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.

The Claims Floor: Limits of Administrative Claims Data for Dynamic Prediction of Pregnancy Loss and Stillbirth

Sobhan Ebrahimi Azar
University of Nevada, Las Vegas
Jennifer Vanderlaan
University of Nevada, Las Vegas
Mohsen Larni
University of Nevada, Las Vegas
Kazem Taghva
kazem.taghva@unlv.edu
University of Nevada, Las Vegas

Research Article

Keywords: pregnancy loss, stillbirth, administrative claims, competing risks, discrete-time survival, dynamic prediction, data leakage, all-payer claims database

Posted Date: September 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10781005/v1

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

Additional Declarations: No competing interests reported.

The Claims Floor: Limits of Administrative Claims Data for Dynamic Prediction of Pregnancy Loss and Stillbirth

Sobhan Ebrahimi Azar[1], Jennifer Vanderlaan[2], Mohsen Larni[1], Kazem Taghva[1*]

[1]Department of Computer Science, University of Nevada, Las Vegas, 4505 S Maryland Pkwy, Las Vegas, 89154, NV, US.
[2]School of Nursing, University of Nevada, Las Vegas, 4505 S Maryland Pkwy, Las Vegas, 89154, NV, US.

*Corresponding author(s). E-mail(s): kazem.taghva@unlv.edu;
Contributing authors: sobhan.ebrahimiazar@unlv.edu;
jennifer.vanderlaan@unlv.edu; mohsen.larni@unlv.edu;

Abstract

Background. Dynamic, gestation-anchored risk prediction could give clinicians lead time to act on pregnancy loss and stillbirth. Administrative claims databases are attractive substrates for such models because of their scale and population coverage, but whether claims support above-chance dynamic prediction of these outcomes at clinically actionable gestational time points is unestablished.
Methods. Using the Virginia All-Payer Claims Database (2018–2023; 344,560 anchored pregnancy episodes), we built a discrete-time competing-risks survival framework distinguishing live birth, early pregnancy loss ($<$20 weeks), and stillbirth ($\geq$20 weeks), excluding induced termination. A gestational clock was anchored via ICD-10-CM Z3A codes; episodes were expanded to person-periods; predictions were made at gestational landmarks with a two-week gap/blanking window to prevent peri-event leakage. An evidence-based, code-level feature set was compared against a broad 3-character-prefix set using penalized discrete-time models, and evaluated with competing-risks cumulative incidence, time-dependent AUC, IPCW Brier / IPA, and calibration, under patient-level random and temporal splits.
Results. Three findings emerged. Naive weekly-hazard models were severely inflated by same-week coding contamination: early-loss AUC fell from **0.833** to **0.61** once a two-week gap was enforced, with the events in the cell dropping

from **378** to **48**. An apparent "worse-than-nothing" index of prediction accuracy proved to be a two-part artifact (a per-period-versus-cumulative scale mismatch together with rare-cause magnitude inflation from balanced class weights), and resolved fully under proper competing-risks cumulative-incidence computation and recalibration (corrected IPA $\approx$ **0**). With contamination removed and evidence-based features properly constructed, neither outcome was forecastable above chance at actionable landmarks: stillbirth time-dependent AUC peaked at **0.56** with a flat discrimination curve across gestation, and broad granularity did not help (**0.52**). A marker of recognized high-risk surveillance (ICD-10-CM O09) was *anti*-predictive, consistent with effective surveillance biasing toward live outcomes.

Conclusions. In this APCD extract, linear competing-risks models on properly constructed claims features did not predict early loss or stillbirth above chance at clinically actionable gestational time points. The result delineates a "claims floor" that contrasts with the biophysical ceiling of ultrasound/Doppler-based models, because the antecedents that carry predictive signal are coded at the event rather than antenatally. Beyond the substantive null, we provide two reusable methodological cautions for claims-based dynamic prediction.

Keywords: pregnancy loss, stillbirth, administrative claims, competing risks, discrete-time survival, dynamic prediction, data leakage, all-payer claims database

1 Introduction

Static, single-time-point obstetric risk tools do not capture the evolving nature of gestation. Dynamic prediction anchored to gestational age is attractive because it could open a clinically actionable window: low-dose aspirin confers most benefit when begun before 16 weeks, and antenatal fetal surveillance for recognized risk conditions is generally initiated between 32 and 36 weeks (American College of Obstetricians and Gynecologists Committee Opinion 828 [1]). A model that updates risk as a pregnancy progresses could, in principle, flag deterioration while intervention is still possible.

