Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-Assisted Pregnancy Episode Identification from All Payers Claim Dataset

Yash Devkota
University of Nevada, Las Vegas
Sobhan Ebrahimi Azar
University of Nevada, Las Vegas
Sepideh Farivar
University of Nevada, Las Vegas
Jesse Ortega
University of Nevada, Las Vegas
Jennifer Vanderlaan
University of Nevada, Las Vegas
Jorge Fonseca
University of Nevada, Las Vegas
Kazem Taghva
kazem.taghva@unlv.edu

University of Nevada, Las Vegas

Research Article

Keywords: All-Payer Claims Database (APCD), Transformer Architecture, Electronic Health Records (EHR), Artificial Intelligence in Nursing

Posted Date: June 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9382980/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Advances in Computational Intelligence on October 27th, 2026. See the published version at https://doi.org/10.1007/s43674-026-00095-3.

AI-Assisted Pregnancy Episode Identification from All Payers Claim Dataset

Yash Devkota[1], Sobhan Ebrahimi Azar[1], Sepideh Farivar[1], Jesse Ortega[1], Jennifer Vanderlaan[2], Jorge Fonseca[1], Kazem Taghva[1*]

[1*]Department of Computer Science, University of Nevada, Las Vegas, Las Vegas, NV, USA.
[2]School of Nursing, University of Nevada, Las Vegas, Las Vegas, NV, USA.

*Corresponding author(s). E-mail(s): kazem.taghva@unlv.edu;

Abstract

Identification of complete pregnancy episodes in All Payer Claims Datasets (APCD) is essential for maternal health research. These data sets store maternal-related encounters based on provider specialty and pregnancy-specific ICD codes. In this paper, we report on a two-stage approach that combines deep learning with a post-processing step to identify and number pregnancies for individual patients based on Virginia APCD.

The training and testing of our approach was done on a subset of Virginia APCD consisting of 397,000 individuals. On the test set, 93.90% of individuals' pregnancies were labeled with perfect accuracy. On a per-event basis, the model achieved a weighted average F1-score of 0.99 indicating high accuracy in classifying each clinical claim into its specific, numbered pregnancy episode (e.g., ***PR*#1** for first prenatal, **3#*B*** for third birth). Further analysis reveals that the model excels at identifying common first and second pregnancies, but fails in the post-processing stage of correctly labeling rare and complex cases such as third or fourth pregnancies. This finding is significant, as it demonstrates the success of our hybrid approach and suggests that future improvements can be achieved by refining the deterministic post-processing logic.

Keywords: All-Payer Claims Database (APCD), Transformer Architecture, Electronic Health Records (EHR), Artificial Intelligence in Nursing

Introduction

Severe Maternal Morbidity (SMM), a set of life-threatening conditions and complications arising during pregnancy and childbirth, has become a critical focus of public health research in the United States [1]. Understanding the prevalence of SMM and its association with pre-existing maternal comorbidities is essential to develop effective clinical interventions, shape health policy, and address persistent health disparities across racial and ethnic groups [2–5]. Researchers have long demonstrated the power of large administrative datasets to uncover significant disparities in healthcare access and utilization among vulnerable populations, highlighting the critical need for robust analytical methods [6].

Large administrative health datasets, particularly All-Payer Claims Databases (APCDs), have emerged as powerful resources for this type of population-level research [7–9]. The integration of artificial intelligence into nursing research provides powerful new methods to analyze such large-scale data, offering insights that can inform clinical practice and health policy. As comprehensive repositories of claims from various payers, APCDs are increasingly recognized by policymakers for their potential to provide a holistic view of healthcare utilization and cost [10]. Such databases enable the creation of citywide or regional health information systems that offer unique insights into population health patterns [11]. These integrated datasets are foundational for developing targeted interventions, such as identifying and supporting high-utilizing patients to improve outcomes and reduce costs. However, realizing this potential is hindered by significant technical and logistical hurdles, reflecting the persistent challenges of health information exchange and data integration [12].

