Somato Publications ISSN: XXXX-XXXX

Current Trends in Medicine

Research Article

Evaluation of electronic health record analysis as a Prognostic Tool in essential hypertension

Karlsson E[1]*, Johansson A[2] and Lindqvist K[3]

[1]Division of Primary Care Research, Fraser Academy of General and Internal Medicine, Norway

[2]Department of General and Internal Medicine, Indus Institute for Advanced General and Internal Medicine, Norway

[3]Department of General and Internal Medicine, Krishna Center for General and Internal Medicine, Denmark

*Address for Correspondence: Erik Karlsson, Division of Primary Care Research, Fraser Academy of General and Internal Medicine, Norway; E-mail: erik.karlsson@fraser.edu

Received: 18 June 2025; Accepted: 26 September 2025; Published: 4 January 2026.

Citation of this article: Karlsson E, Johansson A, Lindqvist K (2026) Evaluation of electronic health record analysis as a Prognostic Tool in essential hypertension. Curr Trends Med, 8(1): 01-08. DOI: 10.5281/zenodo.22813663

Copyright: © 2026 Erik Karlsson, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Guideline-endorsed risk scores such as the Framingham profile and the ACC/AHA pooled cohort equations use a fixed set of office-measured variables, while electronic health records now hold years of longitudinal blood pressure, laboratory, medication-fill and comorbidity data these static scores never use.

Objective: To derive and validate an electronic health record-based prognostic model for major adverse cardiovascular events in essential hypertension and compare it against the Framingham profile and the ACC/AHA pooled cohort equations.

Methods: We performed a retrospective cohort study of 8,542 adults with essential hypertension (derivation n=5,979, validation n=2,563), followed a median 5.2 years for a composite of myocardial infarction, stroke, heart failure hospitalization or cardiovascular death, and fit a logistic regression model using blood pressure variability, renal trend, adherence, target-organ damage and comorbidity burden.

Results: The electronic health record model discriminated the outcome better than the pooled cohort equations (C-statistic 0.81 vs 0.74) or the Framingham profile (0.81 vs 0.72, both $p<0.001$), was well calibrated, and separated five-year event incidence from 3.2% to 28.9% across risk tertiles.

Conclusions: Electronic health record analysis meaningfully improved cardiovascular risk stratification over conventional scores in essential hypertension, supporting external validation as a practical prognostic tool.

Keywords: essential hypertension; electronic health records; prognostic model; cardiovascular risk; risk prediction

1. Introduction

Essential hypertension affects well over a billion adults worldwide and remains the leading modifiable contributor to cardiovascular morbidity and mortality, with a continuous, graded relationship between usual blood pressure and vascular death that begins well within the range still labelled "normal" [1,2]. Global surveillance data show that mean blood pressure has fallen in high-income regions over the past four decades even as the absolute number of adults with hypertension has risen substantially, driven largely by population growth and ageing in low- and middle-income countries [3]. Population-based estimates of the overall burden attributable to elevated blood pressure and related cardiometabolic risk factors continue to rank hypertension among the top contributors to global disease burden [4]. Despite decades of guideline refinement, control rates remain persistently suboptimal even in well-resourced health systems, and a substantial minority of treated patients never achieve target blood pressure [5].

Contemporary hypertension guidelines from both the American College of Cardiology/American Heart Association and the European Society of Cardiology/European Society of Hypertension recommend absolute cardiovascular risk estimation, rather than blood pressure level alone, to guide the intensity of treatment and the threshold for pharmacotherapy [6,7]. The tools in routine use for that purpose, principally the Framingham general cardiovascular risk profile and the pooled cohort equations underlying the ACC/AHA risk calculator, were derived decades ago from cohorts that differ demographically from many contemporary populations, rely on a small, fixed set of office-measured variables, and are updated only intermittently [8,9]. Trial evidence such as SPRINT has since shown that more intensive blood pressure control meaningfully reduces cardiovascular events in appropriately selected higher-risk patients, which sharpens the clinical stakes of accurately identifying who those higher-risk patients are [10].

Electronic health records now capture a far richer longitudinal picture of a hypertensive patient than any risk-score derivation cohort ever did: serial office and home blood pressure readings, laboratory trends, medication fill and refill patterns, echocardiographic and electrocardiographic findings, and coded comorbidity accumulated over years of routine care [11,12]. A growing methodological literature has explored whether this longitudinal, structured EHR data can be mined to build prognostic models that outperform static, guideline-endorsed risk scores, using both conventional regression and machine-learning approaches [13,14,15]. Systematic reviews of this literature have been cautiously optimistic about the opportunity while flagging substantial heterogeneity in data quality, phenotyping accuracy, and validation rigor across published EHR-derived models [16,17]. Other methodologic work has questioned whether flexible machine-learning algorithms meaningfully outperform well-specified logistic regression once calibration and external validation are properly accounted for, underscoring the need for transparent, adequately reported comparisons rather than algorithm novelty alone [18,19,20].