All-payer claims databases (APCDs) are an appealing substrate for such models because they offer population-scale, multi-payer coverage well suited to risk stratification at the population and care-management level. The present study builds directly on a validated Virginia APCD maternal-care foundation: a reproducible, benchmark-validated rule-based pipeline that extracts pregnancy episodes [2], a complementary transformer model that performs episode-level identification and parity-aware numbering [3], and a birth-versus-loss outcome classifier whose stated aim is a system for predicting at-risk deliveries [4]. Claims nonetheless trail real time by an adjudication lag and, in this extract, carry no paid date; the models we evaluate are therefore dynamic in *gestational* time rather than wall-clock time.

What remains unestablished is whether *time-varying* claims support above-chance *dynamic* prediction of loss or stillbirth at the gestational landmarks where a clinician could still act. The next section reviews the substrate, the prior prediction literature, and the methodological tools on which this study builds.

This paper reports a substantive null together with two methodological cautions that accompany it. The first caution concerns a leakage mechanism, same-week coding contamination, that inflates naive dynamic claims models, whose magnitude we quantify and whose remedy we demonstrate. The second concerns an apparent "worse-than-nothing" calibration result, which we decompose and show to be an artifact of competing-risks cumulative-incidence computation combined with class weighting rather than genuine miscalibration. The substantive contribution is that, once contamination is removed and an evidence-based feature set is properly constructed, neither early loss nor stillbirth is forecastable above chance at actionable landmarks, a "claims floor" that we frame against the biophysical prediction ceiling.

2 Background

All-payer claims databases aggregate enrollment, medical, and, where available, pharmacy records from public and private payers across a state or region, and have become a standard substrate for population-scale maternal-health research [2]. Their appeal is breadth. Unlike hospital discharge files, which capture only the delivery encounter, claims follow a member longitudinally across outpatient, inpatient, and professional settings, and they encode clinical history in standardized diagnosis and procedure vocabularies. That breadth comes with structural limits: coverage is incomplete for self-insured plans, race and ethnicity are recorded inconsistently and largely through the payer, residential geography is often absent, and the lack of a paid or adjudication date in many extracts means that claims describe gestational rather than wall-clock time [5]. These properties shape both what can be measured and what can be predicted.

Claims do not mark pregnancies explicitly, so episodes must be inferred from diagnosis, procedure, and gestational-age codes. The methodological literature is dominated by deterministic, rule-based pipelines that combine these signals with temporal heuristics to delineate prenatal, delivery, and postpartum windows [6, 7], and comparative work has shown that the choice of ascertainment algorithm materially changes reported outcome rates [8]. We rely on a reproducible rule-based pipeline for the Virginia APCD that derives gestational age from ICD-10-CM Z3A codes, validates birth and loss terminations with temporal exclusion windows, and reproduces state-level birth and loss benchmarks [2]; a complementary transformer model extends this to episode-level identification and parity-aware numbering for higher-order pregnancies [3]. A separate step classifies each validated episode as a birth or a loss, prioritizing recall on the rare loss class so that adverse outcomes are not misread as healthy deliveries [4]. Together these establish the cohort, the gestational clock, and the terminal labels used in the present study.

Prediction of stillbirth and pregnancy loss has been pursued mainly on clinical and registry data rather than on claims. Machine-learning models trained on demographic and clinical features report high discrimination, with tree-based and neural models reaching areas under the curve above 0.85 for stillbirth before 24 weeks [9] and, in large national-registry studies, above 0.90 when rich antenatal measurements are available [10, 11]. These results depend on variables that claims do not contain.

The most informative models add biophysical measurement: a validated first- and second-trimester model combining maternal factors with ultrasound biometry and uterine-artery Doppler reaches an area under the curve of approximately 0.84 [12]. Models restricted to maternal history and prior outcomes, which is closer to what claims encode, reach more modest discrimination of roughly 0.64 to 0.66 [13], consistent with the recurrence structure of loss and stillbirth documented in large cohorts [14, 15]. Validated claims-based phenotypes for loss and stillbirth exist [16], but they are designed to ascertain outcomes after the fact rather than to forecast them.

