ISSN-e: 2737-6419 Revista Athenea
Período: octubre-diciembre de 2026 Vol.7, Número 26, (pp. 60–71)

Artículo de investigación https://doi.org/10.47460/athenea.v7i26.158

Modelado estocástico y simulación computacional de trayectorias de recuperación en cirugía estética

Franyelit Suárez-Carreño*
https://orcid.org/0000-0002-8763-5513
fsuarez@mgu.us
Modern Global University
Puerto Rico, Estados Unidos

Wilfredo Fariñas-Coronado
https://orcid.org/0000-0003-2095-5755
wfarinas@mgu.us
Modern Global University
Puerto Rico, Estados Unidos

Autor de correspondencia: fsuarez@mgu.us

Recibido: (05/08/2026), Aceptado: (01/11/2026)

Resumen. Se desarrolló un modelo matemático-computacional para generar datos longitudinales simulados de recuperación posterior a una liposucción aislada hipotética. El sistema integró procesos estocásticos correlacionados, heterogeneidad individual, eventos adversos competitivos y mecanismos selectivos de observación. Se generó una cohorte de 10.000 individuos virtuales con 80.000 evaluaciones programadas durante 180 días. Las trayectorias mostraron disminuciones progresivas de los índices de edema y dolor, acompañadas de incrementos de función y cicatrización. Se registraron 121 primeros eventos adversos agudos, 219 irregularidades del contorno y 6.999 observaciones faltantes. La disponibilidad de estados latentes completos permitió establecer referencias computacionales para futuras evaluaciones estadísticas. Los resultados demostraron la reproducibilidad interna del sistema, sin constituir evidencia clínica. SIM-LIPO 3.2 ofrece un recurso documentado para estudiar métodos longitudinales y desarrollar experimentos computacionales bajo supuestos controlados.

Palabras clave: simulación computacional, modelado estocástico, datos longitudinales, liposucción, eventos competitivos, reproducibilidad científica.

Stochastic Modeling and Computational Simulation of Recovery Trajectories in Cosmetic Surgery

Abstract. A mathematical and computational model was developed to generate simulated longitudinal data on recovery following hypothetical isolated liposuction. The system integrated correlated stochastic processes, individual heterogeneity, competing adverse events, and selective observation mechanisms. A cohort of 10,000 virtual individuals was generated, with 80,000 scheduled assessments over 180 days. The trajectories showed progressive decreases in edema and pain indices, accompanied by improvements in function and healing indices. A total of 121 first acute adverse events, 219 contour irregularities, and 6,999 missing observations were recorded. The availability of complete latent states made it possible to establish computational reference values for future statistical evaluations. The results demonstrated the internal reproducibility of the system without constituting clinical evidence. SIM-LIPO 3.2 provides a documented resource for studying longitudinal methods and conducting computational experiments under controlled assumptions.

Keywords: computational simulation, stochastic modeling, longitudinal data, liposuction, competing risks, scientific reproducibility.

I. INTRODUCCIÓN

La liposucción y los procedimientos de remodelación del contorno corporal plantean problemas de investigación que abarcan la recuperación posoperatoria, la aparición de eventos adversos y la variación temporal de los resultados. Estas dimensiones no siempre pueden resumirse mediante una única medición final, pues describen procesos diferentes y potencialmente relacionados. Una revisión sistemática y metaanálisis de complicaciones de liposucción documentó la diversidad de desenlaces que se investigan, entre ellos seroma, hematoma, infección, tromboembolismo venoso y alteraciones del contorno [1]. Tales hallazgos justifican representar de manera diferenciada el curso longitudinal de la recuperación y los mecanismos de ocurrencia de eventos. No obstante, las estimaciones publicadas dependen de poblaciones y procedimientos específicos y no autorizan extrapolaciones automáticas a escenarios computacionales hipotéticos.

El análisis de sistemas posoperatorios mediante simulación permite formular hipótesis estructurales y examinar el comportamiento de modelos bajo supuestos controlados sin recurrir a información individual de pacientes. Su utilidad reside en disponer de estados internos conocidos, modificar parámetros y observar cómo cambian los resultados. La literatura de verificación y validación de modelos de simulación distingue entre comprobar que una implementación ejecuta correctamente sus especificaciones y demostrar que representa adecuadamente el sistema de interés [2]. Esta distinción resulta esencial en investigación biomédica computacional: una base reproducible y coherente puede ser valiosa para comparar métodos estadísticos sin constituir por ello evidencia clínica sobre frecuencias de complicaciones o eficacia terapéutica.

Los registros longitudinales añaden dificultades analíticas cuando las mediciones de un mismo individuo están correlacionadas y existen observaciones faltantes. La teoría de Rubin sobre inferencia con datos incompletos advierte que el mecanismo de ausencia condiciona la interpretación de las estimaciones [3]. Por otra parte, cuando se estudia el primer evento adverso entre varios tipos mutuamente excluyentes, la estructura de riesgos competitivos requiere distinguir las funciones de incidencia acumulada y los riesgos asociados con causas específicas [4]. La incorporación explícita de estos fenómenos a un simulador crea un entorno conocido para ensayar métodos de estimación, verificar sesgos computacionales y estudiar la sensibilidad a distintos supuestos de observación.

La posibilidad de reproducir resultados constituye otro requisito central. Las prácticas de investigación computacional reproducible recomiendan conservar el código, los parámetros y el origen de cada resultado [5], mientras que los principios FAIR favorecen que los recursos de investigación sean localizables, accesibles, interoperables y reutilizables [6]. Bajo estos criterios se desarrolló SIM-LIPO 3.2, un recurso longitudinal generado mediante procedimientos matemáticos y publicado con sus archivos de datos y componentes de reproducción en Zenodo [7]. El repositorio ofrece una referencia estable para identificar el recurso; esa disponibilidad, por sí sola, no certifica la pertinencia clínica de su proceso generador.

