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.

Early Detection of Maternal Supraphysiological Hypercholesterolemia: Development and Internal Validation of a Machine Learning Model and Characterization of First- Trimester Atherogenic Biomarkers

Ramon Serra
Hospital FF.AA Cirujano Guzmán
Claudette Cantin
San Sebastián University
David Medina-Ortiz
Universidad de Magallanes
Alejandro Sepúlveda-Peñaloza
Universidad de Los Andes, Chile
Andrea Morales
San Sebastián University
Vanessa Ríos
San Sebastián University
Reyna Peñailillo
Universidad de Los Andes, Chile
Lara J. Monteiro
Universidad de Los Andes, Chile
Sean W.D Carter
National University of Singapore
Matthew W. Kemp
National University of Singapore
Sebastián E. Illanes
Universidad de Los Andes, Chile
Andrea Leiva
andrea.leiva@uss.cl

San Sebastián University

Research Article

Keywords: maternal hypercholesterolemia, machine learning, pregnancy, first-trimester prediction, lipid metabolism, cardiovascular risk, explainable artificial intelligence

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11193070/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Maternal supraphysiological hypercholesterolemia (MSPH), defined by total cholesterol ≥ 290 mg/dL at term in women without pre-existing dyslipidemia, is associated with a more atherogenic profile and adverse cardiovascular consequences for mother and offspring. Because MSPH is recognized only late in pregnancy, early markers identifying at-risk women are lacking. We aimed to develop and internally validate supervised machine learning models using first-trimester clinical and biochemical measurements to predict MSPH at term, and to characterize early atherosclerosis-related biomarkers in women who developed MSPH.

Methods

This prospective observational cohort included 160 women with singleton, uncomplicated pregnancies recruited at two prenatal centers in Santiago, Chile. Predictors comprised 11 first-trimester anthropometric, metabolic, hemodynamic, and lipid-related variables. ApoB, ApoA1, oxidized LDL, and malondialdehyde were measured in a subgroup (n = 78). Third-trimester lipid measurements, used to define MSPH, were excluded as predictors. A complete-case dataset (n = 148) and a median-imputed full-cohort (n = 160) were evaluated. Multiple classifier families, preprocessing strategies, and internal-validation schemes, including stratified train-test splits, k-fold cross-validation, and leave-one-out validation, were benchmarked within leakage-controlled pipelines. Stochastic partitioning strategies were evaluated across fixed random seeds, and model selection prioritized predictive performance, stability, and parsimony. Interpretability was assessed using SHAP and permutation importance.

Results

Of 160 women, 44 (27.5%) developed MSPH. First-trimester total and LDL cholesterol showed the strongest between-group differences, while multivariable exploratory analyses indicated substantial overlap between groups. The selected Linear Discriminant Analysis pipeline achieved a ROC-AUC of 0.77 and a PR-AUC of 0.52. At the fixed operating threshold of 0.60, sensitivity was 0.62, specificity 0.79, accuracy 0.75, and MCC 0.40. Outputs were interpreted as discrimination scores rather than calibrated absolute risk estimates. SHAP and permutation importance identified first-trimester total cholesterol as the dominant predictor. Women who subsequently developed MSPH also showed higher first-trimester ApoB and malondialdehyde.

Conclusions

First-trimester clinical and lipid measurements contain discriminatory information for subsequent MSPH, with total cholesterol providing the strongest contribution. Atherogenic and oxidative alterations are also detectable in the first trimester of pregnancies that later develop MSPH. These findings support further evaluation of interpretable first-trimester risk assessment approaches; external validation is required before clinical application.

Trial registration

: Not applicable. Analytical code: https://github.com/kren-ai-lab/MSPH_classification_models.

Introduction

Hypercholesterolemia (HC), a condition characterized by high levels of total cholesterol (TC) and low-density lipoproteins (LDL) in plasma, is a well-known risk factor for the development and progression of atherosclerotic cardiovascular disease, the leading cause of death worldwide [1]. Pregnancy is a unique state marked by substantial metabolic adaptations, including alterations in lipid metabolism to accommodate the cholesterol requirements of the developing fetus [2, 3]. Despite maternal cholesterol levels exceeding those observed in non-pregnant women, particularly during the third trimester (T3) [4, 5], maternal HC is widely recognized as a physiological condition during pregnancy; however, its contribution to long-term dyslipidemia and cardiovascular risk remains unclear [6–10].

Over the past decade, several studies have examined the impact of elevated cholesterol during pregnancy, identifying two groups based on TC levels at term. Women with TC levels below 290 mg/dL or <p75 are classified as having maternal physiological hypercholesterolemia (MPH), whereas those with TC levels at or above 290 mg/dL or ≥p75 are classified as having maternal supraphysiological hypercholesterolemia (MSPH) [5]. Large-scale studies to determine the exact prevalence of MSPH are lacking, as TC is not routinely measured during pregnancy. Nonetheless, different studies have shown that a significant proportion of pregnant women (25–35%) develop this condition at the end of pregnancy [5, 11].

Deleterious consequences of MSPH during pregnancy have been reported both in the mother and her offspring. Specifically, this condition has been linked to endothelial dysfunction in the placental macro- [5, 11, 12] and micro-vasculature [13], and it can lead to cholesterol accumulation and changes in its transfer across the placenta [14]. MSPH has also been found to increase oxidative stress markers in the placenta as well as in the neonatal and maternal circulation [8, 12, 15]. Furthermore, it has been reported to increase levels of proprotein convertase subtilisin kexin 9 (PCSK9), oxidized LDL (oxLDL), and lipoproteins with a more significant pro-atherogenic profile in the maternal circulation at delivery [8, 16]. Maternal circulation has also been shown to have elevated markers of endothelial dysfunction and increased cardiovascular risk factors such as increased lipid peroxidation and apolipoprotein B (ApoB) [16]. Interestingly, follow-up studies have shown that infants born to mothers with a history of MSPH exhibit more extensive atheromatous lesions at birth [17] and later in life [18], with a higher propensity to

develop severe and complex cardiovascular disease in adulthood [7, 9]. It is important to note that these studies focus primarily on late pregnancy or later in life, as MSPH is typically identified only at the end of gestation, when TC levels exceed physiological values. Earlier in pregnancy, TC levels remain within the normal range (below 200 mg/dL) [11], limiting opportunities for timely monitoring. Consequently, these findings support earlier monitoring of maternal cholesterol from the first trimester (T1) to identify women who may be at increased risk of developing MSPH later in pregnancy [5, 8, 9, 11–13, 15–17, 19–32].

Artificial intelligence (AI) technologies, and particularly machine learning (ML), are increasingly used to support clinical prediction and decision-making. Significant progress has been made in the field of obstetrics and gynecology, where ML approaches have been applied to a range of perinatal complications [33–35]. To date, most studies have focused on predicting conditions such as prematurity and preeclampsia [33, 36]. Maternal hypercholesterolemia, by contrast, remains largely unexplored from a predictive-modeling perspective. Previous work has characterized MSPH cross-sectionally or examined associations between individual maternal variables and term lipid levels [5, 8, 16], without constructing and internally validating multivariable prediction models.