The clinical value of any antenatal prediction depends on whether it arrives in time to act. Low-dose aspirin reduces preterm preeclampsia when started before 16 weeks [17], and antenatal fetal surveillance for recognized risk conditions is generally initiated between 32 and 36 weeks [1]. A useful dynamic model must therefore discriminate at these gestational landmarks, not merely at or after the event.

Methodologically, our analysis sits within discrete-time competing-risks survival modeling, in which a pregnancy can end in distinct, mutually exclusive ways and time is measured in gestational weeks. Dynamic, landmark-based prediction updates risk at successive gestational time points and is the natural framework for this question; deep-learning approaches such as DeepHit and its dynamic extension formalize competing-risks prediction from longitudinal data [18, 19], while inverse-probability-of-censoring-weighted scoring rules provide proper accuracy and calibration measures under censoring [20]. Because claims-based models often treat diagnosis codes as a vocabulary, the statistical structure of that vocabulary matters: ICD-10-CM codes in this corpus follow a heavy-tailed, non-Zipfian distribution with weak head dominance, which informs how features should be thresholded and weighted [21].

What is missing from this literature is a direct test of whether time-varying claims, as opposed to biophysical measurements or static maternal history, support above-chance dynamic prediction of loss or stillbirth at clinically actionable landmarks. That is the question we address.

3 Methods

3.1 Data source

We used the Virginia APCD for 2018–2023, restricted to the ICD-10-CM era. The extract comprises professional and facility claims; it contains no pharmacy file, and records carry only a service date (`Incurred_Date`), with no adjudication or paid date. This study was determined to be exempt from review by the University of Nevada, Las Vegas Institutional Review Board because the study uses de-identified data and therefore does not meet the definition of human subjects research.

3.2 Cohort and outcome definitions

Pregnancy episodes were identified using the established, benchmark-validated rule-based pipeline [2], which matches reversal claims, collapses ICD fields, parses gestational age from Z3A codes, applies birth and loss validity windows, and truncates

at 1 January 2018; a transformer-based model for episode identification and parity-aware numbering offers a learned alternative for the same delineation task [3]. Each episode was assigned one of four competing-risks labels: birth, early loss (before 20 weeks), stillbirth (20 weeks or later), or censored. The terminal birth-versus-loss distinction was taken from the validated outcome classifier of the same pipeline [4] and then resolved further by gestational age.

Induced and elective terminations were excluded (O04; O05 and O06 were confirmed empty in the Virginia APCD; Z33.2). Spontaneous abortion (O03) was treated as early loss, and intrauterine fetal death (O36.4) was resolved to loss and then split by gestational age. The gestational clock was anchored to the last menstrual period, estimated from the most plausible Z3A code as the termination date minus the maximum plausible Z3A week expressed in days. An episode entered the anchored cohort only if it carried at least one plausible Z3A code; episodes with only the gestational-age-of-one sentinel (Z3A.01), with no Z3A code, or representing induced termination formed an explicit excluded arm.

3.3 Discrete-time competing-risks framework

Each episode was expanded into person-periods, one row per gestational week at risk, with left truncation at the first anchoring claim and right censoring at the end of enrollment. We modeled cause-specific discrete-time hazards, $h_k(t) = P(\text{event of cause } k \text{ in week } t \mid \text{at risk at } t)$. Dynamic prediction was framed as a landmark–horizon problem: at a landmark week ℓ we predicted the cumulative incidence of each cause over the interval $(\ell + g, \ell + H]$, using a gap of $g = 2$ weeks as a blanking window so that coding occurring around the event itself could not enter the prediction. Episodes experiencing an event within the gap, $(\ell, \ell + g]$, were removed from the landmark risk set. Cumulative incidence was computed in the proper competing-risks form rather than as $1 - \prod(1 - h_k)$,

$$\mathrm{CIF}_k(\ell, \ell + H) = \sum_{t=\ell+1}^{\ell+H} S(t-1)\, h_k(t), \qquad S(t-1) = \prod_{u=\ell+1}^{t-1} \Big(1 - \textstyle\sum_j h_j(u)\Big), \quad (1)$$

where $S(t-1)$ is overall survival to week $t-1$.