El objetivo de este estudio fue diseñar, implementar y describir un modelo matemático-computacional capaz de producir trayectorias longitudinales simuladas de recuperación tras una liposucción aislada hipotética, incorporando heterogeneidad individual, dependencia estocástica entre dimensiones, primeros eventos adversos agudos competitivos, un desenlace estético separado y ausencia selectiva de mediciones. Como aportación metodológica, se presenta una cohorte computacional de 10.000 individuos virtuales con ocho momentos programados de observación durante 180 días, junto con reglas de generación y procedimientos de comprobación reproducibles. El estudio evalúa propiedades del sistema simulado y su utilidad como banco de pruebas estadístico; no estima resultados en pacientes reales ni valida recomendaciones clínicas.

II. FUNDAMENTOS TEÓRICOS

A. Modelado computacional y validez de los modelos

Un modelo de simulación es una representación formal de un sistema cuya conducta se obtiene mediante reglas matemáticas y algoritmos. En un modelo estocástico, los estados resultantes dependen de parámetros deterministas y realizaciones aleatorias. Para que sus productos sean interpretables, deben explicitarse las variables de entrada, las ecuaciones de transición, las restricciones y los mecanismos de observación. La verificación examina la correspondencia entre código y especificaciones, mientras que

la validación externa contrasta el modelo con evidencia del fenómeno que se pretende representar [2]. Por consiguiente, las pruebas de integridad de archivos, límites de índices o reproducción por semilla son necesarias, pero no equivalen a calibración con pacientes reales.

En SIM-LIPO, cada individuo virtual posee atributos basales y estados longitudinales latentes. Se consideran cuatro dimensiones adimensionales de recuperación: edema, dolor, función y cicatrización. Sus escalas normalizadas constituyen índices computacionales definidos por el diseño y no instrumentos clínicos validados. Los valores que surgen del modelo son consecuencias de sus supuestos, por lo que su interpretación apropiada se refiere al comportamiento del algoritmo y a la evaluación de herramientas analíticas que operen sobre él [7].

B. Dependencia longitudinal y procesos estocásticos

Las mediciones repetidas no suelen ser independientes: el valor previo de una dimensión puede contribuir a explicar su estado siguiente. Para representar ese carácter dinámico, SIM-LIPO utiliza una estructura discreta inspirada en procesos de reversión a un objetivo temporal, próximos a la familia de Ornstein–Uhlenbeck, con perturbaciones aleatorias correlacionadas y heterogeneidad entre individuos [7]. La transformación logística del estado subyacente mantiene los índices dentro del intervalo abierto $(0, 1)$, sin atribuirles unidades clínicas ni umbrales diagnósticos. La dependencia entre perturbaciones se construye mediante una matriz de correlación y una factorización matricial que permite generar variaciones relacionadas entre las cuatro dimensiones.

Esta formulación permite examinar cómo la variabilidad individual, la amplitud del ruido y la velocidad de acercamiento a objetivos modifican las trayectorias computacionales. Sin embargo, representar dependencia temporal no demuestra que el proceso matemático reproduzca mecanismos biológicos particulares. Su ventaja en este proyecto consiste en proporcionar secuencias estructuradas y verificables con las que se puedan estudiar procedimientos longitudinales.

C. Eventos adversos y riesgos competitivos

Los eventos adversos no constituyen necesariamente desenlaces intercambiables. El metaanálisis de Comerci y colaboradores evidencia la diversidad de complicaciones investigadas tras liposucción [1]. Desde la teoría de riesgos competitivos, cuando se considera únicamente el primer evento agudo de cada individuo, la ocurrencia de un tipo impide registrar otro como primer evento en ese mismo esquema. Fine y Gray formularon un enfoque para analizar la subdistribución de un riesgo competitivo, destacando la diferencia entre el riesgo de una causa y su incidencia acumulada en presencia de causas alternativas [4].

El simulador utiliza una regla de asignación de primer evento agudo entre seroma, hematoma, infección y tromboembolismo venoso. Esta exclusividad pertenece a la definición del resultado simulado, no a una afirmación de que una persona real no pueda experimentar complicaciones múltiples. Las irregularidades del contorno se registran aparte, como otro desenlace del sistema. Las frecuencias obtenidas son productos del mecanismo generador y no deben compararse directamente con las prevalencias clínicas del metaanálisis como si procedieran de muestras equivalentes.

D. Observaciones incompletas y mecanismo de ausencia

La falta de una medición longitudinal puede alterar las distribuciones observadas y afectar modelos predictivos o estimadores de trayectorias. Rubin introdujo una distinción fundamental entre mecanismos de ausencia vinculados o no a valores no observados [3]; Little y Rubin desarrollaron posteriormente estrategias de análisis e inferencia para este tipo de problemas [8]. En simulación resulta especialmente útil conservar la tabla completa de estados latentes y generar por separado la tabla visible, pues así puede medirse el error de reconstrucción respecto de valores que se conocen por construcción.

SIM-LIPO incorpora un mecanismo de observación dependiente de información del proceso simulado. La denominación estadística precisa de ese mecanismo debe justificarse respecto de las variables efectivamente disponibles para el analista; no corresponde clasificarlo automáticamente como ausencia al azar (MAR) o no al azar (MNAR) solo por depender de estados internos. Esta cautela es importante para evitar conclusiones incorrectas sobre el comportamiento de procedimientos de imputación.

