



Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / abril – junio 2026

Volumen 1, Número 2

<https://doi.org/10.63862/rhs-v1n2-66-80-2026>

Modelo analítico longitudinal aplicado a la identificación de riesgo de abandono y deterioro clínico en programas de riesgo cardiovascular en población del Caribe colombiano

*Longitudinal analytical model applied to identifying risk of
abandonment and clinical deterioration in cardiovascular risk
programs in a Colombian Caribbean population*

Jair de Jesús Díaz Lora
Investigador independiente en analítica de datos aplicada en salud



DOI: <https://doi.org/10.63862/rhs-v1n2-66-80-2026>

**Modelo analítico longitudinal aplicado a la identificación de riesgo
de abandono y deterioro clínico en programas de riesgo
cardiovascular en población del Caribe colombiano**

Jair de Jesús Díaz Lora

Investigador independiente en analítica de datos aplicada en salud

jairdiazlora@gmail.com

<https://orcid.org/0009-0006-0871-5096>

Barranquilla, Colombia

Recibido: 2026-05-16

Aceptado: 2026-05-18

Publicado: 2026-05-18



Resumen

Las enfermedades cardiovasculares representan una de las principales causas de morbimortalidad y carga asistencial en Colombia, particularmente en pacientes diagnosticados con hipertensión arterial (HTA), diabetes mellitus (DM) y enfermedad renal crónica (ERC). Uno de los principales desafíos de los programas de riesgo cardiovascular corresponde a las dificultades relacionadas con continuidad del cuidado, adherencia terapéutica y seguimiento longitudinal de pacientes en escenarios poblacionales complejos. El objetivo del presente estudio fue desarrollar un modelo analítico longitudinal aplicado a datos reales de atención en salud para identificar patrones de abandono y deterioro clínico en programas de riesgo cardiovascular mediante integración de información clínica, operativa y de laboratorio. Se realizó un estudio observacional retrospectivo con enfoque cuantitativo sobre una población consolidada de 26.732 pacientes pertenecientes a las regiones Atlántico y Cesar durante el periodo abril 2025 – marzo 2026. La integración y procesamiento de información fue desarrollada mediante SQL Server, Python y técnicas de Machine Learning basadas en Random Forest con interpretabilidad SHAP. Los resultados evidenciaron diferencias relevantes entre escenarios poblacionales masivos y focalizados. Atlántico presentó crecimiento acelerado de cohortes incidentes y presión operativa sostenida, mientras Cesar mostró un comportamiento más estable y focalizado. El modelo predictivo obtuvo un AUC aproximado de 0.659, identificando como variables de mayor impacto la edad, el control de HTA, la continuidad asistencial y la adherencia. Se concluye que la integración de analítica descriptiva, evolutiva y predictiva fortalece procesos de continuidad del cuidado, priorización clínica y toma de decisiones basadas en datos dentro de programas de riesgo cardiovascular.

Palabras clave: Analítica de datos en salud, riesgo cardiovascular, abandono asistencial, Machine Learning, salud digital, continuidad del cuidado.

Longitudinal analytical model applied to identifying risk of abandonment and clinical deterioration in cardiovascular risk programs in a Colombian Caribbean population

Abstract

Cardiovascular diseases represent one of the leading causes of morbidity, mortality, and healthcare burden in Colombia, particularly among patients diagnosed with hypertension, diabetes mellitus, and chronic kidney disease. One of the main challenges of cardiovascular risk programs involves continuity of care, therapeutic adherence, and longitudinal patient follow-up in complex population scenarios. The aim of this study was to develop a longitudinal analytical model applied to real-world healthcare data to identify patterns of abandonment and clinical deterioration in cardiovascular risk programs through the integration of clinical, operational, and laboratory information. A retrospective observational study with a quantitative approach was conducted on a consolidated population of 26,732 patients from the Atlántico and Cesar regions during the period April 2025 – March 2026. Data integration and analytical processing were performed using SQL Server, Python, and Machine Learning techniques based on Random Forest with SHAP interpretability. The results showed significant differences between large-scale and focused population scenarios. Atlántico presented accelerated growth of incident cohorts and sustained operational pressure, while Cesar showed a more stable and focused behavior. The predictive model achieved an AUC of approximately 0.659, identifying age, hypertension control, continuity of care, and adherence as the variables with the greatest impact. It is concluded that integrating descriptive, evolutionary, and predictive analytics strengthens continuity of care, clinical prioritization, and data-driven decision-making processes within cardiovascular risk programs.