We therefore asked whether a single lipid determination, added to the exams already performed at the first prenatal visit, could help identify which women will later develop MSPH while their cholesterol concentrations are still within the normal range. Because these exams are inexpensive and widely available in routine prenatal care, such an approach could support earlier risk assessment and targeted lipid monitoring during pregnancy.

This study had two objectives. The primary objective was to develop and internally validate supervised ML models using T1 clinical and biochemical measurements to predict MSPH at term. The secondary objective was to characterize atherosclerosis-related biomarkers in women who later developed MSPH, compared with women with MPH.

Material and Methods

Study Design and Ethical Considerations

This prospective observational cohort study was conducted at two prenatal care centers in Santiago, Chile: Clínica Universidad de los Andes (CUA) and Hospital Parroquial San Bernardo (HPSB), which serve different populations within the city. Participants were recruited between 2022 and 2023 at their first prenatal visit, where written informed consent was obtained and blood samples were collected; clinical and biochemical variables were subsequently extracted from maternity records. The study protocol was approved by the respective Institutional Review Boards (CEC2024122 and CEC2024036) and adhered to the Declaration of Helsinki.

Study Groups

The analytical cohort comprised 160 pregnant women with singleton, uncomplicated pregnancies resulting in live births at ≥ 37 weeks' gestation. Pregnancies complicated by fetal aneuploidy, major structural anomalies, gestational diabetes, preeclampsia, fetal growth restriction, or pregnancy loss were excluded. Women with preexisting dyslipidemia or receiving lipid-lowering treatment were not eligible.

Based on TC levels at the end of pregnancy, women were assigned to the maternal physiological hypercholesterolemia (MPH) group if TC was < 290 mg/dL, and to the maternal supraphysiological hypercholesterolemia (MSPH) group if TC ≥ 290 mg/dL. This cut-off value reflects levels at which human fetoplacental endothelial and vascular dysfunction has been previously reported [5, 8, 11, 13–17]. The outcome was determined from laboratory measurement of T3 TC (see Supplementary Section S1.1), independently of T1 predictor values. The final cohort included 44 MSPH (27.5%) and 116 MPH cases (72.5%) (MSPH-positive and MSPH-negative cases, respectively). No lipid-lowering treatment or dietary intervention targeting lipid levels was indicated during pregnancy in either group.

During T1, maternal characteristics (age, weight, height, body mass index (BMI)) and biochemical parameters (fasting glycemia and lipid profile) were recorded, together with systolic (SBP) and diastolic blood pressure (DBP). TC, high-density lipoprotein (HDL) cholesterol, LDL cholesterol and triglyceride levels were determined in the complete cohort. Blood samples were obtained after an overnight fast. BMI was calculated from measured weight and height. Anthropometric and blood pressure measurements were performed according to standard clinical protocols at each center.

Cardiovascular Risk Markers

In a representative T1 case-control subgroup of the full cohort (n = 78; 46 women who later developed MPH and 32 who developed MSPH), apolipoprotein B (ApoB), apolipoprotein A1 (ApoA1), the ApoB/ApoA1 ratio, oxidized LDL (oxLDL), and malondialdehyde (MDA) were quantified. This subgroup size is comparable to that used in prior T3 MSPH studies that successfully detected meaningful differences in these cardiovascular risk and lipid peroxidation biomarkers [8, 16]. Detailed protocols are provided in Supplementary Sections S1.2–S1.4.

Comparisons between MPH and MSPH groups were performed using the Wilcoxon Mann-Whitney test. Biomarker figures were generated using GraphPad Prism.

Machine Learning Framework

The ML analysis followed a structured workflow designed to minimize information leakage and support reproducible internal validation. As summarized in Fig. 1, the framework comprised cohort harmonization, exploratory characterization of the predictor space, preprocessing, systematic benchmarking across model and validation configurations, stability-oriented model selection, final model evaluation, and post hoc interpretability. Detailed definitions of the analytical cohort, leakage-control

procedures, benchmarking strategy, final model evaluation, and interpretability analyses are provided in Supplementary Sections S2–S6.

Prediction Task and Dataset Construction

MSPH prediction was framed as a supervised binary classification task, using T1 clinical and biochemical variables as predictors and T3 MSPH status as the outcome. The predictor set comprised maternal age, weight, height, BMI, fasting glycemia, SBP, DBP, TC, HDL, LDL, and triglycerides. All predictors were obtained during T1, before ascertainment of the pregnancy outcome. T3 lipid measurements were excluded from the predictor matrix. T1 TC was retained because it represents an earlier measurement of the same analyte used to define MSPH at term and is temporally distinct from the outcome-defining T3 measurement. Detailed definitions of the prediction task, outcome, and predictor set are provided in Supplementary Sections S2.1 and S2.3.

Data from the two recruitment cohorts were harmonized through standardized variable definitions, numerical consistency checks, verification of derived variables, and assessment of predictor missingness and redundancy. The analytical cohort comprised 160 participants, including 44 MSPH cases and 116 MPH controls. The available cohort size motivated the use of comparatively parsimonious classifiers together with repeated internal-validation strategies aimed at reducing sensitivity to individual data partitions. Additional details on cohort construction and harmonization are provided in Supplementary Sections S2.2 and S2.3.

Missingness was limited to 21 predictor values distributed across 12 of 160 participants (7.5%). Six participants had one missing predictor, three had two, and three had three. Fasting glycemia was the most frequently missing variable (n = 8), followed by SBP and DBP (n = 4 each); age and all T1 lipid variables were complete.

Two complementary analytical representations were generated. The STRICT dataset retained the 148 participants with complete measurements across all 11 predictors, whereas the IMPUTED representation retained all 160 participants after median replacement of missing numerical values during dataset construction. This stored IMPUTED dataset was treated as a sensitivity representation and should be distinguished from median imputation implemented within leakage-controlled modeling pipelines. The complete missing-data structure and construction of both dataset variants are described in Supplementary Sections S2.2 and S2.4.

Leakage prevention was enforced by excluding all T3 lipid measurements from the predictor space and by fitting data-dependent preprocessing and resampling operations only within the corresponding training partitions during model validation. Participant identifiers and recruitment-source identifiers were also excluded from the predictor matrix. Further details on leakage control are provided in Supplementary Section S2.5.

Exploratory Data Analysis and Preprocessing Strategies

Exploratory analyses were performed to characterize the statistical and multivariable structure of the T1 predictor space before model benchmarking, without using these analyses for outcome-driven feature selection. Univariate comparisons between MPH and MSPH groups were assessed using distributional statistics, effect-size estimates, and multiple-testing correction. Associations among predictors were characterized using Pearson and Spearman correlations, while principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were used exclusively to visualize multivariable structure and class overlap. These analyses were descriptive and were not used to establish predictive performance or to select variables for the final model. Detailed univariate, effect-size, correlation, PCA, and t-SNE analyses are provided in Supplementary Sections S3.1–S3.4.

The same prespecified set of 11 first-trimester predictors was retained throughout model benchmarking. Alternative preprocessing configurations, including numerical scaling and PCA, were evaluated as components of the ML pipelines rather than selected from the exploratory analyses. The evaluated preprocessing strategies are described in Supplementary Sections S3.5 and S3.6.