E. Exposiciones de seguimiento e interpretación causal

La incorporación de una variable de seguimiento o apoyo posoperatorio a un conjunto longitudinal no demuestra un efecto terapéutico. En los estudios con exposiciones que pueden variar o seleccionarse a lo largo del tiempo, la historia previa del individuo puede relacionarse tanto con la asignación de la exposición como con su desenlace, lo que dificulta atribuir diferencias observadas a una causa específica [9]. El recurso SIM-LIPO contiene una asignación de seguimiento de apoyo dependiente de antecedentes simulados, pero fija en cero su efecto verdadero sobre las trayectorias. En consecuencia, esta variable puede utilizarse para ensayar métodos frente a asociaciones inducidas por selección, pero no para recomendar una intervención ni cuantificar eficacia clínica.

F. Reproducibilidad, reutilización y límites de extrapolación

La transparencia metodológica requiere conservar semillas, versiones, especificaciones y comprobaciones de integridad, de acuerdo con las recomendaciones para investigación computacional reproducible [5]. Los principios FAIR aportan una orientación complementaria para facilitar el descubrimiento, acceso e interpretación de conjuntos de datos [6]. SIM-LIPO 3.2 se distribuye en Zenodo con un identificador persistente y materiales de reproducción [7], lo que permite revisar las reglas de construcción y desarrollar análisis secundarios sobre un escenario controlado.

La reutilización científicamente responsable exige declarar el propósito del conjunto. Es idóneo como entorno experimental para explorar métodos computacionales, depuración de análisis longitudinales, tratamiento de ausencia de observaciones y modelado de eventos competitivos. En cambio, no permite inferir incidencia poblacional, probabilidades individuales de complicación, seguridad relativa de procedimientos o eficacia de cuidados clínicos. La validación externa futura demandaría evidencia observacional independiente, criterios éticos apropiados y análisis explícitos de sensibilidad.

III. METODOLOGÍA

A. Diseño y alcance del estudio

Se desarrolló un estudio metodológico de simulación computacional destinado a construir y documentar un conjunto de datos longitudinales para experimentos estadísticos y de reproducibilidad sobre recuperación posoperatoria simulada tras liposucción aislada. El diseño no incluyó reclutamiento, intervención ni observación de seres humanos. Todos los registros se generaron mediante distribuciones probabilísticas y transiciones estocásticas especificadas en código Python. El recurso definitivo se depositó como SIM-LIPO, versión 3.2, con identificador DOI 10.5281/zenodo.23460437 [7].

La unidad individual fue el paciente virtual, identificado mediante una clave única; la unidad longitudinal fue el registro correspondiente a un paciente y un tiempo programado. Se construyeron $N = 10\,000$ individuos virtuales, con ocho evaluaciones por individuo realizadas en los días 0, 1, 3, 7, 14, 30, 90 y 180. Esto representó un total de $80\,000$ registros longitudinales programados. La población se restringió a liposucción aislada: no se simularon abdominoplastia, liposucción de alta definición, transferencia de grasa ni intervenciones combinadas. La elección de este ámbito se correspondió con las restricciones temáticas del metaanálisis de Comerci y colaboradores [1], utilizado exclusivamente como fuente orientadora de algunos porcentajes de escenarios adversos.

B. Arquitectura y variables del simulador

Cada paciente i se caracterizó mediante un vector $z_i = (A_i, B_i, S_i, C_i, V_i, D_i)$, donde A_i representó edad (años), B_i índice de masa corporal (kg/m^2), S_i tabaquismo actual (indicador 0/1), C_i comorbilidad simulada (indicador 0/1), V_i volumen aspirado (mL) y D_i duración del procedimiento (minutos). La información se distribuyó en archivos relacionados por `patient_id`, con un registro de procedimiento por individuo y un calendario longitudinal fijo.

La evolución se representó mediante cuatro dimensiones normalizadas: edema (E), dolor (P), función (F) y cicatrización (H). Se reservaron dos vistas de estos estados: `latent_truth.csv`, que almacena los valores computacionales completos, y `followup.csv`, que conserva la misma estructura temporal, pero deja en blanco los índices cuando el generador marca una visita como no observada. Estos estados son índices matemáticos adimensionales y no corresponden a escalas clínicas validadas.

Tabla 1. Distribuciones y supuestos utilizados en la generación de variables basales.

Variable	**Regla computacional**	**Naturaleza**
Edad	$A_i = \text{round}(18 + 52U_i, 1)$; $U_i \sim \text{Beta}(2.4, 3.0)$	Supuesto exploratorio
IMC	$B_i = \text{round}(18.5 + 15.5W_i, 2)$; $W_i \sim \text{Beta}(2.7, 4)$	Supuesto exploratorio
Tabaquismo	$S_i \sim \text{Bernoulli}(0.13)$	Supuesto exploratorio
Comorbilidad	$C_i \sim \text{Bernoulli}(0.14)$	Supuesto exploratorio
Duración	$D_i = \text{round}(65 + 140R_i)$; $R_i \sim \text{Beta}(2.5, 3)$	Supuesto exploratorio
Aspirado	$V_i = \text{round}(450 + 3350Q_i)$; $Q_i \sim \text{Beta}(2.5, 3)$	Supuesto exploratorio

Las variables beta se generaron a partir de extracciones separadas del generador seudoaleatorio, con una semilla fija. Se utilizó la intensidad procedimental adimensional $I_i = (V_i - 450)/3350$. No se estimaron distribuciones, regresiones ni parámetros individuales a partir de historias clínicas. La edad y las demás covariables se redondearon en el momento de la simulación con las precisiones indicadas.