Keywords: Healthcare analytics, cardiovascular risk, clinical abandonment, machine learning, digital health, continuity of care

Introducción

Las enfermedades cardiovasculares constituyen una de las principales causas de morbimortalidad a nivel mundial y representan un desafío prioritario para los sistemas de salud debido a su impacto clínico, operativo y financiero. Según la Organización Mundial de la Salud (OMS, 2023), las enfermedades cardiovasculares continúan siendo la principal causa de muerte en el mundo, especialmente en poblaciones con factores de riesgo asociados como hipertensión arterial, diabetes mellitus y enfermedad renal crónica.

En Colombia, las patologías cardiovasculares generan una elevada carga asistencial y un incremento progresivo de costos relacionados con hospitalización, complicaciones cardiovasculares, progresión renal y eventos cerebrovasculares. Adicionalmente, los programas de riesgo cardiovascular enfrentan dificultades relacionadas con continuidad del cuidado, adherencia terapéutica y seguimiento longitudinal de pacientes en contextos institucionales complejos.

Tradicionalmente, muchos modelos de gestión clínica se han desarrollado bajo esquemas reactivos sustentados principalmente en análisis descriptivos convencionales y seguimiento operativo limitado. Esto reduce la capacidad institucional para anticipar escenarios de abandono asistencial, deterioro clínico y progresión de enfermedad en poblaciones de alto riesgo (Rajkomar et al., 2019).

De manera complementaria, Beam y Kohane (2018) señalan que uno de los principales retos de la inteligencia artificial aplicada en salud corresponde a la integración de múltiples fuentes de información y la generación de modelos interpretables y operacionalmente utilizables dentro de entornos clínicos reales.

En escenarios latinoamericanos, los desafíos relacionados con fragmentación de información, heterogeneidad poblacional y limitaciones en calidad de datos dificultan el desarrollo de estrategias analíticas longitudinales orientadas a gestión poblacional y continuidad del cuidado. En consecuencia, surge la necesidad de implementar modelos analíticos capaces de integrar información clínica, operativa y longitudinal que permitan fortalecer procesos de priorización y gestión del riesgo.

El objetivo del presente estudio fue desarrollar un modelo analítico longitudinal orientado a identificar patrones de abandono y deterioro clínico en programas de riesgo cardiovascular mediante integración de información clínica, administrativa y de laboratorio provenientes de escenarios reales de atención en salud en población del Caribe colombiano.

Estado del arte

La literatura científica reciente evidencia un crecimiento progresivo del uso de modelos analíticos y técnicas de inteligencia artificial aplicadas a gestión del riesgo cardiovascular y continuidad del cuidado en enfermedades crónicas. Diversos estudios han demostrado que la integración de información longitudinal permite fortalecer procesos de seguimiento clínico, priorización y prevención de eventos adversos.

Durante los últimos años, la analítica predictiva aplicada en salud ha adquirido creciente relevancia como herramienta orientada a fortalecer procesos de continuidad del cuidado, priorización clínica y gestión poblacional en enfermedades crónicas complejas.

Rajkomar et al. (2019) destacan que la implementación de Machine Learning en medicina permite desarrollar modelos capaces de identificar patrones clínicos complejos y mejorar procesos de toma de decisiones. Sin embargo, los autores advierten que los modelos predictivos aplicados en escenarios reales presentan importantes desafíos relacionados con calidad de datos, integración clínica y heterogeneidad poblacional.

De forma similar, Beam y Kohane (2018) sostienen que el aprovechamiento de Big Data en salud depende no solamente de la capacidad computacional, sino también de la calidad estructural de la información clínica y de la posibilidad de generar modelos interpretables y operativamente útiles.

En el contexto de enfermedades cardiovasculares, diversos estudios han demostrado la utilidad de modelos predictivos para identificación temprana de riesgo, progresión clínica y adherencia terapéutica. Topol (2019) plantea que la convergencia entre inteligencia artificial y medicina de precisión representa uno de los principales escenarios de transformación de los sistemas de salud contemporáneos.

Por otra parte, Lundberg y Lee (2017) introdujeron el enfoque SHAP como mecanismo de interpretabilidad de modelos predictivos complejos, permitiendo identificar el impacto relativo de las variables sobre los resultados analíticos y facilitando la comprensión operativa de los modelos de Machine Learning.