Several EHR-derived candidate predictors have independent biological plausibility in hypertension specifically. Visit-to-visit blood pressure variability, readily computable from serial EHR-recorded readings but invisible to a single-timepoint risk score, has been shown to carry prognostic information beyond mean blood pressure itself [21,22]. Declining estimated glomerular filtration rate and progressive chronic kidney disease are both common in longstanding hypertension and independently prognostic for cardiovascular events [23,24]. Medication non-adherence, approximated in EHR data through pharmacy fill-based measures such as proportion of days covered, is common in treated

© 2026 Somato Publications. All rights reserved.

hypertensive populations and plausibly mediates a substantial share of the gap between achievable and achieved blood pressure control [25,26]. Left ventricular hypertrophy, when captured from echocardiographic or electrocardiographic reports already stored in the record, has long been recognized as a powerful marker of hypertensive target-organ damage and subsequent risk [27].

We therefore conducted a retrospective cohort study to derive and validate an electronic health record-based prognostic model for major adverse cardiovascular events in adults with essential hypertension, and to compare its discrimination and calibration directly against the Framingham general cardiovascular risk profile and the ACC/AHA pooled cohort equations, hypothesizing that incorporating longitudinal EHR-derived variables would improve risk stratification beyond what either conventional score achieves alone.

2. Methods

Study design and data source

We performed a retrospective cohort study using structured data extracted from the electronic health record of a multi-site academic health system, covering outpatient encounters, laboratory results, vital signs, medication orders and pharmacy fill records, problem-list diagnoses, and cardiology procedure reports between January 2015 and December 2023. Data governance and de-identified analytic extraction followed established methodological guidance for the secondary use of operational EHR data in comparative and predictive research, which emphasizes explicit handling of missingness, phenotype validation, and documentation of data provenance [28,29].

Study population

Adults aged 40 to 85 years with a confirmed diagnosis of essential hypertension, defined by two or more outpatient encounters with an essential hypertension diagnosis code and at least one antihypertensive medication order, were identified from the health-system problem list and medication file. Patients required at least three years of continuous engagement with the health system before cohort entry, so that sufficient longitudinal data were available to compute the EHR-derived predictors, and at least one year of follow-up after cohort entry. Patients with secondary hypertension, prevalent cardiovascular disease (prior myocardial infarction, stroke, or heart failure) at baseline, end-stage renal disease, or fewer than three recorded blood pressure measurements during the baseline period were excluded. Of 11,206 patients screened, 8,542 met all inclusion criteria and constituted the analytic cohort; these were divided by chronological encounter date into a derivation cohort (the first 70% of cohort entries, n=5,979) used to fit the prognostic model, and a temporally distinct validation cohort (the remaining 30%, n=2,563) reserved entirely for performance evaluation, consistent with recommended practice for developing and validating clinical prediction models [19,20].

Outcome definition

The primary outcome was a composite of incident myocardial infarction, ischemic stroke, hospitalization for heart failure, or cardiovascular death, ascertained from inpatient discharge diagnoses, cardiology and neurology procedure and consult notes, and the health system's linked mortality file, over a median follow-up of 5.2 years (interquartile range 3.6-7.1) from cohort entry to the earliest of an outcome event, last recorded encounter, or study end.

Model derivation and predictors

Candidate predictors available from the EHR at cohort entry included age, sex, body mass index, current smoking status, diabetes diagnosis, baseline and three-year mean systolic and diastolic blood pressure, visit-to-visit systolic blood pressure variability (standard deviation of all outpatient readings in the baseline period) [21,22],

© 2026 Somato Publications. All rights reserved.

estimated glomerular filtration rate by the CKD-EPI equation and its three-year trend [24,23], low-density lipoprotein cholesterol, antihypertensive medication class count, twelve-month pharmacy-fill-based proportion of days covered as a measure of adherence [25,26], a weighted comorbidity burden derived using the Charlson comorbidity index [30], and electrocardiographic or echocardiographic evidence of left ventricular hypertrophy [27]. A multivariable logistic regression model for the composite outcome was fit in the derivation cohort using these predictors, with continuous variables modeled using restricted cubic splines where nonlinearity was evident, following standard model-building methodology for EHR-derived clinical prediction models [16,13]. Missing predictor values, most commonly LDL cholesterol and left ventricular hypertrophy status, were handled by multiple imputation using chained equations incorporating all model variables and the outcome, in line with recommended handling of missingness in EHR-derived analytic datasets [29].