C. Estructura de dependencia y condiciones iniciales

La dependencia entre las cuatro dimensiones del estado se incorporó mediante una matriz simétrica de correlaciones de referencia, definida antes de generar los datos:

$$R = \begin{pmatrix} 1 & 0.36 & -0.20 & -0.15 \\ 0.36 & 1 & -0.27 & -0.12 \\ -0.20 & -0.27 & 1 & 0.35 \\ -0.15 & -0.12 & 0.35 & 1 \end{pmatrix} \tag{1}$$

La matriz se factorizó mediante Cholesky, $LL^{\mathsf{T}} = R$. Para cada individuo se obtuvo un vector de efectos aleatorios de dimensión cuatro, $b_i = 0.24L\eta_i$, con $\eta_i \sim \mathcal{N}(0, I_4)$. Las condiciones iniciales del estado latente se fijaron como

$$q_i(0) = \begin{pmatrix} 1.3 \\ 1.1 \\ -0.7 \\ -1.1 \end{pmatrix} + b_i + I_i \begin{pmatrix} 0.65 \\ 0.40 \\ -0.25 \\ -0.25 \end{pmatrix} \tag{2}$$

En consecuencia, los índices iniciales fueron heterogéneos y correlacionados. Tanto R como las constantes de la ecuación (2) se establecieron como parámetros de diseño, sin pretensión de representar correlaciones clínicas observadas.

D. Modelo dinámico estocástico de recuperación

Sea $q_{i,k,j}$ el estado latente del paciente i, para la dimensión $k \in \{E, P, F, H\}$ en el instante programado t_j, y $\Delta_j = t_j - t_{j-1}$. La evolución se calculó mediante una transición autorregresiva de reversión hacia un objetivo dependiente del tiempo, inspirada en la estructura de Ornstein–Uhlenbeck:

$$\begin{aligned} q_{i,k,j} = \mu_{i,k}(t_j) \\ + e^{-\kappa_k \Delta_j} \left[q_{i,k,j-1} - \mu_{i,k}(t_j) \right] \\ + \sigma_k \sqrt{\frac{1 - e^{-2\kappa_k \Delta_j}}{2\kappa_k}}\, \varepsilon_{i,k,j} \end{aligned} \tag{3}$$

La perturbación multivariada se obtuvo mediante $\varepsilon_{i,j} = L\xi_{i,j}$, con $\xi_{i,j} \sim \mathcal{N}(0, I_4)$ independiente

entre pasos de tiempo. El esquema coincide con la forma exacta de una transición de Ornstein–Uhlenbeck con objetivo constante dentro de cada intervalo; sin embargo, el objetivo $\mu_i(t_j)$ cambia entre momentos y depende de estados anteriores. Por ello se describió como un esquema discreto inspirado en Ornstein–Uhlenbeck, sin atribuirle una solución exacta global de una ecuación diferencial estocástica con deriva variable.

Los parámetros de reversión y difusión fueron, respectivamente,

$$\begin{aligned} \kappa &= (0.090, 0.110, 0.055, 0.045)\ \mathrm{d}^{-1}, \\ \sigma &= (0.27, 0.25, 0.22, 0.20) \end{aligned} \tag{4}$$

El objetivo de estado se actualizó mediante funciones deterministas diseñadas para inducir disminución simulada de edema y dolor, e incremento de función y cicatrización. Con $G_i(t) = \mathrm{logit}^{-1}(q_{i,E}(t))$ y con indicadores $J^{SH}_{i,j-1}$ de seroma o hematoma ocurrido hasta t_{j-1}, e $J^{IN}_{i,j-1}$ de infección ocurrida hasta ese momento, las cuatro componentes fueron:

$$\begin{aligned} \mu_{i,E}(t_j) = &-2.7 + 3.5e^{-0.038t_j} + 0.45I_i \\ &+ 0.35J^{SH}_{i,j-1} + b_{i,E}, \\ \mu_{i,P}(t_j) = &-3.0 + 3.6e^{-0.055t_j} + 0.25I_i \\ &+ 0.30J^{IN}_{i,j-1} + b_{i,P}, \\ \mu_{i,F}(t_j) = &\ 2.8 - 3.8e^{-0.030t_j} \\ &- 0.35G_i(t_{j-1}) + b_{i,F}, \\ \mu_{i,H}(t_j) = &\ 3.2 - 4.4e^{-0.020t_j} \\ &- 0.50J^{IN}_{i,j-1} + b_{i,H} \end{aligned} \tag{5}$$

La variable G_i corresponde al edema latente transformado del instante anterior. Los indicadores de eventos no se activaron hasta que el evento había ocurrido al cierre del tiempo anterior; de este modo, el programa evitó usar un evento futuro para modificar estados previos. No se modelaron mortalidad, reintervenciones, gravedad del evento ni trayectorias específicas de tratamientos reales.

Los estados latentes se proyectaron al intervalo abierto $(0, 1)$ mediante una transformación logística:

$$\begin{aligned} X_{i,k,j} &= \mathrm{logit}^{-1}(q_{i,k,j}) \\ &= \frac{1}{1 + e^{-q_{i,k,j}}}, \qquad 0 < X_{i,k,j} < 1 \end{aligned} \tag{6}$$

Los valores se almacenaron con seis posiciones decimales. Esta normalización garantiza acotación matemática, pero no otorga a los índices significado diagnóstico ni equivalencia con resultados de instrumentos clínicos.