Breiman (2001) describió Random Forest como una técnica robusta para clasificación y predicción basada en árboles de decisión múltiples, ampliamente utilizada en escenarios clínicos debido a su capacidad para manejar grandes volúmenes de variables y relaciones no lineales.

En Latinoamérica, la evidencia relacionada con modelos longitudinales aplicados específicamente a programas de riesgo cardiovascular aún es limitada, particularmente en escenarios institucionales con integración de múltiples fuentes clínicas y operativas. La mayoría de estudios disponibles se concentra en análisis descriptivos tradicionales o cohortes limitadas, con baja incorporación de analítica predictiva aplicada a entornos reales de atención.

Adicionalmente, Collins et al. (2015) enfatizan la importancia de fortalecer transparencia metodológica y validación de modelos predictivos aplicados en medicina, promoviendo estándares reproducibles para investigación clínica basada en datos.

En este contexto, el presente estudio busca aportar evidencia aplicada sobre integración analítica longitudinal y modelos predictivos orientados a continuidad del cuidado y gestión del riesgo cardiovascular en población del Caribe colombiano, incorporando información clínica, operativa y longitudinal proveniente de escenarios reales de atención.

Metodología

Diseño del estudio

Se realizó un estudio observacional retrospectivo con enfoque cuantitativo, longitudinal, evolutivo y predictivo aplicado a pacientes pertenecientes a programas de riesgo cardiovascular.

Población de estudio

La población analizada correspondió a 26.732 pacientes pertenecientes a dos regiones del Caribe colombiano.

Región	Pacientes
Atlántico	17.922
Cesar	8.810
Total	26.732

El periodo de análisis comprendió desde abril de 2025 hasta marzo de 2026.

Variables analizadas

Variables clínicas

- Hipertensión arterial.
- Hemoglobina glicosilada (HbA1c).
- Creatinina sérica.
- Tasa de filtración glomerular (TFG).
- Clasificación de enfermedad renal crónica.

Variables operativas

- Frecuencia de atención.
- Continuidad asistencial.
- Adherencia.
- Recencia clínica.
- Cohortes prevalentes e incidentes.

Variables analíticas

- Riesgo clínico.
- Días sin atención.
- Intensidad de atención.

- Variables predictivas longitudinales.

Integración y procesamiento de datos

La integración analítica fue desarrollada mediante SQL Server utilizando consultas longitudinales y consolidación de múltiples fuentes clínicas y operativas. Se integraron bases de atención, laboratorios clínicos y seguimiento operativo mediante reglas de negocio orientadas a trazabilidad longitudinal y consistencia poblacional.

Posteriormente, el procesamiento analítico y predictivo fue desarrollado en Python utilizando bibliotecas orientadas a análisis de datos, clasificación y evaluación de modelos predictivos.

Modelo predictivo

Se implementó un modelo Random Forest orientado a identificación de riesgo de abandono y deterioro clínico.

El desempeño analítico fue evaluado mediante:

- ROC AUC.
- Validación cruzada.
- Interpretabilidad SHAP.

Consideraciones éticas

La información utilizada fue procesada bajo criterios de anonimización y confidencialidad, preservando la protección de datos personales y el uso analítico institucional conforme a principios éticos aplicables a investigación observacional basada en datos secundarios.

Resultados

Caracterización poblacional

Los resultados evidenciaron diferencias relevantes entre escenarios poblacionales masivos y focalizados.

Atlántico presentó crecimiento acelerado de cohortes incidentes y mayor presión operativa asociada al seguimiento clínico. Por su parte, Cesar mostró una operación más estable y focalizada con menor proporción de pacientes incidentes.

Tabla 1. Características regionales consolidadas

Indicador	Atlántico	Cesar
Pacientes totales	17.922	8.810
Cohorte prevalente	9.922	7.215
Cohorte incidente	7.940	1.515
Riesgo alto (%)	13.9	4.8
Adherencia promedio	0.09	0.11
AUC modelo	0.659	0.659

Hallazgo metodológico crítico

Uno de los hallazgos más relevantes correspondió a la identificación de pacientes clasificados como incidentes que previamente ya presentaban trazabilidad clínica dentro del sistema analizado. Este comportamiento evidenció limitaciones relacionadas con integración, depuración y caracterización longitudinal de cohortes poblacionales.