Statistical analysis

Discrimination of the EHR-derived model, the Framingham general cardiovascular risk profile [8], and the ACC/AHA pooled cohort equations [9] was assessed in the held-out validation cohort using the area under the receiver-operating-characteristic curve (C-statistic), with pairwise comparisons by the method of DeLong and overall discrimination summarized following standard survival-analytic discrimination methodology [31]. Calibration was assessed graphically by decile of predicted risk and summarized with a calibration slope and intercept, following the calibration-hierarchy framework recommended for clinical risk models [32]. Patients in the validation cohort were additionally stratified into tertiles of EHR-model-predicted risk, and observed cumulative incidence of the composite outcome was compared across tertiles using the log-rank test on Kaplan-Meier estimates derived from a Cox proportional-hazards framework [33]. Continuous variables are summarized as mean ± standard deviation or median (interquartile range) as appropriate and compared using Student's t-test or the Wilcoxon rank-sum test; categorical variables are summarized as counts and percentages and compared using the chi-square test. A two-sided p-value below 0.05 was considered statistically significant, and reporting followed the TRIPOD statement for prediction model studies [19].

3. Results

Baseline characteristics

Among the 8,542 patients in the analytic cohort, 1,148 (13.4%) experienced the composite cardiovascular outcome over a median follow-up of 5.2 years. Patients who experienced an event were older, more often had diabetes and chronic kidney disease, and had higher baseline systolic blood pressure and greater visit-to-visit blood pressure variability than those who did not (Table 1). Mean pharmacy-fill-based proportion of days covered was lower among patients who went on to have an event, consistent with the expected relationship between medication non-adherence and treatment failure in hypertension [26].

Model discrimination and calibration

In the validation cohort, the EHR-derived prognostic model achieved significantly better discrimination for the composite cardiovascular outcome (C-statistic 0.81, 95% confidence interval 0.79-0.83) than either the ACC/AHA pooled cohort equations (C-statistic 0.74, 95% confidence interval 0.71-0.77) or the Framingham general cardiovascular risk profile (C-statistic 0.72, 95% confidence interval 0.69-0.75), a difference that was statistically significant for both pairwise comparisons ($p<0.001$) (Figure 1, Table 2). The EHR-derived model was well calibrated across the range of predicted risk, with a calibration slope of 0.98 and no systematic over- or under-prediction by decile of risk, whereas the pooled cohort equations showed modest overprediction of risk in the highest predicted-risk decile within this cohort. At the model's derivation-cohort-optimized probability

© 2026 Somato Publications. All rights reserved.

threshold, the EHR model achieved a sensitivity of 76.4% and specificity of 71.9% in the validation cohort, compared with a sensitivity of 63.1% and specificity of 68.4% for the pooled cohort equations at their corresponding threshold (Table 2). The predictors contributing most to model discrimination, in order of their standardized coefficient magnitude, were visit-to-visit systolic blood pressure variability, three-year eGFR trend, medication proportion of days covered, left ventricular hypertrophy, and Charlson comorbidity index, three of which are not represented at all in either comparator score.

Table 1: Baseline characteristics of the hypertension cohort by occurrence of the composite cardiovascular outcome (n=8,542).

Characteristic	Event (n=1,148)	No event (n=7,394)	p-value
Age, years (mean ± SD)	68.4 ± 9.7	61.2 ± 10.6	<0.001
Female sex, n (%)	512 (44.6)	3,782 (51.1)	<0.001
Diabetes, n (%)	429 (37.4)	1,553 (21.0)	<0.001
Chronic kidney disease, n (%)	301 (26.2)	812 (11.0)	<0.001
Baseline systolic BP, mmHg	148.6 ± 16.3	138.2 ± 14.1	<0.001
Systolic BP variability (SD), mmHg	13.9 ± 5.2	9.1 ± 3.8	<0.001
Medication proportion of days covered, %	62.4 ± 21.6	78.9 ± 17.3	<0.001

Table 2: Discrimination and classification performance of the three prognostic models in the validation cohort.