In our work, we are interested in the use of administrative claims data to link individual clinical events to a specific pregnancy episode explicitly. A patient's record is a time-stamped sequence of diagnosis and procedure codes, from which distinct pregnancy journeys must be inferred. Traditional methods for constructing these episodes rely on complex, handcrafted rule-based algorithms [13–15]. These systems, while built on deep clinical expertise, are labor-intensive to develop, brittle in the face of evolving coding practices, and struggle to generalize across different healthcare systems. They often fail to capture the nuanced, nonlinear patterns of care present in messy, real-world data, especially for patients with complex comorbidities or poorly coded outcomes like early-term miscarriages.

This paper addresses this methodological gap by proposing a deep learning framework to identify and delineate pregnancy episodes from claims data automatically. Our contributions are:

1. A custom Transformer architecture that effectively integrates sparse clinical event codes with temporal and clinical context features to understand patient trajectories (i.e., the sequence and timing of a patient's pregnancy events over time).
2. A two-stage labeling framework that decouples the complex pattern recognition task from the structured, sequential numbering task, specifically designed to handle the data sparsity common in higher-order pregnancies.

3. A comprehensive evaluation on a large-scale, real-world claims dataset, demonstrating the high accuracy of our approach and identifying key areas for future refinement in post-processing logic.

Background: Transformers for Sequential Health Data

The Bidirectional Encoder Representations from Transformers (BERT) model has revolutionized the field of Natural Language Processing (NLP) [16]. Unlike previous sequential models like Recurrent Neural Networks (RNNs) that process text in one direction, BERT's Transformer architecture uses a self-attention mechanism. This mechanism allows the model to weigh the importance of all other "words" in a sentence when encoding a given word, enabling it to learn deep, bidirectional relationships and understand context with unprecedented accuracy. The success of BERT inspired its adaptation for structured, sequential data, such as Electronic Health Records (EHR) and administrative claims. The BEHRT model [17], BERT for EHR, is an effort in this area. It represents a patient's health record as a "sentence" composed of "words" that represent clinical concepts such as diagnosis or procedure code. By pre-training on large amounts of unlabeled patient data, BEHRT learned clinically meaningful embeddings of medical concepts and patient trajectories. These learned representations could then be fine-tuned for specific downstream tasks like disease prediction or risk stratification, often outperforming traditional statistical models.

Our work builds on this foundation. We adopt the core principle of BEHRT; treating a sequence of clinical events as a language to be understood, but we make critical adaptations for our specific task. Instead of relying solely on pre-training, we fine-tune a Transformer model directly for the sequence labeling task of identifying pregnancy-related events. Furthermore, we enhance the input representation by creating a composite embedding that explicitly incorporates temporal and clinical context features alongside the event codes. This tailored approach is designed to meet the unique challenge of delineating pregnancy episodes, which requires understanding not only the clinical meaning of events, but also their precise timing and context within a patient's broader health journey.

Methods

Our methodology for automatically identifying and numbering pregnancy episodes is structured into four main parts. In the Dataset and Cohort Selection subsection, we give an overview of dataset preparation and splitting. The Feature Engineering subsection presents our feature engineering for the creation of model inputs. Finally, the Two-Stage Episode Labeling Framework and Experimental Setup subsections summarize our learning model to predict generic event types and the deterministic algorithm to convert generic predictions into specific numbered episodes. This project was deemed exempt from ethical review by the University of Nevada Institutional Review Board because the de-identified data does not meet the definition of human subjects research.

Dataset and Cohort Selection

The study was conducted on a large, de-identified administrative claims dataset containing the longitudinal health records of 397,391 unique individuals. The data is owned by a third party and can be purchased for research use from Virginia Health Information at https://www.vhi.org/apcd/. We have developed a rule-based pipeline for identifying complete pregnancy episodes in all-payer claims data, combining facility and professional records into a unified, diagnosis-normalized dataset [15]. Maternal-related encounters were retained based on provider specialty and pregnancy-specific ICD-10 codes with non-informative diagnoses excluded. Each record was flagged as inpatient or outpatient to support later outcome analysis.