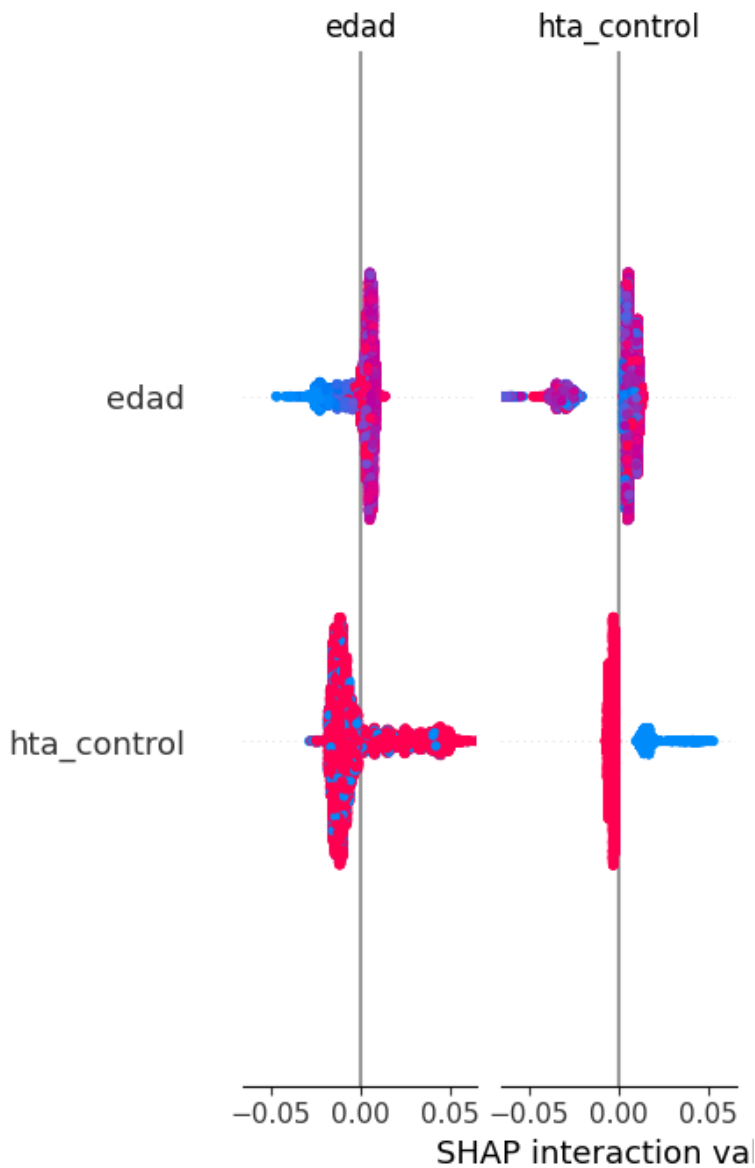
Analítica predictiva

El modelo predictivo alcanzó un AUC aproximado de 0.659, reflejando un desempeño moderado consistente con escenarios poblacionales complejos y datos provenientes de entornos reales de atención en salud.

Las variables de mayor impacto identificadas mediante interpretabilidad SHAP fueron:

- Edad.
- Control de hipertensión arterial.
- Continuidad asistencial.
- Adherencia.

Figura 1. Variables de mayor impacto identificadas mediante SHAP



Fuente: Elaboración propia mediante modelo Random Forest y análisis SHAP.

Los resultados permitieron identificar patrones relevantes de continuidad del cuidado y comportamiento poblacional asociados a deterioro clínico y riesgo de abandono asistencial.

Discusión

La aplicación de modelos analíticos longitudinales en programas de riesgo cardiovascular representa una estrategia emergente orientada a fortalecer la continuidad del cuidado y la toma de decisiones clínicas basadas en datos. En el presente estudio, la integración de múltiples fuentes clínicas y operativas permitió identificar patrones poblacionales complejos relacionados con adherencia, continuidad asistencial y deterioro clínico.

Los hallazgos obtenidos evidencian el valor de la analítica longitudinal aplicada a escenarios reales de atención en salud para fortalecer procesos de continuidad del cuidado y gestión del riesgo cardiovascular.

Uno de los principales aportes del estudio corresponde a la integración de múltiples fuentes clínicas, operativas y longitudinales dentro de un entorno poblacional complejo, permitiendo identificar limitaciones estructurales relacionadas con trazabilidad y caracterización de cohortes incidentes.

Los resultados obtenidos son consistentes con lo descrito por Rajkomar et al. (2019), quienes señalan que los modelos predictivos aplicados a entornos clínicos reales suelen presentar desempeños moderados debido a heterogeneidad poblacional, calidad variable de datos y fragmentación de información.