Data Partitioning and Preprocessing

Internal validation included stratified 80/20 train-test splits, non-stratified random 80/20 splits as a sensitivity analysis, stratified 5-fold and 10-fold cross-validation, and leave-one-out cross-validation. Validation strategies involving stochastic partitioning were repeated across a fixed random-seed grid, whereas leave-one-out validation was deterministic. Detailed validation procedures are provided in Supplementary Sections S4.3–S4.6.

Feature scaling, PCA, and class-imbalance procedures were implemented within the corresponding modeling pipelines and fitted exclusively using training observations. The modeling framework also supported training-partition median imputation when missing values were present. However, the stored IMPUTED analytical representation had already undergone median replacement during dataset construction and was therefore treated as a complementary sensitivity representation, as described in Supplementary Section S2.4. Scaling alternatives and class-imbalance strategies were evaluated systematically during benchmarking rather than selected a priori. SMOTE and other resampling procedures were applied only to training observations, leaving validation and test partitions unchanged. Further details on preprocessing and resampling are provided in Supplementary Sections S3.5 and S3.6.

Distance- or similarity-aware partitioning was not required because the dataset comprised unique participants represented by low-dimensional clinical variables, with no repeated individuals or predefined similarity structure expected to generate identity-based leakage. Leakage control therefore focused on separating outcome-defining variables from predictors and preventing preprocessing or resampling information from crossing validation boundaries, as detailed in Supplementary Section S2.5.

Model Selection and Stability Criteria

Seven classifier families were evaluated under the common benchmarking framework, including logistic regression, linear discriminant analysis (LDA), quadratic discriminant analysis, k-nearest neighbors, linear support vector classification, random forest, and Gaussian naive Bayes. Model selection was designed to avoid prioritizing configurations that achieved high performance only under individual data partitions. Benchmark results were therefore summarized using Matthews correlation coefficient (MCC), and a stability-oriented generalization score was calculated as mean MCC − SD MCC. Candidate configurations were restricted to the upper 10% of the mean-MCC distribution and subsequently ranked according to higher generalization score, higher mean MCC, lower MCC variability, and lower pipeline complexity. The complete benchmarking space and selection procedure are described in Supplementary Sections S4.1–S4.8.

For the final evaluation, an LDA configuration using the IMPUTED analytical representation, standard scaling, and SMOTE was retained. This configuration was evaluated separately using repeated stratified holdout partitions, as detailed in Supplementary Section S5.

Clinical Operating Threshold

Model outputs were interpreted as discrimination scores rather than calibrated absolute risk estimates. For the final evaluation, a fixed operating threshold of 0.60 was used to summarize threshold-dependent classification performance and the associated sensitivity–specificity trade-off. This operating point yielded greater specificity than sensitivity, but it was not derived through a formal threshold-optimization procedure and should not be interpreted as an externally validated clinical cut-off. Detailed threshold-dependent performance is reported in Supplementary Section S5.

Model Interpretability and Calibration

Model interpretability was assessed using SHapley Additive exPlanations (SHAP) and permutation-based feature importance. SHAP was used to quantify the contribution of individual predictors to model outputs, while permutation importance measured the reduction in predictive performance after perturbation of each feature. These analyses were used to characterize model behavior and variable contribution without implying causal relationships. Detailed interpretability procedures and results are provided in Supplementary Section S6.

The Brier score was recorded as an auxiliary probability-based metric within the broader benchmarking framework, as described in Supplementary Section S4.7. However, no formal probability recalibration of the final model was performed. Model outputs were therefore interpreted as discrimination scores rather than calibrated estimates of absolute MSPH risk, consistent with the final evaluation described in Supplementary Section S5.

Implementation and Reproducibility

All analyses were implemented in Python 3.12 using modular and reproducible workflows for data preprocessing, exploratory analysis, model benchmarking, performance evaluation, and interpretability. Core computations were performed using NumPy 1.26.4, pandas 2.2.2, SciPy 1.13.1, scikit-learn 1.5.1, and imbalanced-learn 0.12.3. Model interpretation and visualization were conducted using SHAP 0.46.0, Matplotlib 3.9.1, and Seaborn 0.13.2.

Dataset variants, preprocessing configurations, model hyperparameters, validation schemes, and random seeds were defined through configuration files to support consistent execution across experiments. Leakage-controlled preprocessing and resampling operations used during model validation were implemented within the corresponding training pipelines. The project repository provides the configuration files, dependency specifications, benchmarking outputs, performance summaries, and figure-generation scripts required to inspect and reproduce the reported computational analyses: https://github.com/kren-ai-lab/MSPH_classification_models.

Results

Overall national mortality burden

Across 1999–2024, 604,331 U.S. deaths had I60 or I61 recorded as the underlying cause. Of these, 551 also carried a J84 multiple-cause code, corresponding to an aggregate co-certification proportion of 0.091% (95% Wilson CI, 0.084%–0.099%). ICH accounted for 437 co-certified deaths (79.3%) and SAH for 114 (20.7%); the corresponding aggregate co-certification proportions were 0.096% for ICH and 0.076% for SAH (Table 2). These unadjusted subtype proportions summarize the full 26-year period and should not be interpreted as subtype-specific risk estimates.

The combined crude population mortality rate was 0.050 per 1,000,000 (14 deaths) in 1999 and 0.100 per 1,000,000 (34 deaths) in 2024. The 2024 AAMR was 0.087 (95% CI, 0.062–0.121); the 1999 AAMR was designated unreliable and was not used for direct age-adjusted comparison. In the complete combined J84 series, BIC selected one change point at 2012: the crude population mortality rate declined during 1999–2012 (APC, −3.0%; 95% CI, −5.4% to −0.5%) and increased during 2012–2024 (APC, 6.8%; 95% CI, 4.2%–9.5%); full-period AAPC was 1.6% (95% CI, 0.5%–2.7%) (Figure 1; Table 2). These estimates describe crude population mortality involving both conditions, not the annual probability that a fatal SAH/ICH death carried J84.

Intracerebral versus subarachnoid hemorrhage

ICH accounted for the larger absolute number of co-certified deaths (437 vs 114) and had a higher unadjusted aggregate co-certification proportion (0.096% vs 0.076%). Formal annual subtype trends were not estimated because the ICH series contained a suppressed 2009 cell and SAH cells were suppressed for all years in 1999–2017 and five of seven years in 2018–2024. Because suppression is

informative and the subtype populations differ substantially in age and clinical profile, these aggregate differences do not establish greater susceptibility to co-certified ILD in ICH.

Sex and age patterns

Women accounted for 291 co-certified deaths and men for 260. Crude population rates were 0.071 per 1,000,000 person-years among women and 0.065 among men, whereas era-specific age-adjusted rates were descriptively higher in men (0.060 vs 0.043 in 1999–2017 and 0.076 vs 0.059 in 2018–2024) (Table 3). Because crude and age-adjusted comparisons differed and annual sex-specific cells were frequently suppressed, no sex effect was inferred. Adults aged ≥85 years had the highest reportable age-specific rates (0.893 per 1,000,000 in 1999–2017 and 1.403 in 2018–2024), followed by those aged 75–84 years (0.558 and 0.671, respectively). Age 45–64 results were not consistently reportable under suppression constraints.