Gestational age was derived from `Z3A.*` ICD-10 codes where available, enabling back-calculation of pregnancy start dates. In their absence, the method applied a 36-week fixed interval or adjusted to dataset boundaries. Pregnancy terminations were identified from ICD-10 delivery and loss codes, validated by temporal rules ($\geq$ 6 months between births, $\geq$ 56 days between losses) to avoid miscoding. Delivery episodes were extended to cover the full admission–discharge window, ensuring inclusion of all related inpatient services.

Postpartum intervals of 56 days were enforced to separate consecutive pregnancies, and prenatal intervals were labeled from the estimated start date to termination. The approach produced non-overlapping pregnancy episodes with clear prenatal, termination, and postpartum labels. This framework offers a reproducible, transparent baseline for maternal health research and for developing more advanced predictive models.

Throughout this rule-based data set preparation, our team relied heavily on the subject matter expert of the school of nursing faculty. This phase of our work has taken six months. One of our long-term objectives is to classify pregnancy episodes and other pattern discoveries with machine learning methods. There are two main reasons to build an AI-based classification in parallel to our rule-based dataset construction:

- There are many states beyond Virginia that require the construction of APCD to help with the state policies for healthcare. There is no standard set of rules that these states follow to build their version of APCD.
- The ICD codes are constantly revised with different versions.

These obstacles will require that we manually change our rules to adapt to the new version of ICD codes and the variety of data formats from different states.

A key challenge in using insurance billing data for pregnancy research is that it records individual services rather than telling a continuous patient story. Each claim is for a single event, like a doctor's visit or a test, and lacks a clear connection to the full course of a pregnancy. A patient's history appears as a chronological list of billing codes, which includes definite pregnancy care alongside routine appointments, unrelated illnesses, and other vague medical events. As a result, pinpointing when a pregnancy truly begins and ends is a difficult task. The start could be a general check-up, a visit for a missed period, or the first positive pregnancy test—all happening at different times. Similarly, the endpoint is unclear; it could be the delivery, the standard

six-week follow-up visit, or later, if complications arise after birth. Table 1 illustrates a typical, complex patient journey that highlights these challenges.

Table 1 A Mimicked Example of Patients Journey from Claims Data Illustrating Delineation Challenges

Date	Code	General Description	Interpretation Challenge
2023-01-15	D_Z01.419	Routine gynecological exam	Is this pre-conception care or unrelated?
2023-03-20	D_N91.2	Amenorrhea (missed period)	Possible start of pregnancy symptoms.
2023-04-05	P_81025	Urine pregnancy test, positive	Strong signal, but not the conception date.
2023-05-10	D_Z34.00	Supervision of normal pregnancy	Confirms ongoing pregnancy.
2023-07-22	D_J02.9	Sore throat	Interleaved, unrelated care to be ignored.
2023-11-15	D_O02.1	Missed abortion (Loss)	Definitive end of pregnancy, but not the episode.
2023-11-29	D_Z39.2	Routine postpartum follow-up	Part of the episode, but is this the end?
2024-02-10	D_M54.5	Low back pain	Is this a late postpartum issue or unrelated?

Source: Table recreated from clinical claims data examples.

The rule-based prepared data was then used as the ground truth for our transformer model. To ensure a robust evaluation and prevent data leakage, where information about a single individual could appear in both training and testing sets, we partitioned the entire dataset randomly at the person level.

The data was split as follows:

- **Training Set:** Approximately 320,000 individuals, used to train the deep learning model.
- **Validation Set:** Approximately 40,000 individuals, used to tune model hyperparameters and select the best-performing model epoch to prevent overfitting.
- **Test Set:** The remaining individuals, used for the final, unbiased evaluation of the chosen model's performance.

Feature Engineering

To prepare the raw claims data for our model, we engineered a set of features for each clinical event. Each feature was designed to provide a different type of context.