El AUC aproximado de 0.659 observado en el presente estudio refleja un comportamiento coherente con escenarios reales de atención en salud donde las variables clínicas y operativas presentan alta variabilidad longitudinal. Aunque el desempeño predictivo no corresponde a un modelo altamente discriminativo, sí permite identificar tendencias útiles para priorización clínica y seguimiento poblacional.

De igual manera, los resultados coinciden con Beam y Kohane (2018), quienes destacan que uno de los principales desafíos de la inteligencia artificial aplicada en medicina corresponde a la capacidad de transformar datos heterogéneos en información operacionalmente útil.

La incorporación de interpretabilidad mediante SHAP permitió identificar variables determinantes del riesgo clínico y facilitar comprensión operativa de los resultados predictivos, aspecto relevante para favorecer adopción institucional de modelos analíticos aplicados en salud.

Otro hallazgo relevante correspondió a la identificación de pacientes catalogados como incidentes pese a presentar antecedentes de trazabilidad clínica previa. Este comportamiento sugiere limitaciones institucionales relacionadas con integración de información y consistencia longitudinal de cohortes, situación que puede impactar procesos de seguimiento, continuidad del cuidado y priorización clínica.

Desde una perspectiva operativa, el estudio evidencia que los programas de riesgo cardiovascular requieren modelos analíticos capaces de integrar información evolutiva y longitudinal para anticipar escenarios de abandono asistencial y deterioro clínico.

Entre las principales limitaciones del estudio se encuentran la dependencia de calidad de datos institucionales, variabilidad operacional entre regiones y ausencia de algunas variables sociodemográficas y farmacológicas potencialmente relevantes para mejorar desempeño predictivo.

No obstante, el estudio aporta evidencia aplicada sobre implementación de analítica longitudinal y predictiva en escenarios reales de atención en salud dentro de población colombiana, aspecto aún limitado en la literatura regional.

Conclusiones

La integración de analítica longitudinal aplicada a datos reales de programas de riesgo cardiovascular permitió identificar patrones relevantes de abandono asistencial, deterioro clínico y progresión de riesgo en poblaciones complejas.

El estudio evidenció que las limitaciones relacionadas con calidad, integración y trazabilidad de datos afectan directamente la caracterización poblacional y continuidad del seguimiento clínico en programas de riesgo cardiovascular.

La implementación de modelos descriptivos, evolutivos y predictivos facilita la transición desde enfoques reactivos hacia esquemas de gestión basada en riesgo y priorización clínica sustentada en información integrada y longitudinal.

Los resultados obtenidos demuestran que variables operativas como continuidad asistencial y adherencia poseen un impacto relevante sobre la identificación de riesgo clínico y abandono asistencial.

Finalmente, el estudio evidencia el potencial de la analítica aplicada y el Machine Learning como herramientas complementarias para fortalecer procesos institucionales de gestión del riesgo, continuidad del cuidado y toma de decisiones basadas en datos dentro de escenarios reales de atención en salud.

Como línea futura de investigación se recomienda incorporar variables sociodemográficas, farmacológicas y determinantes sociales que permitan mejorar capacidad predictiva y robustez longitudinal de los modelos analíticos aplicados a programas de enfermedades crónicas.

Referencias bibliográficas

- American Diabetes Association. (2025). Standards of medical care in diabetes—2025. *Diabetes Care*, 48(Suppl. 1), S1–S350. <https://doi.org/10.2337/dc25-SINT>
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/jama.2017.18391>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162(1), 55–63. <https://doi.org/10.7326/M14-0697>

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

Organización Mundial de la Salud. (2023). Enfermedades cardiovasculares. [https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMra1814259>

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>

Anexos

Anexo 1. Variables longitudinales utilizadas en el modelo analítico

- Edad.
- Cohorte clínica.
- Control de hipertensión arterial.
- HbA1c.
- Continuidad asistencial.
- Días sin atención.
- Frecuencia de atención.
- Adherencia.
- Riesgo clínico.
- Clasificación ERC.

Anexo 2. Componentes tecnológicos utilizados

- SQL Server.
- T-SQL.
- Python.
- Random Forest.
- SHAP.
- Procesamiento longitudinal de datos.
- Integración de múltiples fuentes clínicas y operativas.

Declaraciones finales:

Conflicto de intereses: Los autores declaran que no existe conflicto de interés posible.

Financiamiento: No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento: N/A

Nota editorial: El artículo no es producto de una publicación anterior.