Sensitivity analyses

Restricting the pulmonary phenotype to J84.1-coded other interstitial pulmonary diseases with fibrosis identified 410 deaths (74.4% of the broad J84 cohort), including 333 ICH and 77 SAH deaths (Figure 2; Table 4). The 2009 J84.1 annual count was suppressed. Because that missingness directly denotes a small protected count, no full-period J84.1 trend coefficient was estimated. For 2020–2024, all annual J84-plus-U07.1 counts and the aggregate five-year count were suppressed, so exact COVID-excluded rates and slopes were not estimable.

Discussion

In this study, we addressed two objectives. First, we used a ML approach to develop and internally validate a model predicting MSPH at term from T1 clinical and biochemical variables. Second, we characterized early cardiovascular risk by comparing atherogenic biomarkers between women who did and did not later develop MSPH. The predictive model showed that discriminatory information associated with subsequent MSPH is already present in T1, involving TC together with a broader set of correlated lipid, anthropometric, and hemodynamic variables. Additionally, the biomarker comparison revealed that women who later developed MSPH had higher T1 ApoB and MDA, even while their cholesterol levels remained within normal ranges, suggesting atherogenic and oxidative changes already detectable at this early stage. These findings show that clinical, lipid, and biomarker differences associated with subsequent MSPH are detectable early in pregnancy, a period that is clinically important as this condition is associated with increased cardiovascular risk for both mother and offspring later in life [7, 17, 18, 25], yet is typically identified only late in gestation, when there is limited opportunity for preventive action and processes such as fetal programming are already established.

In T1, women who later developed MSPH differed from those who remained within physiological cholesterol ranges. They were older and had higher SBP and lower BMI, together with higher total, LDL, and HDL cholesterol, although these lipid concentrations remained within the normal adult range [37]. The prevalence of MSPH at term was 27.5%, consistent with earlier reports [5, 11]. These baseline differences were accompanied by a specific structure in the early signal. Univariately, the strongest differences were carried by total and LDL cholesterol, which were closely correlated and reflect a shared biological relationship. HDL cholesterol and triglycerides showed weaker between-group differences. When all variables were examined jointly, the two groups were not clearly separated by their overall T1 profile, indicating a modest and correlated multivariable signal with substantial class overlap.

When the different model families were compared, linear classifiers showed competitive performance with comparatively stable behavior, while more flexible algorithms did not consistently improve on them. This is expected when the predictive signal is modest and the sample is limited, as added model complexity can increase instability without necessarily improving generalization, particularly given the moderate class imbalance of the cohort [38]. We therefore prioritized robustness over isolated peak performance, selecting the final model on the basis of stable and reproducible behavior across repeated evaluations. The resulting model, a linear discriminant analysis (LDA) classifier, achieved a ROC-AUC of 0.77, with a sensitivity of 0.62 and specificity of 0.79 at the predefined operating threshold.

Interpretability analyses identified T1 TC as the dominant contributor to the model's predictions. Because MPH and MSPH differ in the magnitude of this analyte, the prominence of TC may be in part a consequence of how the outcome is defined. It does not reflect information leakage, as T1 and term measurements are separated by approximately 26 weeks and T1 concentrations remained within the normal range; rather, it indicates that women who later developed MSPH already had higher, though still normal, cholesterol concentrations at the beginning of gestation. Other lipid, anthropometric, and hemodynamic variables also contributed to the fitted multivariable model. The reduced contribution of LDL, despite its clear univariate association with MSPH, is consistent with its strong correlation with TC and the resulting redundancy in predictive information. These attributions are predictive, not causal as SHAP and permutation analyses describe variables that influence model output rather than mechanistic determinants of MSPH. The moderate sensitivity noted above indicates that early clinical and biochemical variables alone do not fully capture the complexity of gestational lipid regulation. The model is therefore best understood as an internally validated prediction model rather than a test ready for clinical use. Because its predictors are inexpensive and rely on measurements obtainable at a single early visit, the approach may provide a basis for earlier risk assessment, subject to external validation before clinical implementation.

To our knowledge, this is the first ML model developed to predict maternal hypercholesterolemia during pregnancy. Previous ML work in obstetrics has concentrated on other conditions, such as preeclampsia, gestational diabetes, and preterm birth [33, 36, 39], and extending the predictive modeling to maternal hypercholesterolemia broadens the clinical reach of ML in prenatal care. Model interpretability remains particularly important in clinical prediction because similar discrimination can arise from models that

rely on very different predictor structures [40]. Here, SHAP and permutation importance allowed the contribution of individual variables to be examined directly and supported a consistent interpretation of TC as the dominant predictor while also revealing contributions from additional clinical variables [33, 40, 41].

For the secondary outcome, we compared T1 lipid concentrations and cardiovascular risk markers in a subgroup of women who did and did not develop MSPH. Both groups showed marked lipid increases between T1 and delivery, greater in women who later developed MSPH, particularly for total and LDL cholesterol, a pattern previously reported in MSPH cohorts [11]. Women who developed MSPH also had higher ApoB and MDA in T1, even though their cholesterol concentrations remained within normal ranges at the beginning of gestation. Elevated ApoB, MDA, and oxLDL have previously been reported at term in MSPH pregnancies [8, 16, 42]. The finding that MDA was elevated in T1 while oxLDL did not differ between groups is compatible with early differences in lipid peroxidation, although the temporal relationship between these markers cannot be established from the present data. The ApoB difference is also relevant, as this apolipoprotein reflects the number of atherogenic lipoprotein particles and is an established cardiovascular risk marker [43, 44]. During pregnancy, changes in lipoprotein metabolism affect ApoB-containing particles, including small, dense LDL, a subfraction that previous work suggests predominates at term in MSPH [8, 45, 46]. Our findings are compatible with early alterations in this atherogenic profile, but the mechanisms underlying these differences and their relationship to later vascular changes in MSPH require further investigation.

Several limitations should be considered. First, the cohort was of moderate size, and no independent external validation cohort was available. Although repeated validation and robustness-aware model selection reduce sensitivity to individual data partitions, they do not replace external validation. Second, subgroup-specific model performance was not evaluated because sociodemographic variables were unavailable and the number of MSPH events was insufficient for reliable stratified analyses. Third, questions of usability within current prenatal care, including how input data would be handled at the point of use and how the model would be integrated into clinical workflows, were outside the scope of this development study and would require a dedicated implementation study after external validation. Moreover, because the final model was not formally recalibrated or externally validated for absolute risk estimation, its outputs should be interpreted as discrimination scores rather than calibrated probabilities of MSPH. Finally, the biomarker findings were based on a smaller subgroup. Although its size is comparable to previous MSPH biomarker studies, these results warrant confirmation in larger samples. Addressing these limitations in larger prospective cohorts, and testing whether the early elevation of ApoB and MDA contributes to the vascular changes associated with MSPH or simply accompanies them, would clarify both how early the condition can be predicted and how it develops.

Conclusion

In this study, an LDA model built on first-trimester clinical and biochemical variables captured discriminatory information associated with the subsequent development of MSPH at term. First-

trimester total cholesterol was the dominant predictor, while additional lipid, anthropometric, and hemodynamic variables also contributed to the fitted multivariable model. Higher ApoB and MDA concentrations further showed that atherogenic and oxidative differences are already detectable during the first trimester in women who later develop MSPH.