- **Event Code:** This is the primary feature representing the clinical activity. We combined diagnosis codes and procedure codes into a single vocabulary. To preserve their origin, codes were prefixed with 'D_' for diagnoses and 'P_' for procedures (e.g., 'D_Z3400', 'P_10907ZC'). This unified vocabulary allows the model to learn from all recorded clinical codes.
- **Time Difference (in days):** To provide the model with a sense of time and visit frequency, we have calculated the number of days that had passed since the patient's previous event. For an individual's first event in the dataset, this value was set to 0.
- **Hospitalization Status:** We have created a binary feature ('0' or '1') to indicate whether an event occurred as part of an inpatient hospital stay. This helps the model distinguish between routine outpatient visits and more significant hospitalization events.

- **Pregnancy Week:** The gestational week of the pregnancy at the time of the event was included as a numerical feature. A special value of -1 was used to represent events where this information was missing, allowing the model to learn to handle such cases.

Two-Stage Episode Labeling Framework

Our core contribution is a two-stage framework that splits the complex task of pattern recognition from the structured task of episode numbering.

Stage 1: Generic Event Labeling with a Transformer Model

A significant challenge in this task is data sparsity, particularly for higher-order pregnancies. Directly training a model to predict specific labels like 'PR#3', '4#L', or '5#B' would result in a vast number of classes, many with very few examples in the training data. This would likely lead to poor model performance and an inability to generalize to unseen data.

To address this, we simplified the problem in the first stage by training the Transformer model to predict a generic clinical category for each event, ignoring the specific pregnancy number. This allows the model to learn the fundamental patterns of prenatal care, birth, loss, and postnatal care from a much larger pool of aggregated examples.

- **Model Task:** For each event in a person's timeline, the model predicts one of five broad labels: **PR** (Prenatal), **B** (Birth), **L** (Loss), **PO** (Postnatal), or **O** (Other/non-pregnancy related).
- **Model Architecture:** We designed a model based on the Transformer architecture, which excels at understanding context in sequential data. The core of our model is a multi-layer Transformer encoder, whose base weights were initialized from a standard 'bert-base-uncased' checkpoint. The key innovation lies in how we represent each clinical event before it is processed by the Transformer's self-attention layers as seen in Fig 1. Instead of relying on a single embedding for the event code, we have constructed a rich **composite input embedding** by summing four separate, concurrently trained embedding vectors:

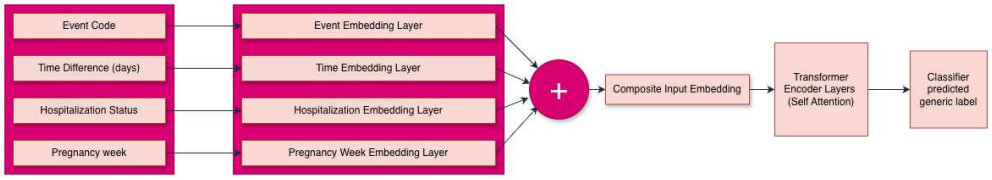


Fig. 1 Model Architecture

1. **Event Code Embedding:** This is the foundational embedding, representing the clinical meaning of each unique diagnosis or procedure code. It is a high-dimensional vector learned from a vocabulary of all event codes present in the training data. This allows the model to understand the semantic similarity between different clinical events (e.g., different types of prenatal screenings).
2. **Time Difference Embedding:** This vector learns the temporal patterns associated with the number of days since the patient's last visit. By converting the continuous time difference into a learned embedding, the model can understand the significance of visit frequency and identify meaningful gaps in care that might signal the end of one episode or the start of another.
3. **Hospitalization Status Embedding:** This vector learns the distinct context of inpatient versus outpatient care. It allows the model to assign greater importance to significant events like a delivery hospitalization compared to a routine outpatient check-up.
4. **Pregnancy Week Embedding:** This vector learns the significance of the gestational age at which an event occurs. This provides a powerful signal, as certain procedures and diagnoses are highly specific to particular stages of pregnancy. The embedding layer is designed to handle missing gestational week data by mapping it to a dedicated vector.