E. Generación de eventos adversos con riesgos competitivos

Se definieron cuatro causas mutuamente excluyentes para el primer evento agudo: seroma, hematoma, infección y tromboembolia venosa. Las proporciones p_c empleadas como puntos de partida exploratorios fueron 0.0065, 0.0027, 0.00020 y 0.00017, respectivamente. La irregularidad del contorno se simuló por separado, con $p_{\mathrm{contorno}} = 0.0235$. Estos números coincidieron en escala con las proporciones publicadas por Comerci et al. [1], pero no constituyeron una calibración de tasas a 180 días: la fuente no establece una ventana de seguimiento uniforme de esa duración y presenta heterogeneidad entre estudios.

Para introducir heterogeneidad matemática, se asignó a cada individuo una susceptibilidad latente $U_i \sim \mathcal{N}(0, 1)$ y un multiplicador de riesgo exclusivamente hipotético:

$$r_i = \exp\left[0.15\frac{B_i - 25}{5} + 0.12S_i + 0.10C_i + 0.18U_i - \frac{0.18^2}{2}\right] \tag{7}$$

Para cada causa aguda c se especificó una intensidad constante en el horizonte hipotético de $T = 180$ días:

$$\lambda_{ic} = r_i \frac{-\ln(1 - p_c)}{T}, \quad \Lambda_i = \sum_{c \in \mathcal{C}} \lambda_{ic} \tag{8}$$

En presencia de al menos una intensidad positiva, el tiempo del primer evento se extrajo de $T_i^* \sim \mathrm{Exp}(\Lambda_i)$. Si $T_i^* \leq 180$, se seleccionó la causa con probabilidades condicionadas

$$\Pr\left(C_i^* = c \mid T_i^* \leq 180, \lambda_i\right) = \frac{\lambda_{ic}}{\Lambda_i} \tag{9}$$

El evento se guardó con su tiempo redondeado a dos decimales. Para el desenlace estético separado se calculó

$$\pi_{i,\mathrm{contorno}} = 1 - (1 - p_{\mathrm{contorno}})^{r_i}, \quad T_{i,\mathrm{contorno}} \sim \mathcal{U}(30, 180) \tag{10}$$

con extracción de tiempo condicionada a la ocurrencia del desenlace binario. La irregularidad del contorno podía coexistir con un evento agudo; no se incorporó como quinta causa competitiva. La suma de probabilidades marginales por causa no se interpretó como riesgo global clínico, ni la ocurrencia simulada se presentó como evidencia epidemiológica.

F. Asignación de la exposición abstracta de apoyo

En cada visita posterior al inicio, una persona virtual todavía no expuesta podía pasar al estado de exposición de apoyo mediante una regla probabilística dependiente del edema latente del instante anterior y del indicador de tabaquismo:

$$\Pr(Z_{i,j} = 1 \mid Z_{i,j-1} = 0) = \mathrm{logit}^{-1}\left[-3.4 + 2.4G_i(t_{j-1}) + 0.4S_i\right] \tag{11}$$

Una vez activada, la exposición se mantuvo activa hasta el final del horizonte simulado. El tiempo de inicio registrado correspondió al instante anterior. El efecto causal de esta variable se fijó exactamente en cero tanto para estados longitudinales como para eventos adversos; por lo tanto, no constituye un tratamiento clínico evaluado ni sustenta inferencias de efectividad. Aunque el código contiene comentarios sobre historia observada, la implementación utiliza el valor latente de edema del paso anterior, incluso cuando dicho paso fue marcado como no observado.

G. Mecanismo de observaciones faltantes y datos completos

El estado de observación de la visita basal se fijó en uno. En visitas posteriores, el indicador $O_{i,j}$ se obtuvo con una regla de Bernoulli basada en el edema latente actual, tabaquismo y tiempo:

$$O_{i,0} = 1, \quad \Pr(O_{i,j} = 1) = \mathrm{logit}^{-1}\left[2.6 - 0.85X_{i,E,j} - 0.25S_i + 0.002t_j\right], \quad j > 0 \tag{12}$$

Si $O_{i,j} = 0$, los cuatro índices se dejaron vacíos en `followup.csv`, aunque los campos de tiempo, identificador y exposición permanecieron registrados. El archivo `latent_truth.csv` preservó los valores generados antes del enmascaramiento y mantuvo el mismo indicador $O_{i,j}$. Debido a que la ecuación (12) utiliza el edema latente del momento actual —que queda sin observar cuando $O_{i,j} = 0$—, no se afirmó que el mecanismo satisficiera estrictamente la condición MAR (*missing at random*) bajo los datos visibles. El recurso permite estudiar el comportamiento de métodos estadísticos con conocimiento del estado latente simulado, pero no representa un registro clínico real de pérdidas de seguimiento.

H. Implementación, reproducibilidad y auditoría técnica

Se utilizó Python con NumPy, un generador seudoaleatorio inicializado con semilla 261021, y el modo `competing` para los primeros eventos agudos. El procedimiento generó las tablas `patients.csv`, `procedures.csv`, `treatments.csv`, `followup.csv`, `latent_truth.csv` y `complications.csv`, junto con diccionario de datos, parámetros, código y verificaciones. Se mantuvieron identificadores relacionales consistentes. La versión 3.2 corrigió un truncamiento material del archivo `followup.csv` en la versión 3.1, reconstruyendo exactamente las 80.000 filas desde `latent_truth.csv` y su máscara de observación, sin alterar los pacientes, trayectorias latentes, procedimientos ni desenlaces estocásticos publicados [7].

El programa `reproduce_dataset.py` de la versión 3.2 incorpora una verificación del número de filas realmente escritas y reconstruye la vista observada a partir del archivo completo de estados. Como controles de integridad se verificaron: unicidad de claves primarias, ocho tiempos por paciente, pertenencia de los índices a $(0, 1)$, ausencia de registros huérfanos, exclusividad del primer evento agudo y correspondencia exacta entre los índices visibles y los estados completos. Se recalcularon las sumas criptográficas SHA-256 de los archivos depositados. Estas pruebas establecieron consistencia computacional, no ajuste de parámetros a pacientes reales ni validación predictiva externa.