Although this internally validated model is not yet ready for clinical use, it relies on inexpensive and easily obtained measurements and provides an interpretable framework for early MSPH risk identification. External validation in larger prospective cohorts will be required to determine its clinical utility and to clarify the biological significance of the early ApoB and MDA differences observed in this study.

Abbreviations

AAPC, average annual percentage change

AAMR, age-adjusted mortality rate

APC, annual percentage change

BIC, Bayesian information criterion

CDC, Centers for Disease Control and Prevention

CI, confidence interval

ICD-10, International Classification of Diseases, Tenth Revision

ICH, intracerebral hemorrhage

ILD, interstitial lung disease

IPF, idiopathic pulmonary fibrosis

SAH, subarachnoid hemorrhage

STROBE, Strengthening the Reporting of Observational Studies in Epidemiology

WONDER, Wide-ranging Online Data for Epidemiologic Research

Declarations

Ethics approval and consent to participate

This study analyzed publicly available, deidentified, aggregate mortality data from CDC WONDER and involved no direct participant contact or identifiable private information. Institutional review board

approval and informed consent were therefore not required for this analysis of publicly available aggregate data.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are publicly available through the CDC WONDER Multiple Cause of Death database (https://wonder.cdc.gov/mcd.html). Derived analytic data and statistical code are available from the corresponding author on reasonable request, subject to CDC WONDER data-use restrictions and preservation of suppression labels.

Competing interests

The authors declare that they have no competing interests.

Funding

No specific funding was received for this study.

Authors' contributions

CC conceived the study, developed the methodology, performed the investigation, curated and analyzed the data, prepared the visualizations, and drafted and revised the manuscript. TH provided clinical supervision, contributed to interpretation of the pulmonary findings, and critically revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Martin SS, Aday AW, Allen NB, Almarzooq ZI, Anderson CAM, Arora P, et al. 2025 Heart Disease and Stroke Statistics: A Report of US and Global Data from the American Heart Association. Circulation. 2025. https://doi.org/10.1161/CIR.0000000000001303.

2. Brizzi P, Tonolo G, Esposito F, Puddu L, Dessole S, Maioli M, et al. Lipoprotein metabolism during normal pregnancy. Am J Obstet Gynecol. 1999;181:430–4. https://doi.org/10.1016/S0002-9378(99)70574-0.
3. Chatuphonprasert W, Jarukamjorn K, Ellinger I. Physiology and pathophysiology of steroid biosynthesis, transport and metabolism in the human placenta. Front Pharmacol. 2018;9:1–29. https://doi.org/10.3389/fphar.2018.01027.
4. Herrera E. Lipid metabolism in pregnancy and its consequences in the fetus and newborn. Endocrine. 2002;19:43–55. https://doi.org/10.1385/ENDO:19:1:43.
5. Leiva A, Salsoso R, Sáez T, Sanhueza C, Pardo F, Sobrevia L, Placenta. W.B. Saunders Ltd; 2015;36:895–902. https://doi.org/10.1016/j.placenta.2015.05.012.
6. Mendelson MM, Lyass A, O'Donnell CJ, D'Agostino RB, Levy D. Association of Maternal Prepregnancy Dyslipidemia With Adult Offspring Dyslipidemia in Excess of Anthropometric, Lifestyle, and Genetic Factors in the Framingham Heart Study. JAMA Cardiol. 2016;1:26. https://doi.org/10.1001/jamacardio.2015.0304.
7. Cacciatore F, Bruzzese G, Abete P, Russo G, Palinski W, Napoli C. Maternal hypercholesterolaemia during pregnancy affects severity of myocardial infarction in young adults. Eur J Prev Cardiol. 2022;29:758–65. https://doi.org/10.1093/eurjpc/zwab152.
8. Cantin C, Garchitorena MJ, Escalona R, Carvajal JA, Illanes SE, Gutierrez J, et al. Increased Circulating Levels of PCSK9 and Pro-Atherogenic Lipoprotein Profile in Pregnant Women with Maternal Supraphysiological Hypercholesterolemia. Antioxidants. 2022;11. https://doi.org/10.3390/antiox11050869.
9. Kolovou G. Can pregnancy hypercholesterolaemia cause myocardial infarction in young adults? Eur J Prev Cardiol. 2022;29:756–7. https://doi.org/10.1093/eurjpc/zwab162.
10. van Lennep JER, Nerenberg KA. Delivering evidence to prevent recurrent venous thromboembolism in pregnancy. Lancet. 2022;400:1743–5. https://doi.org/10.1016/S0140-6736(22)02030-X.
11. Leiva A, De Medina CD, Salsoso R, Sáez T, San Martín S, Abarzúa F, Arteriosclerosis, et al. Thromb Vascular Biology. 2013;33:2444–53. https://doi.org/10.1161/ATVBAHA.113.301987.
12. Leiva A, Fuenzalida B, Westermeier F, Toledo F, Salomón C, Gutiérrez J et al. Role for Tetrahydrobiopterin in the Fetoplacental Endothelial Dysfunction in Maternal Supraphysiological Hypercholesterolemia. Steinert J, editor. Oxidative Medicine and Cellular Longevity. Hindawi Limited; 2016;2016. https://doi.org/10.1155/2016/5346327
13. Fuenzalida B, Sobrevia B, Cantin C, Carvajal L, Salsoso R, Gutiérrez J, et al. Maternal supraphysiological hypercholesterolemia associates with endothelial dysfunction of the placental microvasculature. Sci Rep Nat Publishing Group. 2018;8. https://doi.org/10.1038/s41598-018-25985-6.
14. Fuenzalida B, Cantin C, Kallol S, Carvajal L, Pastén V, Contreras-Duarte S, et al. Cholesterol uptake and efflux are impaired in human trophoblast cells from pregnancies with maternal

supraphysiological hypercholesterolemia. Sci Rep. 2020;10:1–16. https://doi.org/10.1038/s41598-020-61629-4.