This composite representation is then fed into the multi-layer Transformer encoder. The self-attention mechanism within the Transformer weighs the influence of all other events in the patient's sequence to produce a final, contextually-aware representation for each event. This output vector is then passed to a final linear classification layer with a SoftMax activation function, which outputs a probability distribution over the five generic labels.

Stage 2: Deterministic Episode Renumbering

After the model generates a sequence of generic labels for a person (e.g., 'PR, PR, O, B, PO, PR, L'), our renumbering algorithm processes this sequence to assign the final, specific episode numbers.

- **Goal:** To convert generic labels into specific, numbered episodes (e.g., 'PR#1, PR#1, O, 1#B, PO#1, PR#2, 2#L').
- **Logic:** The algorithm processes a person's events chronologically and maintains a state based on the last significant non-'O' label encountered. A pregnancy episode counter is advanced based on the following rules:

  1. The **pregnancy counter** is incremented only when a 'PR' (Prenatal) event is encountered and the last significant event marked the end of a previous pregnancy (i.e., was a 'B', 'L', or 'PO'), or if it's the first significant event for that person.
  2. 'B' (Birth), 'L' (Loss), and 'PO' (Postnatal) events use the currently active pregnancy counter. They do not increment the counter themselves. This ensures that multiple 'B' events for a single delivery are all correctly labeled as part of the same episode (e.g., '1#B').

3. 'O' (Other) labels are ignored for state-tracking purposes, allowing the context of the last significant event to persist across gaps of non-pregnancy-related care.

Experimental Setup

The model was trained for a total of 5 epochs[18] using the AdamW[19] optimizer with a learning rate of 3×10^{-5} and a batch size of 32. To select the best version of the model, we monitored the validation loss at the end of each epoch. The model checkpoint from **Epoch 4** was chosen for the final evaluation on the test set, as it achieved the lowest validation loss, indicating the best generalization performance before the onset of overfitting.

Results and Discussion

Overall Model Performance

The performance of our two-stage model was evaluated on the held-out test set, composed of patient data never seen during training or validation. The model demonstrated high efficacy. For 93.90% of individuals, all their events were classified successfully. This indicates that, for nearly 94% of individuals in the test set, the model's final predicted specific labels for their clinical events exactly matched the original ground truth labels.

The detailed classification metrics, shown in Table 2, further underscore the model's strong performance. The weighted average and micro average F1-scores are both 0.99, reflecting excellent accuracy on a per-event basis, driven by strong performance on the most frequent classes (i.e., first and second pregnancies). The macro average F1-score of 0.86, while still high, suggests variability in performance across different label classes, which merits a deeper analysis.

Error Analysis and Interpretation

The results indicate that the model is extremely proficient at identifying first- and second-pregnancy episodes, which constitute the vast majority of cases in the dataset. The primary source of error, as highlighted by the lower macro average F1-score, is concentrated in higher-order pregnancies (3rd, 4th, and beyond).

To investigate this, we analyzed the confusion matrix and calculated the false negatives for each class (Table 3). A false negative occurs when the model fails to predict the correct label for an event. The analysis reveals that while the model is highly accurate for first-time pregnancy events (e.g., only 108 false negatives for '1#B' out of over 237,000 instances), the number of misclassified events becomes more significant for higher-order pregnancies. For instance, the model misclassified 1,687 true '2#B' events, and 2,600 true 'PR#3' events were not correctly labeled (many being predicted as 'PR#2').