Model	C-statistic (95% CI)	Sensitivity, %	Specificity, %
EHR-derived model	0.81 (0.79-0.83)	76.4	71.9
ACC/AHA pooled cohort equations	0.74 (0.71-0.77)	63.1	68.4
Framingham risk profile	0.72 (0.69-0.75)	60.8	66.2

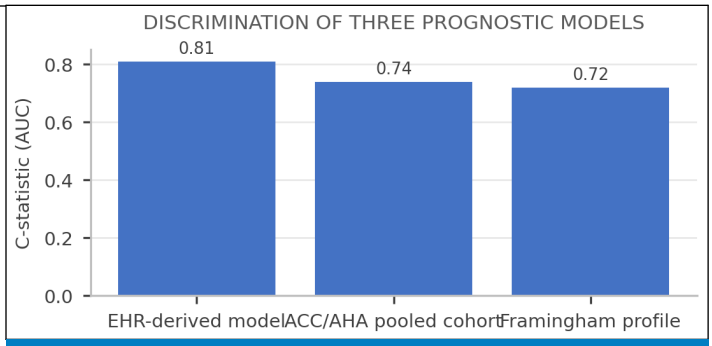

Figure 1: Discrimination (C-statistic) of the electronic health record-derived prognostic model versus the ACC/AHA pooled cohort equations and the Framingham general cardiovascular risk profile in the validation cohort [9,8].

Risk stratification

When validation-cohort patients were divided into tertiles of EHR-model-predicted risk, observed five-year cumulative incidence of the composite outcome rose sharply across tertiles, from 3.2% in the lowest tertile to 11.6% in the middle tertile and 28.9% in the highest tertile (log-rank p<0.001) (Figure 2). This roughly nine-fold gradient between the lowest and highest predicted-risk tertiles was numerically wider than the gradient produced by tertiles of Framingham-predicted risk in the same validation cohort, consistent with the model's superior discrimination on the continuous C-statistic.

4. Discussion

In this retrospective cohort of 8,542 adults with essential hypertension managed in routine care, a prognostic model built from longitudinal, structured electronic health record data discriminated incident major adverse cardiovascular events significantly better than either the Framingham general cardiovascular risk profile or the ACC/AHA pooled cohort equations, while remaining well calibrated across the range of predicted risk. The gain in discrimination was driven substantially by predictors that are simply unavailable to a static, guideline-era risk score, most notably visit-to-visit blood pressure variability, longitudinal renal function trend, and a pharmacy-fill-based adherence measure, echoing prior work showing that each of these carries independent

© 2026 Somato Publications. All rights reserved.

prognostic information in hypertensive populations even though none is captured by conventional risk equations [21,22,23,25].

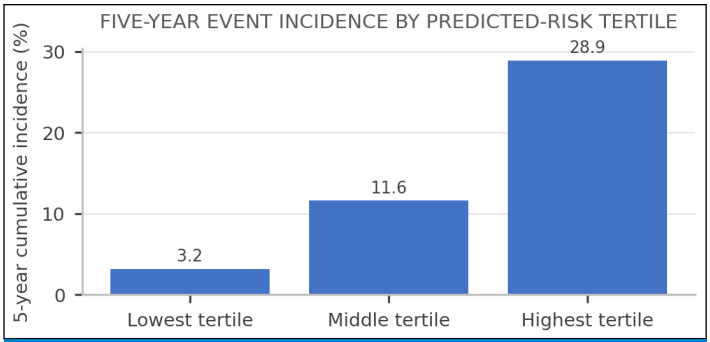

Figure 2: Observed five-year cumulative incidence of the composite cardiovascular outcome by tertile of electronic health record-model-predicted risk in the validation cohort (log-rank p<0.001).

These findings are consistent with, and extend, a broader methodological literature suggesting that EHR data mining can meaningfully improve cardiovascular risk prediction over routinely used clinical scores, while also engaging directly with that literature's central caution [13,16]. Several authors have argued that apparent gains from more flexible modeling approaches often evaporate once calibration and genuinely external validation are enforced, and that well-specified logistic regression, properly built and reported, performs comparably to more complex machine-learning algorithms in most clinical prediction tasks [18]. The present model used conventional multivariable logistic regression rather than a black-box algorithm, and its improvement over existing scores appears attributable primarily to the richer, longitudinal predictor set that EHR data uniquely make available rather than to modeling technique per se, a distinction consistent with recommended practice for transparent, reproducible prediction-model development [19,20].