15. Liguori A, D'Armiento FP, Palagiano A, Balestrieri ML, Williams-Ignarro S, De Nigris F, et al. Effect of gestational hypercholesterolaemia on omental vasoreactivity, placental enzyme activity and transplacental passage of normal and oxidised fatty acids. BJOG: Int J Obstet Gynecol. 2007;114:1547–56. https://doi.org/10.1111/j.1471-0528.2007.01510.x.
16. Cantin C, Morales A, Serra R, Illanes SE, Leiva A. Maternal Supraphysiological Hypercholesterolemia Is Accompanied by Shifts in the Composition and Anti-Atherogenic Functions of Maternal HDL along with Maternal Cardiovascular Risk Markers at Term of Pregnancy. Antioxidants. 2023;12. https://doi.org/10.3390/antiox12101804.
17. Napoli C, D'Armiento FP, Mancini FP, Postiglione A, Witztum JL, Palumbo G, et al. Fatty streak formation occurs in human fetal aortas and is greatly enhanced by maternal hypercholesterolemia. Intimal accumulation of low density lipoprotein and its oxidation precede monocyte recruitment into early atherosclerotic lesions. J Clin Invest. 1997;100:2680–90. https://doi.org/10.1172/JCI119813.
18. Napoli C, Glass CK, Witztum JL, Deutsch R, D'Armiento FP, Palinski W. Influence of maternal hypercholesterolaemia during pregnancy on progression of early atherosclerotic lesions in childhood: Fate of Early Lesions in Children (FELIC) study. The Lancet. Elsevier Ltd. 1999;354:1234–41. https://doi.org/10.1016/S0140-6736(99)02131-5.
19. Leiva A, Contreras-Duarte S, Amigo L, Sepúlveda E, Boric M, Quiñones V, et al. Gugulipid causes hypercholesterolemia leading to endothelial dysfunction, increased atherosclerosis, and premature death by ischemic heart disease in male mice. PLoS ONE. 2017;12:1–19. https://doi.org/10.1371/journal.pone.0184280.
20. Duarte SC, Chen P, Andía M, Uribe S, Irarrázaval P, Kopp S, et al. Attenuation of atherogenic apo B – 48 – dependent hyperlipidemia and high density lipoprotein remodeling induced by vitamin C and E combination and their beneficial effect on lethal ischemic heart disease in mice. Biol Res BioMed Cent. 2018;51:1–11. https://doi.org/10.1186/s40659-018-0183-6.
21. Contreras-Duarte S, Carvajal L, Garchitorena MJ, Subiabre M, Fuenzalida B, Cantin C, et al. Gestational Diabetes Mellitus Treatment Schemes Modify Maternal Plasma Cholesterol Levels Dependent to Women´s Weight: Possible Impact on Feto-Placental Vascular Function. Nutrients. 2020;12:506. https://doi.org/10.3390/nu12020506.
22. Contreras-Duarte S, Cantin C, Farias M, Leiva A. High total cholesterol and triglycerides levels increase arginases metabolism, impairing nitric oxide signaling and worsening fetoplacental endothelial dysfunction in gestational diabetes mellitus pregnancies. Biochimica et Biophysica Acta - Molecular Basis of Disease. Elsevier B V. 2021;1867:166216. https://doi.org/10.1016/j.bbadis.2021.166216.
23. Contreras-Duarte S, Escalona-Rivano R, Cantin C, Valdivia P, Zapata D, Carvajal L, et al. Small extracellular vesicles from pregnant women with maternal supraphysiological hypercholesterolemia

impair endothelial cell function in vitro. Vascul Pharmacol. 2023;150. https://doi.org/10.1016/j.vph.2023.107174.

24. Cantin C, Fuenzalida B, Leiva A. Maternal hypercholesterolemia during pregnancy: Potential modulation of cholesterol transport through the human placenta and lipoprotein profile in maternal and neonatal circulation. Placenta. 2020;94:26–33. https://doi.org/10.1016/j.placenta.2020.03.007.
25. de Nigris F, Cacciatore F, Mancini FP, Vitale DF, Mansueto G, D'Armiento FP, et al. Epigenetic Hallmarks of Fetal Early Atherosclerotic Lesions in Humans. JAMA Cardiol JAMA Cardiol. 2018;3:1184. https://doi.org/10.1001/jamacardio.2018.3546.
26. Cai X, Liang N, Cai X, Zhou Q, Dang Q, Hu Z, et al. Lipid Metabolic Genes and Maternal Supraphysiological Hypercholesterolemia: An Analysis of Maternal-fetal Interaction. J Clin Endocrinol Metab. 2022;107:E3134–44. https://doi.org/10.1210/clinem/dgac317.
27. De Nigris F, Franconi F, Maida I, Palumbo G, Anania V, Napoli C. Modulation by α- and γ-tocopherol and oxidized low-density lipoprotein of apoptotic signaling in human coronary smooth muscle cells. Biochem Pharmacol. 2000;59:1477–87. https://doi.org/10.1016/S0006-2952(00)00275-6.
28. Palinski W, Napoli C. The fetal origins of atherosclerosis: maternal hypercholesterolemia, and cholesterol-lowering or antioxidant treatment during pregnancy influence in utero. FASEB J. 2002.
29. Palinski W, Nicolaides E, Liguori A, Napoli C. Influence of maternal dysmetabolic conditions during pregnancy on cardiovascular disease. J Cardiovasc Transl Res. 2009;2:277–85. https://doi.org/10.1007/s12265-009-9108-7.
30. Liguori A, D'Armiento FP, Palagiano A, Palinski W, Napoli C. Maternal C-reactive protein and developmental programming of atherosclerosis. Am J Obstet Gynecol. 2008;198. https://doi.org/10.1016/j.ajog.2007.11.027. :281.e1-281.e5.
31. Jayalekshmi VS, Ramachandran S. Maternal cholesterol levels during gestation: boon or bane for the offspring? Molecular and Cellular Biochemistry. Springer US. 2021;476:401–16. https://doi.org/10.1007/s11010-020-03916-2.
32. Van Roeters JE, Tokgözoğlu LS, Badimon L, Dumanski SM, Gulati M, Hess CN, et al. Women, lipids, and atherosclerotic cardiovascular disease: a call to action from the European Atherosclerosis Society. European Heart Journal. Volume 44. Oxford University Press; 2023. pp. 4157–73. https://doi.org/10.1093/eurheartj/ehad472.
33. Bertini A, Salas R, Chabert S, Sobrevia L, Pardo F. Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review. Front Bioeng Biotechnol. 2022;9:1–16. https://doi.org/10.3389/fbioe.2021.780389.
34. Pi X, Wang J, Chu L, Zhang G, Zhang W. Prediction of high-risk pregnancy based on machine learning algorithms. Sci Rep. 2025;15:1–8. https://doi.org/10.1038/s41598-025-00450-3.
35. Sufriyana H, Husnayain A, Chen YL, Kuo CY, Singh O, Yeh TY, et al. Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: Systematic review and meta-analysis. JMIR Med Inf. 2020;8:1–26. https://doi.org/10.2196/16503.