This pattern of misclassification, where the error is typically an off-by-one mistake in the episode number (e.g., a true 'PR#3' being labeled 'PR#2'), strongly suggests that the model's underlying generic predictions ('PR', 'B', 'L', 'PO') are largely accurate. The failure point appears to be the renumbering algorithm, which is not robust

Table 2 Test Set Classification Report for Specific Numbered Labels

Label	Precision	Recall	F1-score	Support
1#B	0.99	1.00	1.00	237,231
1#L	0.95	1.00	0.97	12,233
2#B	0.98	0.96	0.97	44,714
2#L	0.95	0.87	0.91	4,837
3#B	0.97	0.90	0.93	7,430
3#L	0.91	0.81	0.86	1,127
4#B	0.92	0.87	0.89	1,194
4#L	0.88	0.78	0.83	207
5#B	0.69	0.77	0.73	66
5#L	1.00	0.83	0.91	47
6#B	1.00	1.00	1.00	25
6#L	0.00	0.00	0.00	1
PO#1	0.98	1.00	0.99	180,165
PO#2	0.96	0.96	0.96	36,794
PO#3	0.94	0.91	0.93	7,374
PO#4	0.90	0.82	0.86	1,206
PO#5	0.68	0.58	0.63	115
PO#6	1.00	0.79	0.88	19
PR#1	0.99	1.00	1.00	799,492
PR#2	0.98	1.00	0.99	193,429
PR#3	0.94	0.93	0.93	36,163
PR#4	0.91	0.89	0.90	6,237
PR#5	0.78	0.87	0.82	505
PR#6	1.00	0.47	0.64	81
micro avg	0.99	0.99	0.99	1,570,692
macro avg	0.89	0.83	0.86	1,570,692
weighted avg	0.99	0.99	0.99	1,570,692

enough to handle all complex patient histories. As shown in the edge cases (Appendix A), sequences involving multiple losses or long gaps of non-specific care between pregnancy episodes can cause the state machine to fail to advance the episode counter correctly.

This finding is significant; it validates our two-stage approach and reframes the remaining challenge. Further substantial gains in accuracy are more likely to come from refining the deterministic post-processing rules rather than from tuning the deep learning model itself. The problem is thus shifted from one of probabilistic pattern recognition to one of creating a more sophisticated, state-aware algorithm for episode construction.

Table 3 False Negatives by True Label Class

True Label Class	False Negatives	True Label Class	False Negatives
1#B	108	PO#1	206
1#L	33	PO#2	1,332
2#B	1,687	PO#3	659
2#L	607	PO#4	218
3#B	716	PO#5	48
3#L	214	PO#6	4
4#B	158	PR#1	24
4#L	46	PR#2	829
5#B	15	PR#3	2,600
5#L	8	PR#4	680
6#B	0	PR#5	68
6#L	1	PR#6	43

Conclusion

In this paper, we have presented a two-stage deep learning framework to automatically identify and number pregnancy episodes from administrative claims data. Our approach successfully uses a custom Transformer-based model to predict generic event types, which are then processed by a deterministic algorithm to construct specifically numbered pregnancy episodes.

Evaluated on a large, real-world dataset, the model demonstrated high accuracy, particularly for the most common pregnancy scenarios. Our analysis reveals that the remaining errors are primarily systematic, stemming from the logical rules of the post-processing stage rather than the predictive capacity of the neural network. This highlights the effectiveness of decoupling pattern recognition from deterministic rule application.

Future research will prioritize the refinement of the episode renumbering algorithm and the application of these episodes to develop predictive models for outcomes such as birth or loss. We plan to explore more sophisticated state machines or potentially a secondary, simpler machine learning model dedicated to identifying the transition points between pregnancy episodes. Concurrently, these refined episodes will be leveraged as structured features for supervised learning models aimed at forecasting specific outcomes, such as live birth or pregnancy loss. By improving this post-processing component, we believe near-perfect episode delineation is an achievable goal, providing a powerful new tool for large-scale maternal health research.

Acknowledgments

We thank professor Lukasz Golab of the University of Waterloo for reading an earlier draft of this paper. His comments and suggestions have greatly improved the quality of this work.