En consecuencia, las inferencias admisibles se restringieron al estudio del comportamiento del simulador, a la comprobación de procesos analíticos, a experimentos sobre datos faltantes y a ejercicios de reproducibilidad. Se excluyeron expresamente aplicaciones destinadas a pronóstico de complicaciones individuales, triage, selección terapéutica, aprobación regulatoria o medición de la eficacia clínica.

IV. RESULTADOS Y DISCUSIÓN

A. Estructura de la cohorte computacional

Se generó una cohorte de 10.000 adultos virtuales sometidos a liposucción aislada, asociada a 10.000 registros de procedimiento. El diseño fijó ocho momentos de seguimiento (días 0, 1, 3, 7, 14, 30, 90 y 180), equivalentes a 80.000 registros longitudinales programados. La edad media generada fue de 41,03 ± 10,19 años y el índice de masa corporal de 24,71 ± 2,74 kg/m^2. Se asignó tabaquismo actual a 1.318 pacientes (13,18 %), y una condición de comorbilidad simulada a 1.420 (14,20 %). La duración media de los procedimientos fue de 128,70 ± 27,38 minutos; el volumen aspirado medio fue de 1.972,44 ± 651,95 mL. Los valores se obtuvieron de distribuciones paramétricas predefinidas, sin ajuste a historias clínicas reales.

Tabla 2. Características descriptivas de la cohorte y los procedimientos simulados.

Indicador	**Resultado**
Pacientes virtuales	10.000
Procedimientos	10.000
Edad, media ± DE (años)	41,03 ± 10,19
IMC, media ± DE (kg/m^2)	24,71 ± 2,74
Tabaquismo simulado, n (%)	1.318 (13,18 %)
Comorbilidad simulada, n (%)	1.420 (14,20 %)
Duración, media ± DE (min)	128,70 ± 27,38
Volumen aspirado, media ± DE (mL)	1.972,44 ± 651,95

B. Trayectorias de recuperación simuladas

Los cuatro índices latentes permanecieron estrictamente dentro del intervalo abierto $(0, 1)$, como exige la transformación logística del modelo. La Tabla 3 presenta las medias de los estados completos generados, incluidos los tiempos marcados como no observados. Estos valores describen las trayectorias impuestas y realizadas por el simulador; no expresan velocidad de recuperación clínica documentada. La Fig. 1 representa los cuatro índices a lo largo de los ocho tiempos.

Tabla 3. Evolución media de los cuatro índices de estado latente simulados.

Día	Edema	Dolor	Función	Cicatrización
0	0,828	0,780	0,309	0,232
1	0,817	0,764	0,306	0,235
3	0,789	0,723	0,307	0,242
7	0,721	0,613	0,321	0,263
14	0,576	0,404	0,381	0,317
30	0,292	0,146	0,589	0,500
90	0,101	0,063	0,898	0,892
180	0,090	0,061	0,927	0,945

La evolución temporal de los indicadores presentada en la Tabla 3 permite identificar las variaciones de las cuatro dimensiones de recuperación durante el período simulado. Para facilitar la interpretación de estas tendencias y visualizar sus diferencias a lo largo de los 180 días de seguimiento, la Fig. 1 representa gráficamente las trayectorias medias de edema, dolor, función y cicatrización.

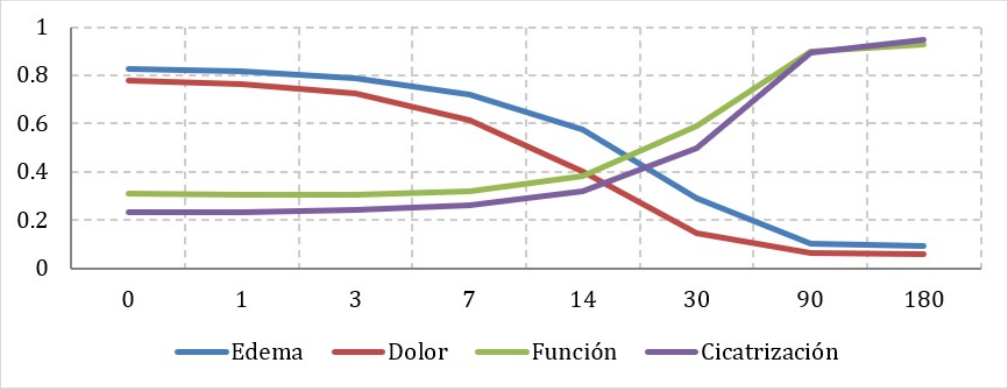


Fig. 1. Trayectorias longitudinales medias simuladas de edema, dolor, función y cicatrización.

C. Observación longitudinal y datos no observados

De las 80.000 visitas programadas, 73.001 (91,25 %) presentaron índices visibles y 6.999 (8,75 %) se marcaron como no observadas. La medición basal fue completa por diseño. Las variables latentes completas continuaron disponibles en un archivo separado, lo que posibilita comparar el estado generado con métodos de imputación o inferencia bajo mecanismos de ausencia simulados. La Tabla 4 desglosa los ocho momentos; la Fig. 2 muestra la fracción visible de visitas por tiempo.

Tabla 4. Disponibilidad de los registros longitudinales según momento de seguimiento.