3.4 Features

Baseline features, known at or before the anchor week, comprised maternal age in bands, prior loss, prior stillbirth, pre-existing diabetes and chronic hypertension constructed from pre-LMP claims (a true baseline, though under-ascertained near the 2018 data floor), obesity, and smoking. Payer and provider three-digit ZIP were carried as explicitly flagged socioeconomic and access proxies rather than as neighborhood measures.

The evidence-based, time-updated feature set (Set A) consisted of conditions with an established association with adverse pregnancy outcome: fetal growth restriction (O36.5), preeclampsia (O14/O11), gestational hypertension (O13/O16), gestational diabetes (O24.4), cholestasis (O26.6), oligohydramnios (O41.0), multifetal gestation

(O30/O31), and high-risk pregnancy supervision (O09). Each was encoded as binary cumulative presence from its onset week rather than as a count, so that the sheer volume of coding could not act as a proxy for acuity. Acute peri-event codes (eclampsia O15, placenta previa O44, and abruption O45) were held out of the headline model and examined only in a labeled sensitivity analysis.

To test whether a richer vocabulary would help, a second feature set (Set B) used every ICD-10-CM three-character prefix appearing in at least 50 cohort episodes, again encoded by onset week. The prevalence floor and the choice of presence over count encoding reflect the heavy-tailed, non-Zipfian frequency structure of ICD-10 codes in this corpus: a flat rank–document-frequency curve with weak head dominance means that low-frequency codes dominate the vocabulary and that raw counts are unstable as feature weights [21].

To prevent the outcome from leaking into the predictors, outcome-defining codes (O03–O06, O36.4, P95, Z37.x, Z33.2, and P96.4) were removed at the level of the individual claim before any aggregation, and the resulting feature matrix was rechecked to confirm that none remained.

3.5 Models and evaluation

Cause-specific hazards were fit with penalized (elastic-net) discrete-time models. We did not apply balanced class weights, for reasons developed under the second finding below; a penalized discrete-time model serves both as the natural baseline and as a sensitive indicator of any residual leakage. Discrimination was summarized by the landmark time-dependent AUC (AUC_{td}), the quantity directly comparable to the literature target of about 0.65. Accuracy and calibration were assessed with the inverse-probability-of-censoring-weighted (Graf) Brier score and the index of prediction accuracy (IPA) relative to an Aalen–Johansen marginal null, followed by isotonic recalibration fit on the validation set and applied to the test set. We evaluated two splitting schemes: a patient-level random split as the primary analysis and a temporal split (training on 2018–2021, validating on 2022, testing on 2023), both grouped by member so that no patient appeared in more than one set. Calibration was prioritized over discrimination, and any landmark–horizon cell with fewer than 50 events was flagged as too thin to interpret.

4 Results

4.1 Cohort

The anchored cohort comprised 344,560 pregnancy episodes: births (327,135; 94.9%), early losses (14,373; 4.2%), and stillbirths (3,052; 0.9%). Anchoring was, however, highly selective: 72.3% of raw losses were excluded (30,235 for having no Z3A code, 14,317 for carrying only the Z3A.01 sentinel, and 545 for being induced terminations), so that only 28.6% of losses could be anchored. This selection is the dominant bound on the external validity of everything that follows (Table 1).

4.2 Same-week contamination inflates naive dynamic models

On the legacy window $[\ell, \ell + H]$, early-loss discrimination at $\ell = 8$, $H = 4$ appeared strong, with an AUC of 0.833 under the random split and 0.798 under the temporal split. Enforcing the two-week gap collapsed these values to 0.61 and 0.54, and the number of events in the cell fell from 378 to 48. In other words, roughly 87% of the losses the model appeared to "predict" were occurring within two weeks of the landmark: the codes were recording a loss already in progress rather than forecasting a future one (Table 2).

4.3 Negative IPA is an artifact, not a signal

The IPA was at first strongly negative (for stillbirth at $\ell = 24$, $H = 8$ it reached -38.3), which would naively suggest a model far worse than useless. The effect decomposed into two parts. The first was a scale mismatch between a per-week hazard and a cumulative label, which the proper competing-risks cumulative-incidence accumulation corrected, moving the IPA to -5.9. The second was a magnitude inflation of the rare cause introduced by balanced class weighting (roughly 2000-fold for stillbirth), which isotonic recalibration corrected, bringing the IPA to approximately zero. After both corrections the model sat at an IPA of about zero with an AUC near 0.5: it matched the Aalen–Johansen marginal and added no incremental accuracy (Figure 1).