Declarations

- Funding: This work was supported by the Faculty Opportunity Award funding from the University of Nevada, Las Vegas (UNLV) Research Foundation, the UNLV Foundation, and the Division of Research at UNLV. The research was also supported by the National Science Foundation, Grant No. 1625677
- Competing interests: None to declare
- Ethics approval: Not applicable
- Consent: Not applicable
- Code availability: https://github.com/UNLVCS/ml-apcd
- Maternity data for 2018-2022 were purchased from the Virginia APCD and paid by the Faculty Opportunity Award at the University of Nevada,Las Vegas.
- Authors' contributions: Fonseca, Vanderlaan, and Taghva contributed to conceptualization, design, and algorithmic approaches. The Implementation and experimentation were performed by Azar, Devkota, Farivar, and Ortega.

References

[1] Fink, D.A., Kilday, D., Cao, Z., Larson, K., Smith, A., Lipkin, C., Perigard, R., Marshall, R., Deirmenjian, T., Finke, A., Tatum, D., Rosenthal, N.: Trends in maternal mortality and severe maternal morbidity during delivery-related hospitalizations in the united states, 2008 to 2021. JAMA Network Open **6**(6), 2317641–2317641 (2023) https://doi.org/10.1001/jamanetworkopen.2023.17641

[2] Metcalfe, A., Wick, J., Ronksley, P.: Racial disparities in comorbidity and severe maternal morbidity/mortality in the united states: an analysis of temporal trends. Acta Obstetricia et Gynecologica Scandinavica **97**(1), 89–96 (2018) https://doi.org/10.1111/aogs.13245

[3] Leonard, S.A., Main, E.K., Lyell, D.J., Carmichael, S.L., Kennedy, C.J., Johnson, C., Mujahid, M.S.: Obstetric comorbidity scores and disparities in severe maternal morbidity across marginalized groups. American Journal of Obstetrics and Gynecology MFM **4**(2), 100530 (2022) https://doi.org/10.1016/j.ajogmf.2021.100530

[4] Li, C., Malapati, S.J., James, L.P., Hutchins, L.F.: Racial disparity in adherence to endocrine therapy among women with early-stage hormone receptor positive breast cancer: An analysis of arkansas all-payers claims database. Clinical Breast Cancer **24**(7), 647–6594 (2024) https://doi.org/10.1016/j.clbc.2024.07.009

[5] Rezaeiahari, M., Brown, C.C., Ali, M.M., Tilford, J.M., Magann, E.F.: Identification of racial/ethnic disparities associated with severe maternal morbidity using the arkansas all-payer claims database (2013-2017). Birth **50**(2), 339–348 (2023) https://doi.org/10.1111/birt.12659

[6] McWilliams, J.M., Meara, E., Zaslavsky, A.M., Ayanian, J.Z.: Use of health services by previously uninsured Medicare beneficiaries. New England Journal of

Medicine **357**(2), 143–153 (2007) https://doi.org/10.1056/NEJMsa067712

[7] APCD Council: All-payer claims databases: State initiatives to improve health care transparency. Technical report, National Association of Health Data Organizations and the University of New Hampshire (2011). https://www.apcdcouncil.org/sites/default/files/media/2023-05/databases-state-initiatives-improve-health-care-transparency.pdf

[8] McCarthy, D.: How states establish an apcd and make it functional. The Commonwealth Fund (2020) https://doi.org/10.26099/06qz-1m31

[9] Gish, R.G., Jacobson, I.M., Lim, J.K., Waters-Banker, C., Kaushik, A., Kim, C., Cyhaniuk, A., Wong, R.J.: Prevalence and characteristics of hepatitis delta virus infection in patients with hepatitis b in the unitedstates: An analysis of the all-payer claims database. Hepatology **79**(5), 1117–1128 (2024) https://doi.org/10.1097/HEP.0000000000000687

[10] Miller, P.B., Love, D., Sullivan, E., Porter, J., Costello, A.: All-payer claims databases: An overview for policymakers. Technical report, National Association of Health Data Organizations and the Regional All-Payer Healthcare Information Council (May 2010). Prepared for State Coverage Initiatives. https://scholars.unh.edu/cgi/viewcontent.cgi?article=1127&context=ihpp

[11] Gross, K., Brenner, J.C., Truchil, A., Post, E.M., Riley, A.H.: Building a citywide, all-payer, hospital claims database to improve health care delivery in a low-income, urban community. Population health management **16**(Suppl 1), 20–25 (2013) https://doi.org/10.1089/pop.2013.0037