36. Islam MN, Mustafina SN, Mahmud T, Khan NI. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. BMC Pregnancy Childbirth BioMed Cent. 2022;22:1–19. https://doi.org/10.1186/s12884-022-04594-2.
37. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, NLA/PCNA Guideline on the Management of Blood Cholesterol. Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol Elsevier. 2019;73:3168–209. https://doi.org/10.1016/j.jacc.2018.11.002. /ABC/ACPM/ADA/AGS/APhA/ASPC/. 2018 AHA/ACC/AACVPR/AAPA.
38. Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. Biom J. 2023;65:2200302. https://doi.org/10.1002/bimj.202200302.
39. Mennickent D, Rodríguez A, Opazo MC, Riedel CA, Castro E, Eriz-Salinas A, et al. Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications. Front Endocrinol. 2023;14:1–22. https://doi.org/10.3389/fendo.2023.1130139.
40. Carvalho DV, Pereira EM, Cardoso JS. Machine learning interpretability: A survey on methods and metrics. Electron (Switzerland). 2019;8:1–34. https://doi.org/10.3390/electronics8080832.
41. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clin Transl Sci. 2024;17:1–15. https://doi.org/10.1111/cts.70056.
42. Marseille-Tremblay C, Ethier-Chiasson M, Forest JC, Giguère Y, Masse A, Mounier C, et al. Impact of maternal circulating cholesterol and gestational diabetes mellitus on lipid metabolism in human term placenta. Mol Reprod Dev Mol Reprod Dev. 2008;75:1054–62. https://doi.org/10.1002/mrd.20842.
43. Sniderman AD, Dufresne L, Pencina KM, Bilgic S, Thanassoulis G, Pencina MJ. Discordance among apoB, non–high-density lipoprotein cholesterol, and triglycerides: implications for cardiovascular prevention. Eur Heart J Oxf Univ Press. 2024;45:2410–8. https://doi.org/10.1093/eurheartj/ehae258.
44. Sehayek D, Cole J, Björnson E, Wilkins JT, Mortensen MB, Dufresne L, et al. ApoB, LDL-C and non-HDL-C as Markers of Cardiovascular Risk. J Clin Lipidol. 2025. https://doi.org/10.1016/j.jacl.2025.05.024.
45. Brunzell JD, Arteriosclerosis. Thromb Vascular Biology. 2005;25:474–5. https://doi.org/10.1161/01.ATV.0000156537.78366.1d.
46. Thongtang N, Diffenderfer MR, Ooi EMM, Barrett PHR, Turner SM, Le NA, et al. Metabolism and proteomics of large and small dense LDL in combined hyperlipidemia: Effects of rosuvastatin. J Lipid Res. 2017;58:1315–24. https://doi.org/10.1194/jlr.M073882.

Tables

Table 1. Study definitions and estimands

Component	Primary definition / estimand	Secondary / sensitivity definition	Interpretive role
Data period	1999–2017 D77 + 2018–2024 D157	—	Uses nonoverlapping WONDER eras and avoids duplicate 2018–2020 records
Hemorrhage phenotype	Underlying cause I60 (SAH) or I61 (ICH)	Separate ICH and SAH descriptions	Restricts cohort to deaths attributed to nontraumatic intracranial hemorrhage
Pulmonary phenotype	J84 listed as a multiple cause	J84.1, other interstitial pulmonary diseases with fibrosis	Administrative co-certification definitions, not adjudicated clinical ILD
Primary estimand	Aggregate proportion of I60/I61 underlying-cause deaths carrying J84, with 95% Wilson CI	Distribution of co-certified deaths by hemorrhage subtype	Measures co-certification among fatal hemorrhagic-stroke deaths
Secondary temporal measure	Annual crude population mortality; reliable AAMRs when available	Segmented Poisson model for complete combined J84 series only	Measures population burden, not annual co-certification proportion
Pandemic sensitivity	All eligible deaths	J84 + U07.1 co-certification in 2020–2024	Exact exclusion was not estimable because cells were suppressed
Suppressed annual cells	Not reconstructed	No formal trend coefficient for affected series	Avoids informative-missingness bias and preserves CDC disclosure rules

Table 2. Co-certification burden and secondary population mortality, 1999–2024

Group	Deaths, n	Aggregate co-certification proportion	1999 crude rate*	2024 crude rate*	AAPC, % (95% CI)
SAH + J84	114	0.076%	Suppressed	0.029	Not estimable
ICH + J84	437	0.096%	0.043	0.071	Not estimated‡
SAH/ICH + J84	551	0.091% (95% CI, 0.084%–0.099%)	0.050	0.100	1.6 (0.5 to 2.7)

*Rates are per 1,000,000 U.S. population. The aggregate co-certification proportion for the combined SAH/ICH + J84 cohort was 0.091% (95% Wilson CI, 0.084%–0.099%). AAPC for the combined series is a secondary exploratory population-rate measure and does not represent an annual J84 co-certification proportion among fatal SAH/ICH deaths. Model-based intervals are conditional on the selected Poisson

model. ‡The ICH annual series contained a suppressed 2009 cell; no trend coefficient was reported because suppression is informative.

Table 3. Descriptive sex and age patterns

Stratum	Co-certified deaths, n	Crude / age-specific rate	Era-specific AAMR	Interpretation
Female	291	0.071†	0.043 / 0.059	Crude and age-adjusted ordering differed; no sex effect inferred
Male	260	0.065†	0.060 / 0.076	Crude and age-adjusted ordering differed; no sex effect inferred
Age 65–74	—	0.168 / 0.185	Not applicable	Age-specific rate by 1999–2017 / 2018–2024 era
Age 75–84	—	0.558 / 0.671	Not applicable	Age-specific rate by 1999–2017 / 2018–2024 era
Age ≥85	—	0.893 / 1.403	Not applicable	Highest observed age-specific rates
Age 45–64	—	Not consistently reportable	Not applicable	Suppression limited consistent era-specific reporting

†Pooled sex-specific crude population rates; age-specific rates are shown separately for 1999–2017 and 2018–2024 because denominators were extracted by database era. Rates are per 1,000,000.

Table 4. Coding-definition and suppression-limited sensitivity analyses

Analysis	Group	Deaths, n	Trend estimate	Interpretation
J84.1 restriction	Combined	410	Not estimated*	2009 annual cell suppressed; aggregate burden reported
J84.1 restriction	ICH	333	Not estimated	Suppressed annual cell(s); descriptive count only
J84.1 restriction	SAH	77	Not estimated	Extensive suppression
COVID-19 exclusion	Combined	Suppressed	Not estimable	All J84 + U07.1 cells were suppressed

*No full-period J84.1 trend coefficient was estimated because the 2009 annual count was suppressed. CDC suppression identifies a small count and is therefore informative missingness; the protected value was neither reconstructed nor imputed.