4.4 The substantive null at actionable landmarks

With contamination removed and the evidence-based features properly constructed, stillbirth discrimination peaked at an AUC_{td} of 0.56 (the best non-thin cell, with O09 removed) and was 0.46 and 0.39 at the actionable landmarks of 28 and 32 weeks; every confidence interval straddled 0.5, and the IPA remained near zero throughout. The broad-vocabulary Set B peaked at 0.515, so granularity was not the limiting factor. Critically, the discrimination curve was flat across gestation: it did not rise toward 28 and 32 weeks, where a genuine placental-dysfunction signal should strengthen as growth restriction and preeclampsia accrue (Table 3). Early loss was similarly unforecastable from time-varying features at clean horizons; only baseline maternal factors carried a modest early signal, with an AUC of about 0.72 at $\ell = 8$ decaying to chance by $\ell = 16$. Because roughly 47% of early losses carry unreliable event timing (an artifact of anchoring when the terminal gestational age is missing), early loss is reported here as a baseline-only finding.

4.5 The O09 inversion

High-risk pregnancy supervision (O09) behaved in a way that is itself informative: it carried a negative coefficient, and removing it improved discrimination by 0.04 to 0.09 across landmarks (Table 4). O09 marks pregnancies that entered surveillance and continued under management: pregnancies whose risk had already been recognized and acted upon, so the marker biases toward live outcomes. The one explicitly encoded clinical risk judgment available in claims is thus inverted by the very intervention it triggers.

5 Discussion

The central result is not that stillbirth is unpredictable. Biophysical models reach an AUC of approximately 0.84 [12] precisely because they draw on ultrasound biometry and uterine-artery Doppler. Our analysis establishes the floor on the claims side, and the distance between that floor and the biophysical ceiling is itself the finding: it locates where the predictive signal actually lives.

The mechanism behind the floor is structural. The antecedents that carry signal (growth restriction, preeclampsia) are coded at the encounter where the resulting outcome is managed rather than antenatally; a two-week prevalence check found no antenatal enrichment of O36.5 or O14 among stillbirths. And the single available marker of recognized risk, O09, is inverted by the effective surveillance it denotes. Both are properties of claims as a prediction substrate, not artifacts of a particular model.

Two methodological cautions generalize beyond this dataset. First, same-week coding contamination silently inflates naive dynamic claims models, so a gap or blanking window, together with claim-level removal of outcome-defining codes before aggregation, is a necessary safeguard. Second, a negative IPA in a discrete-time competing-risks model can be an artifact of cumulative-incidence computation combined with class weighting rather than evidence of genuine miscalibration, and should be investigated as such before it is reported.

These results also clarify the relationship to prior claims-based work. Even the modest maternal-history discrimination band of about 0.64–0.66 [13] was not reached dynamically here, which is consistent with those models being static or baseline rather than time-varying claims predictors. The distinction is not that claims lack outcome signal: retrospective birth-versus-loss classification on this same APCD is highly accurate once the full episode has been observed [4]. The present null concerns the strictly harder, prospective task of forecasting the outcome from antecedent codes at a gestational landmark. That is signal the retrospective classification does not, and cannot, supply.

6 Limitations

The scope of the claim is deliberately narrow: it concerns linear competing-risks models, this particular APCD extract, and claims-derived features, and it is not a claim about non-linear models, other states, or richer data sources. Selection is the most important limitation, as the 72.3% exclusion of losses leaves an anchored cohort skewed toward later gestation and higher acuity. Because the extract carries only service dates and no adjudication date, the models are dynamic in gestational time rather than wall-clock time. The data contain no member residential ZIP (provider three-digit ZIP is only a service-area proxy), race is recorded incompletely and in a payer-dependent way and is reliable only within the Medicaid subgroup, and there is no mother–infant link. Finally, the analysis covers a single state and the ICD-10-CM era only.

Table 1 Cohort composition and loss exclusion.

	N	%
Anchored cohort (total)	344,560	100
Birth	327,135	94.9
Early loss (<20 wk)	14,373	4.2
Stillbirth (≥20 wk)	3,052	0.9
Losses excluded (of raw losses)	45,097	72.3
No Z3A	30,235	
Z3A.01 sentinel only	14,317	
Induced termination	545	

7 Conclusion

In the Virginia APCD, properly constructed, leakage-controlled, evidence-based claims features did not support above-chance dynamic prediction of early pregnancy loss or stillbirth at clinically actionable gestational time points using linear competing-risks models. The study delineates a claims floor that contrasts with the biophysical prediction ceiling, explains the floor mechanistically through antecedents coded at the event and through the inversion of the surveillance marker, and contributes two reusable methodological cautions for dynamic prediction on administrative data.