[12] Vest, J.R., Gamm, L.D.: Health information exchange: persistent challenges and new strategies. Journal of the American Medical Informatics Association **17**(3), 288–294 (2010) https://doi.org/10.1136/jamia.2010.003673

[13] Jones, S.E., Bradwell, K.R., Chan, L.E., McMurry, J.A., Olson-Chen, C., Tarleton, J., Wilkins, K.J., Ly, V., Ljazouli, S., Qin, Q., Faherty, E.G., Lau, Y.K., Xie, C., Kao, Y.-H., Liebman, M.N., Mariona, F., Challa, A.P., Li, L., Ratcliffe, S.J., Haendel, M.A., Patel, R.C., Hill, E.L., Consortium, N.: Who is pregnant? defining real-world data-based pregnancy episodes in the national covid cohort collaborative (n3c). JAMIA Open **6**(3), 067 (2023) https://doi.org/10.1093/jamiaopen/ooad067

[14] Lyu, T., Liang, C., Liu, J., Campbell, B., Hung, P., Shih, Y.-W., Ghumman, N., Li, X., National COVID Cohort Collaborative Consortium: Temporal events detector for pregnancy care (ted-pc): A rule-based algorithm to infer gestational age and delivery date from electronic health records of pregnant women with and without covid-19. PLOS ONE **17**(10), 1–22 (2022) https://doi.org/10.1371/journal.pone.0276923

[15] Azar, S.E., Devkota, Y., Farivar, S., Ortega, J., Vanderlaan, J., Taghva, K.: Transforming Insurance Claims into Maternal Care Insights (2025)

[16] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics, Stroudsburg, PA (2019). https://doi.org/10.18653/v1/n19-1423

[17] Li, Y., Rao, S., Solares, J.R.A., Hassaine, A., Ramakrishnan, R., Canoy, D., Zhu, Y., Rahimi, K., Salimi-Khorshidi, G.: Behrt: Transformer for electronic health records. Scientific Reports **10**(1), 7155 (2020) https://doi.org/10.1038/s41598-020-62922-y

[18] Dodge, J., Ilharco, G., Schwartz, R., Farhadi, A., Hajishirzi, H., Smith, N.: Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping (2020). https://doi.org/https://arxiv.org/abs/2002.06305

[19] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2017). https://api.semanticscholar.org/CorpusID:53592270

Appendix A Edge Case Examples

An appendix contains supplementary information that is not an essential part of the text itself but provides a more comprehensive understanding of the research problem.

A.1 Case 1: Person 1012766775544089922

Table A1 demonstrates an error where a true 2#L is mislabeled as 1#L, causing the subsequent pregnancy (PR#3 and 3#B) to be incorrectly labeled as episode #2.

Table A1 Case 1: Example of error propagation due to mislabeled pregnancy loss

Date	Event Code	Original Label	Predicted Label
...	...	...	...
18-06-2020	D_E039	2#L	1#L
18-06-2020	D_O021	2#L	1#L
26-10-2020	D_O09299	PR#3	PR#2
...	...	...	...
01-07-2021	D_Z6711	3#B	2#B
...	...	...	...

A.2 Case 2: Person 7942822445536094180

Table A2 shows an error propagating through multiple episodes. A true 4#L is mislabeled 3#L, which causes all subsequent episodes (PO#4, 5#L, PR#6, etc.) to be offset by one.

Table A2 Case 2: Example of error propagating through multiple subsequent episodes

Date	Event Code	Original Label	Predicted Label
...	...	...	...
02-04-2021	D_O021	4#L	3#L
09-04-2021	D_Z3201	PO#4	PO#3
...	...	...	...
06-07-2021	D_O021	5#L	3#L
...	...	...	...
19-10-2021	D_N979	PR#6	PR#4

Note: Mislabeled 4#L leads to an offset in all subsequent predicted episode numbers.