Día	Programadas	Visibles	No visibles	Visibles (%)
0	10.000	10.000	0	100,00
1	10.000	8.624	1.376	86,24
3	10.000	8.704	1.296	87,04
7	10.000	8.777	1.223	87,77
14	10.000	8.940	1.060	89,40
30	10.000	9.135	865	91,35
90	10.000	9.362	638	93,62
180	10.000	9.459	541	94,59

Los resultados de la Tabla 4 muestran la distribución temporal de las evaluaciones programadas, los registros disponibles y las observaciones faltantes. La Fig. 2 complementa esta información mediante una representación gráfica que permite apreciar el comportamiento del proceso de observación durante el seguimiento longitudinal de la cohorte virtual.

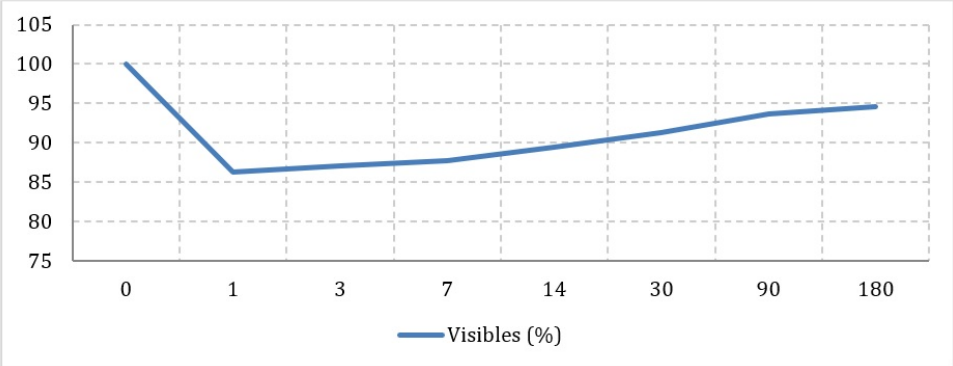


Fig. 2. Proporción de registros longitudinales visibles en cada tiempo programado.

D. Primeros eventos agudos competitivos y desenlace estético

El mecanismo de riesgos específicos por causa produjo 121 primeros eventos agudos (1,21 % de los 10.000 pacientes virtuales). Entre estos se registraron 69 seromas, 48 hematomas, dos infecciones y dos eventos tromboembólicos venosos. Los primeros eventos agudos fueron mutuamente excluyentes por construcción. La irregularidad del contorno se generó como desenlace cosmético independiente del mecanismo competitivo y se presentó en 219 sujetos (2,19 %) (Tabla 5). Por ese motivo no debe sumarse a las cuatro causas como si fuese otra categoría de primer evento agudo. Las cifras son realizaciones de una semilla aleatoria fija; las probabilidades y ventanas temporales son supuestos exploratorios, no incidencias clínicas validadas.

Tabla 5. Eventos adversos y desenlace cosmético generados computacionalmente.

Evento o desenlace	n	Pacientes (%)	Día medio del evento
Seroma	69	0,69	94,78
Hematoma	48	0,48	94,39
Infección	2	0,02	101,56
Tromboembolia venosa	2	0,02	55,70
Irregularidad del contorno	219	2,19	106,97

La Fig. 3 muestra la distribución de los eventos simulados y permite identificar las diferencias en sus frecuencias relativas dentro de la cohorte virtual.

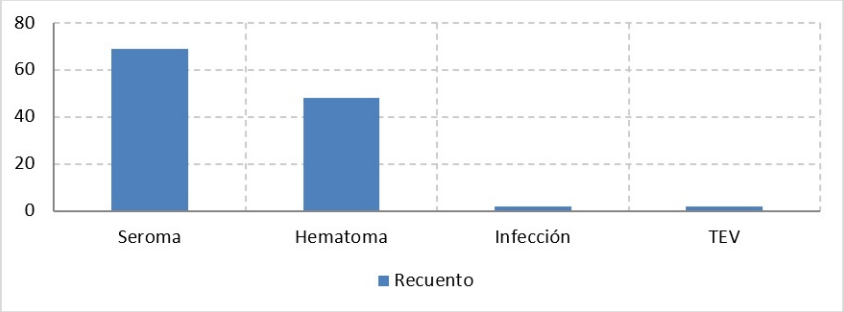


Fig. 3. Recuento de primeros eventos agudos simulados por causa, con el desenlace de contorno separado.

E. Exposición de seguimiento y controles de integridad

Se asignó una exposición abstracta de seguimiento de apoyo a 6.696 pacientes virtuales (66,96 %) mediante una regla dependiente de estados simulados anteriores. Su efecto causal sobre los estados y los eventos fue fijado en cero por construcción. En consecuencia, ninguna comparación entre pacientes expuestos y no expuestos puede interpretarse como una estimación de eficacia terapéutica. Los controles computacionales verificaron la unicidad de los 10.000 identificadores, ocho estados latentes por paciente, rango $(0, 1)$ de los índices y exclusividad del primer evento agudo. Las comprobaciones describen consistencia interna y no validación clínica externa.

F. Discusión

La simulación SIM-LIPO 3.2 permitió generar 80.000 evaluaciones longitudinales correspondientes a 10.000 individuos virtuales, con trayectorias diferenciadas de edema, dolor, función y cicatrización. La evolución de estos indicadores respondió a procesos estocásticos correlacionados y parámetros definidos matemáticamente. Aunque las trayectorias resultaron consistentes con las reglas del modelo, esta coherencia no constituye evidencia de recuperación clínica. Sargent [2] distingue precisamente entre la verificación del funcionamiento de un simulador y su validación frente al sistema real, aspecto que delimita el alcance de los resultados obtenidos.