Declarations

- Funding - This work was supported by the Faculty Opportunity Award funding from the UNLV Research Foundation, the UNLV Foundation, and the Division of Research at the University of Nevada, Las Vegas.
- Conflict of interest/Competing interests: The authors have no conflicts of interest to declare.
- Ethics approval and consent to participate - This study was exempt from IRB review.
- Consent for publication: Not applicable.
- Data availability - Data used in this project can be purchased for research through Virginia Health Information.
- Author contribution: Vanderlaan and Taghva contributed to conceptualization, design, and algorithmic approaches. The Implementation and experimentation were performed by Ebrahimi Azar and Larni.

Table 2 Same-week contamination: early loss, $\ell = 8$, $H = 4$.

Window	Split	AUC	Events in cell
$[\ell, \ell + H]$ (legacy)	random	0.833	378
$(\ell + 2, \ell + H]$ (gap)	random	0.607	48
$[\ell, \ell + H]$ (legacy)	temporal	0.798	308
$(\ell + 2, \ell + H]$ (gap)	temporal	0.540	51

Table 3 Stillbirth four-landmark discrimination (AUC_{td}, test, isotonic-recalibrated; Set A headline). Thin cells (<50 events) omitted.

ℓ	H	Split	AUC_{td} [95% CI]	Events
20	8	random	0.489 [0.425–0.545]	90
24	8	random	0.495 [0.422–0.566]	63
28	8	random	0.469 [0.411–0.521]	74
32	8	random	0.457 [0.399–0.503]	102
20	8	temporal	0.417 [0.350–0.494]	60
24	8	temporal	0.482 [0.415–0.562]	56
28	8	temporal	0.393 [0.316–0.460]	58
32	8	temporal	0.425 [0.366–0.485]	77
Set B (granularity check), best non-thin: 0.515 [0.464–0.576] ($\ell = 28$, random).				

Table 4 O09 inversion. Set A test AUC_{td} with vs. without O09 (random arm).

ℓ	H	With O09	Without O09	Δ
24	8	0.495	0.560	+0.065
28	4	0.477	0.567 (thin)	+0.090
28	8	0.469	0.513	+0.044
32	8	0.457	0.503	+0.046

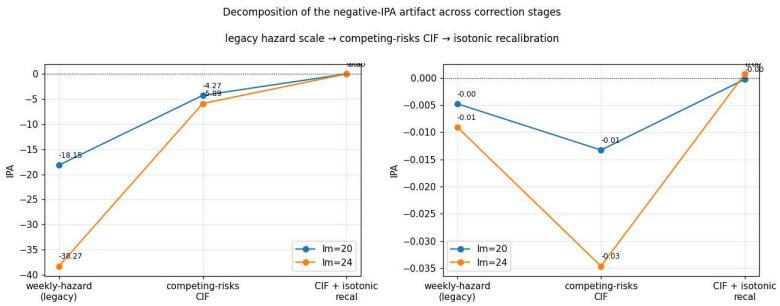

Fig. 1 Decomposition of the negative-IPA artifact, from the raw value through the competing-risks cumulative-incidence correction to the recalibrated result.

References

[1] American College of Obstetricians and Gynecologists: Indications for outpatient antenatal fetal surveillance (Committee Opinion 828). Obstetrics & Gynecology **137**(6), 177–197 (2021)

[2] Ebrahimi Azar, S., Farivar, S., Devkota, Y., Ortega, J., Vanderlaan, J., Taghva, K.: Transforming Insurance Claims into Maternal Care Insights. To appear (2026)

[3] Devkota, Y., Ebrahimi Azar, S., Farivar, S., Ortega, J., Vanderlaan, J., Fonseca, J., Taghva, K.: AI-Assisted Pregnancy Episode Identification from All Payers Claim Dataset. To appear (2026)

[4] Ebrahimi Azar, S., Devkota, Y., Farivar, S., Vanderlaan, J., Taghva, K.: Birth and loss classification based on maternal insurance claims data. Health Information Science and Systems (2026). In press, accepted for publication

