al. Cystic adenomatoid malformation volume ratio predicts outcome in prenatally diagnosed cystic adenomatoid malformation of the lung. *J Pediatr Surg*. 2002;37(3):331–8. <https://doi.org/10.1053/jpsu.2002.30832>
- 32 Lawrence KM, Hedrick HL, Monk HM, Herkert L, Waqar LN, Hanna BD, et al. Treprostinil improves persistent pulmonary hypertension associated with congenital diaphragmatic hernia. *J Pediatr*. 2018;200:44–9. <https://doi.org/10.1016/j.jpeds.2018.04.052>
- 33 De Bie FR, Avitabile CM, Flohr S, Land S, Mathew L, Wang Y, et al. Treprostinil in neonates with congenital diaphragmatic hernia-related pulmonary hypertension. *J Pediatr*. 2023;259:113420. <https://doi.org/10.1016/j.jpeds.2023.113420>