El modelo produjo 121 primeros eventos adversos agudos (1,21 %) y 219 irregularidades del contorno (2,19 %), estas últimas registradas independientemente. La diferenciación responde a la teoría de riesgos competitivos desarrollada por Fine y Gray [4], que establece la necesidad de considerar la ocurrencia de causas alternativas al analizar la incidencia de eventos. Aunque las complicaciones representadas coinciden conceptualmente con las descritas por Comerci y colaboradores [1], sus frecuencias no son directamente comparables, debido a las diferencias entre las condiciones clínicas investigadas y los supuestos computacionales empleados. Por tanto, los resultados constituyen escenarios hipotéticos y no estimaciones epidemiológicas.

Respecto de la integridad longitudinal, se identificaron 6.999 observaciones faltantes (8,75 %), generadas mediante un mecanismo dependiente de características individuales y estados del proceso. Rubin [3] señaló que la naturaleza del mecanismo de ausencia condiciona la validez de las inferencias estadísticas. La conservación de los valores latentes completos representa una ventaja metodológica, pues permite evaluar procedimientos de imputación y estimación frente a referencias computacionales conocidas. Asimismo, la exposición de seguimiento asignada a 6.696 individuos presentó un efecto verdadero nulo, proporcionando un escenario para examinar asociaciones inducidas por selección, cuya interpretación exige las precauciones causales expuestas por Hernán y Robins [9].

La publicación del recurso en Zenodo [7] favorece su trazabilidad y reutilización, en concordancia con las recomendaciones de Sandve y colaboradores [5] y los principios FAIR formulados por Wilkinson y colaboradores [6]. Sin embargo, la principal limitación continúa siendo la ausencia de calibración y validación con registros clínicos reales. La evaluación futura deberá incorporar simulaciones repetidas, análisis de sensibilidad y escenarios alternativos para determinar la estabilidad de los resultados y las propiedades de los métodos estadísticos utilizados. En consecuencia, SIM-LIPO 3.2 constituye un entorno de experimentación computacional reproducible, cuyo valor principal reside en el desarrollo y evaluación de procedimientos matemáticos, sin proporcionar evidencia directa sobre pronóstico o efectividad clínica.

CONCLUSIONES

El desarrollo de SIM-LIPO 3.2 permitió establecer un modelo matemático-computacional reproducible para generar datos longitudinales simulados de recuperación posterior a procedimientos hipotéticos de liposucción aislada. La integración de procesos estocásticos, dependencia temporal, heterogeneidad individual y transformaciones matemáticas acotadas permitió construir trayectorias diferenciadas de edema, dolor, función y cicatrización. La cohorte de 10.000 individuos virtuales y sus 80.000 evaluaciones programadas proporcionó una estructura de datos consistente para estudiar el comportamiento de sistemas longitudinales bajo condiciones previamente definidas.

La incorporación de primeros eventos adversos competitivos, irregularidades del contorno como desenlace independiente y mecanismos selectivos de ausencia de observaciones amplió las posibilidades de experimentación estadística. La disponibilidad simultánea de estados latentes completos y registros observables constituye una característica especialmente relevante, dado que permite comparar estimaciones con valores conocidos y examinar el comportamiento de métodos de imputación, análisis longitudinal y modelado de eventos. Asimismo, la inclusión de una exposición de seguimiento con efecto verdadero nulo ofrece un escenario controlado para estudiar asociaciones inducidas por mecanismos de asignación.

La publicación del recurso en Zenodo, junto con su código, parámetros y documentación, favorece la transparencia, la trazabilidad y la reutilización científica. Sin embargo, las propiedades identificadas corresponden exclusivamente al sistema computacional implementado y no representan evidencia clínica sobre recuperación, incidencia de complicaciones o eficacia terapéutica. Se concluye que SIM-LIPO 3.2 constituye una infraestructura metodológica para investigaciones de simulación, cuya proyección científica dependerá de evaluaciones adicionales de sensibilidad, experimentación computacional repetida y futuras contrastaciones con información clínica independiente.

REFERENCIAS

[1] A. J. Comerci *et al.*, "Risks and complications rate in liposuction: A systematic review and meta-analysis," *Aesthetic Surgery Journal*, vol. 44, no. 7, pp. NP454–NP463, 2024. [Online]. Available: https://doi.org/10.1093/asj/sjae074

[2] R. G. Sargent, "Verification and validation of simulation models," *Journal of Simulation*, vol. 7, no. 1, pp. 12–24, 2013. [Online]. Available: https://doi.org/10.1057/jos.2012.20

[3] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976. [Online]. Available: https://doi.org/10.1093/biomet/63.3.581

[4] J. P. Fine and R. J. Gray, "A proportional hazards model for the subdistribution of a competing risk," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 496–509, 1999. [Online]. Available: https://doi.org/10.1080/01621459.1999.10474144

[5] G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, "Ten simple rules for reproducible computational research," *PLOS Computational Biology*, vol. 9, no. 10, p. e1003285, 2013. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1003285

[6] M. D. Wilkinson *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, p. 160018, 2016. [Online]. Available: https://doi.org/10.1038/sdata.2016.18

[7] F. Suárez-Carreño, "SIM-LIPO: A computationally simulated longitudinal dataset of recovery trajectories and competing first adverse events following isolated liposuction," Zenodo, 2026, version 3.2. [Online]. Available: https://doi.org/10.5281/zenodo.23460437

[8] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 3rd ed. Hoboken, NJ, USA: Wiley, 2019. [Online]. Available: https://doi.org/10.1002/9781119482260

[9] M. A. Hernán and J. M. Robins, *Causal Inference: What If*. Boca Raton, FL, USA: Chapman & Hall/CRC, 2025.