[5] Kranz, A.M., Dworsky, M., Ryan, J., Heins, S.E., Bhandarkar, M.: State all-payer claims databases: Identifying challenges and opportunities for conducting patient-centered outcomes research and multi-state studies. Technical report, RAND Health Care, for the Office of the Assistant Secretary for Planning and Evaluation (ASPE), Washington, DC (October 2023)

[6] Matcho, A., Ryan, P., Fife, D., Gifkins, D., Knoll, C., Friedman, A.: Inferring pregnancy episodes and outcomes within a network of observational databases. PLoS One **13**(2), 0192033 (2018)

[7] Naleway, A.L., Gold, R., Kurosky, S., Riedlinger, K., Henninger, M.L., Nordin, J.D., Kharbanda, E.O., Irving, S.A., Cheetham, T.C., McCarthy, N.L.: Identifying pregnancy episodes, outcomes, and mother–infant pairs in the Vaccine Safety Datalink. Vaccine **31**(27), 2898–2903 (2013)

[8] Margulis, A.V., Huybrechts, K.F.: Identification of pregnancies in healthcare data: A changing landscape. Pharmacoepidemiology and Drug Safety **32**(1), 84–86 (2023)

[9] Cersonsky, T.E., Ayala, N.K., Pinar, H., Dudley, D.J., Saade, G.R., Silver, R.M., Lewkowitz, A.K.: Identifying risk of stillbirth using machine learning. American Journal of Obstetrics and Gynecology **229**(3), 327–132716 (2023) https://doi.org/10.1016/j.ajog.2023.06.017

[10] Khatibi, T., Hanifi, E., Sepehri, M.M., Allahqoli, L.: Proposing a machine-learning based method to predict stillbirth before and during delivery and ranking the features: nationwide retrospective cross-sectional study. BMC Pregnancy and Childbirth **21**(1), 202 (2021)

[11] Koivu, A., Sairanen, M.: Predicting risk of stillbirth and preterm pregnancies

with machine learning. Health Information Science and Systems **8**(1), 14 (2020)

[12] Ashoor, G., Syngelaki, A., Papastefanou, I., Nicolaides, K.H., Akolekar, R.: Development and validation of model for prediction of stillbirth at 19–24 weeks. Ultrasound in Obstetrics & Gynecology **59**(1), 61–68 (2022)

[13] Trudell, A.S., Tuuli, M.G., Colditz, G.A., Macones, G.A., Odibo, A.O.: A stillbirth calculator: development and internal validation of a clinical prediction model. PLoS One **12**(3), 0173461 (2017)

[14] Magnus, M.C., Wilcox, A.J., Morken, N.-H., Weinberg, C.R., Håberg, S.E.: Role of maternal age and pregnancy history in risk of miscarriage. BMJ **364**, 869 (2019)

[15] Lamont, K., Scott, N.W., Jones, G.T., Bhattacharya, S.: Risk of recurrent stillbirth: systematic review and meta-analysis. BMJ **350**, 3080 (2015)

[16] Andrade, S.E., Shinde, M., Moore Simas, T.A., Bird, S.T., Bohn, J., Haynes, K., Taylor, L.G., Lauring, J.R., Longley, E., McMahill-Walraven, C.N., *et al.*: Validation of an icd-10-based algorithm to identify stillbirth in the sentinel system. Pharmacoepidemiology and Drug Safety **30**(9), 1175–1183 (2021)

[17] Rolnik, D.L., Wright, D., Poon, L.C., *et al.*: Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia (ASPRE). New England Journal of Medicine **377**(7), 613–622 (2017)

[18] Lee, C., Zame, W.R., Yoon, J., Schaar, M.: DeepHit: a deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI Conference on Artificial Intelligence (2018)

[19] Lee, C., Yoon, J., Schaar, M.: Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Transactions on Biomedical Engineering **67**(1), 122–133 (2020)

[20] Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M.: Assessment and comparison of prognostic classification schemes for survival data. Statistics in Medicine **18**(17–18), 2529–2545 (1999)

[21] Ebrahimi Azar, S., Vanderlaan, J., Taghva, K.: Revisiting Zipf's law in ICD-10 diagnosis code distributions. In: The International Conference on Innovations in Computing Research (2026). Springer. In press