The magnitude of the observed nine-fold gradient in five-year event incidence across risk tertiles has direct clinical relevance under current guidelines, both of which now recommend that absolute cardiovascular risk, not blood pressure level alone, inform the threshold and intensity of pharmacologic treatment [6,7]. A model this well calibrated and this discriminating, if deployable automatically within the EHR at the point of care, could in principle flag the minority of hypertensive patients carrying disproportionate cardiovascular risk for more intensive management of the kind shown to reduce events in trials such as SPRINT, without requiring any additional data collection beyond what the record already contains [10]. That the EHR model surfaced medication non-adherence, estimated indirectly through pharmacy fill patterns, as one of its stronger predictors is also clinically actionable, since non-adherence is a common and at least partially modifiable contributor to the persistently suboptimal population-level blood pressure control rates reported in contemporary surveillance data [5,26].

This study has several limitations. Its retrospective, single-health-system design leaves open the possibility of unmeasured confounding and limits generalizability to populations with different demographic composition, data completeness, or care patterns, an issue inherent to essentially all EHR-derived prediction research and one that systematic reviews of this literature consistently flag as the leading threat to external validity [16,17]. Outcome ascertainment relied on coded diagnoses and linked mortality data rather than adjudicated event review, which may have introduced misclassification, and predictor variables such as left ventricular hypertrophy status were missing for a meaningful minority of patients and required imputation, however methodologically appropriate [29]. The validation cohort, although temporally distinct from the derivation cohort, was drawn from the same health system rather than an external population, so the reported discrimination and calibration should be regarded as internal-temporal rather than fully external validation, and confirmation in independent health systems with differing EHR platforms and patient populations, consistent with the broader

© 2026 Somato Publications. All rights reserved.

burden of hypertension and its variable control across settings, is an important next step [2,3,4].

5. Conclusion

An electronic health record-based prognostic model incorporating longitudinal blood pressure variability, renal function trend, medication adherence, target-organ damage, and comorbidity burden discriminated major adverse cardiovascular events in adults with essential hypertension significantly better than either the Framingham risk profile or the ACC/AHA pooled cohort equations, while remaining well calibrated. These findings support electronic health record analysis as a practical and clinically meaningful prognostic tool in essential hypertension, and argue for external, multi-site validation and prospective evaluation of EHR-embedded risk stratification as a route toward more precisely targeted antihypertensive management [1,6].

6. References

1. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002;360(9349):1903-1913.

2. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol. 2020;16(4):223-237.

3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19.1 million participants. Lancet. 2017;389(10064):37-55.

4. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1223-1249.

5. Muntner P, Hardy ST, Fine LJ, et al. Trends in Blood Pressure Control Among US Adults With Hypertension, 1999-2000 to 2017-2018. JAMA. 2020;324(12):1190-1200.

6. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults. Hypertension. 2018;71(6):e13-e115.

7. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J. 2018;39(33):3021-3104.

8. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-753.

9. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S49-S73.

10. SPRINT Research Group, Wright JT Jr, Williamson JD, Whelton PK, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. N Engl J Med. 2015;373(22):2103-2116.

11. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13(6):395-405.

12. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. Annu Rev Public Health. 2016;37:61-81.

13. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944.

14. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019;380(14):1347-1358.

15. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.

© 2026 Somato Publications. All rights reserved.

16. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health record data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198-208.

17. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc. 2013;20(1):117-121.

18. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.

19. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Ann Intern Med. 2015;162(1):55-63.

20. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-1931.

21. Rothwell PM, Howard SC, Dolan E, et al. Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. Lancet. 2010;375(9718):895-905.

22. Parati G, Ochoa JE, Lombardi C, Bilo G. Assessment and management of blood-pressure variability. Nat Rev Cardiol. 2013;10(3):143-155.

23. Levey AS, Coresh J. Chronic kidney disease. Lancet. 2012;379(9811):165-180.

24. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-612.

25. Karve S, Cleves MA, Helm M, Hudson TJ, West DS, Martin BC. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. Curr Med Res Opin. 2009;25(9):2303-2310.

26. Burnier M, Egan BM. Adherence in Hypertension: A Review of Prevalence, Risk Factors, Impact, and Management. Circ Res. 2019;124(7):1124-1140.

27. Levy D, Garrison RJ, Savage DD, Kannel WB, Castelli WP. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. N Engl J Med. 1990;322(22):1561-1566.

28. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care. 2013;51(8 Suppl 3):S30-S37.

29. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. EGEMS (Wash DC). 2013;1(3):1035.

30. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-383.

31. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med. 2004;23(13):2109-2123.

32. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167-176.

33. Cox DR. Regression Models and Life-Tables. J R Stat Soc Series B. 1972;34(2):187-220.

© 2026 Somato Publications. All rights reserved.