Figures

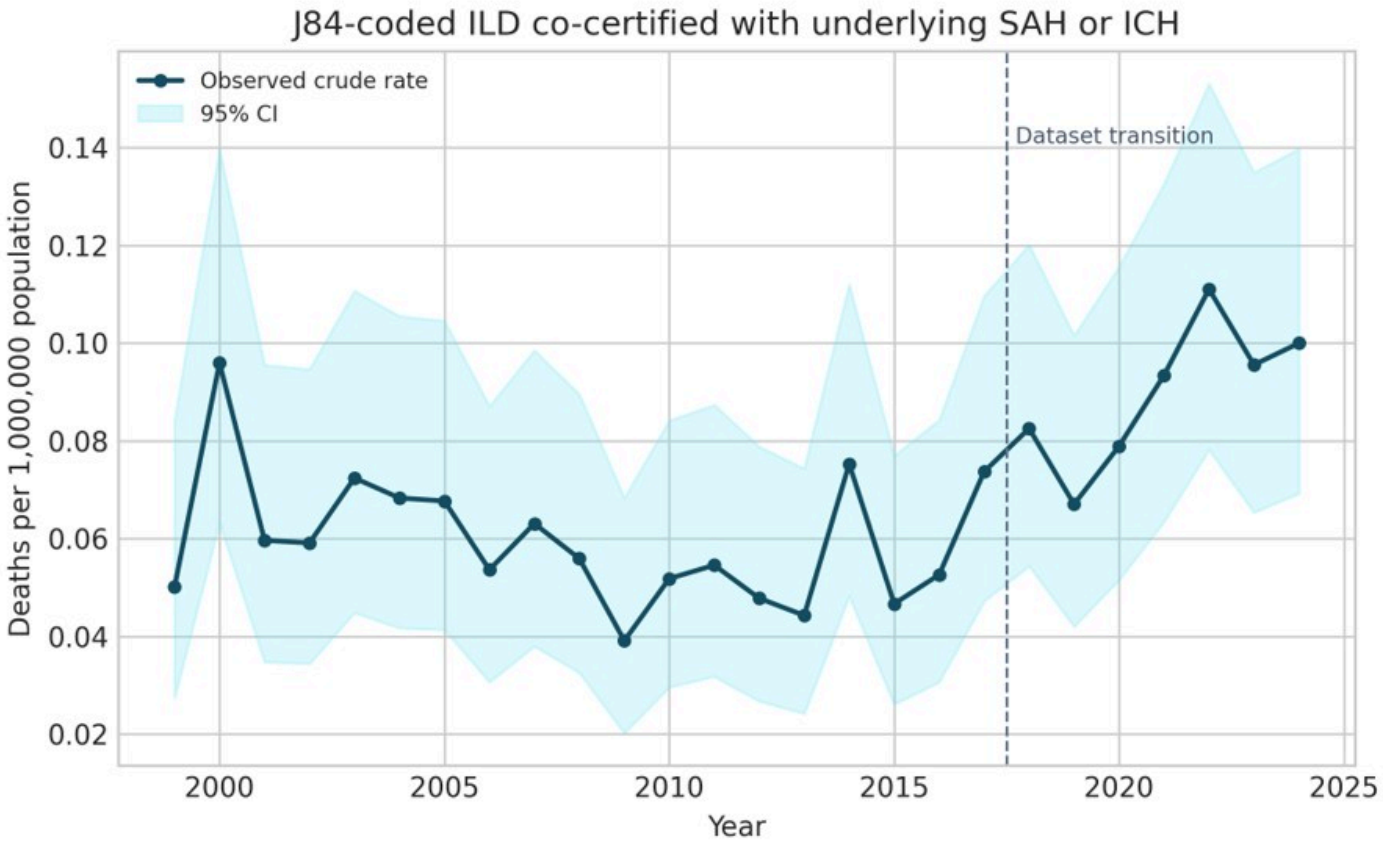


Figure 1

National population mortality involving hemorrhagic stroke and J84-coded ILD, 1999–2024. Points show annual crude population mortality rates for deaths with I60 or I61 as the underlying cause and J84 as a multiple cause; shading shows exact Poisson 95% confidence intervals. The fitted segmented trend is secondary and exploratory. It uses the U.S. population, not annual fatal SAH/ICH deaths, as the denominator and therefore does not measure the annual J84 co-certification proportion among fatal hemorrhagic-stroke deaths. The dashed line marks the transition between the nonoverlapping D77 and D157 data series.

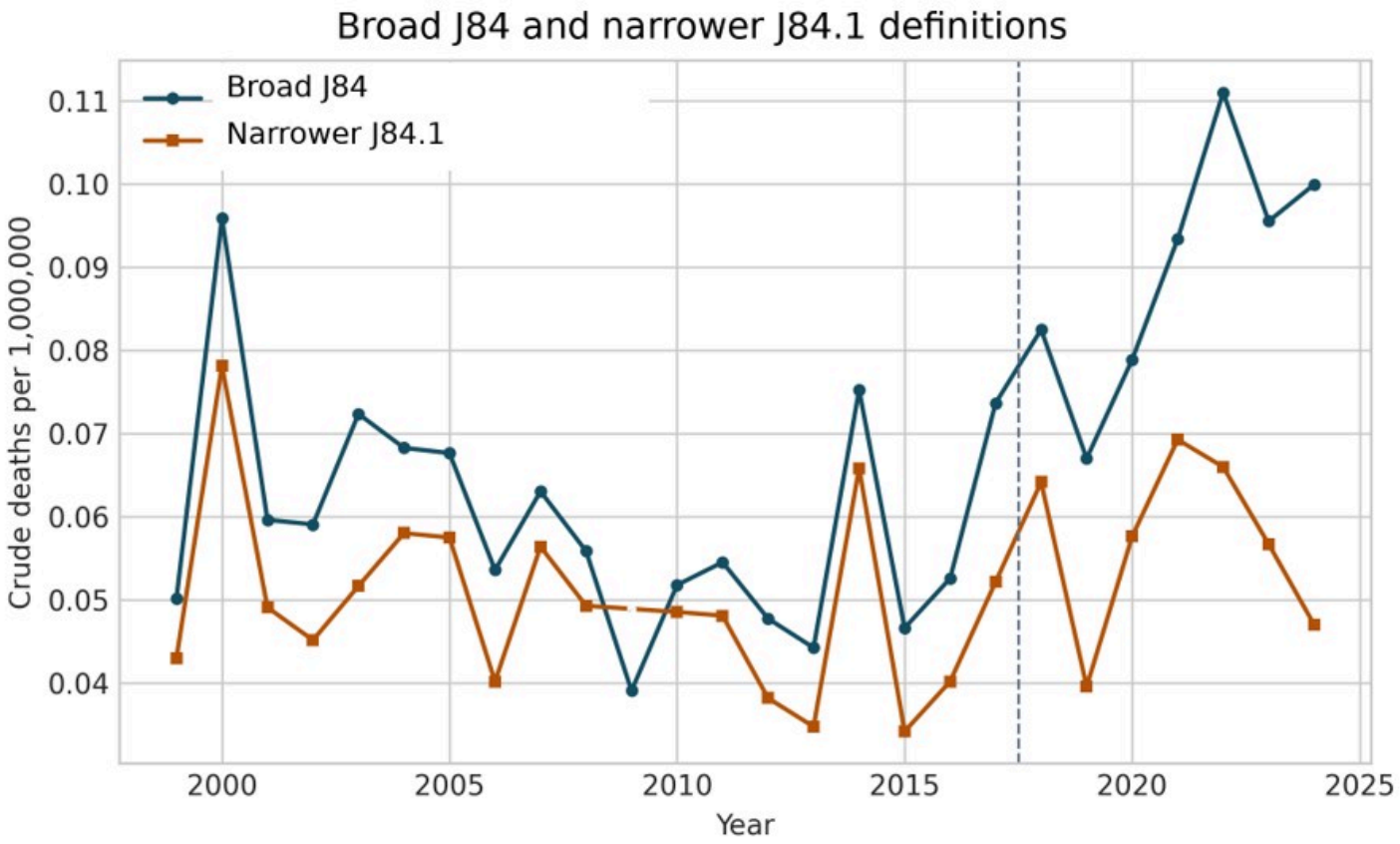


Figure 2

Broad J84 and J84.1 coding-definition sensitivity analysis. Annual crude population mortality rates are shown descriptively for the broad J84 definition and J84.1-coded other interstitial pulmonary diseases with fibrosis. The missing 2009 J84.1 value reflects CDC suppression and was neither reconstructed nor imputed. J84.1 is a narrower administrative coding definition, not a clinically adjudicated fibrotic-ILD phenotype. The dashed line marks the transition between the nonoverlapping D77 and D157 data series.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SerraCantinSupplmaterial.docx