

Modelo de Riesgo de Importación de sarampión en Bogotá

Measles Importation Risk

DOCUMENTO METODOLÓGICO COMPLETO

Fundamentación matemática, decisiones estadísticas y trazabilidad
del modelo espacio-temporal de riesgo de importación y transmisión
local de sarampión en Bogotá D.C.

Secretaría Distrital de Salud

Dirección de Epidemiología – Grupo Análisis

Versión 1.0 — Marzo de 2026

Principios metodológicos y decisiones de diseño

1.0 Definición del modelo:

MIR-BOG es un modelo híbrido de tres etapas que combina técnicas estadísticas y epidemiológicas distintas en cada módulo: un modelo probabilístico de influjo (Poisson con simulación Monte Carlo) para estimar importaciones, un modelo de gravedad espacial paramétrico para distribuir el riesgo entre localidades, y un modelo compartimental SEIR estocástico (algoritmo de Gillespie) para simular la transmisión local post-importación. Se denomina híbrido porque no es un único marco matemático sino la integración en cascada de tres aproximaciones metodológicamente diferentes — estadística (M1), geoespacial (M2) y epidemiológica dinámica (M3) — que se conectan secuencialmente: la salida de cada módulo alimenta el siguiente, y el producto final es un índice integrado (Hazard Outbreak) que sintetiza las tres dimensiones del riesgo.

1.1 ¿Por qué un modelo mecanístico y no aprendizaje automático?

La elección de un marco mecanístico sobre técnicas de aprendizaje automático (ML) responde a tres restricciones fundamentales del problema:

Ausencia de variable de respuesta local. Bogotá no ha registrado casos confirmados de sarampión desde 2019. Un modelo supervisado (Random Forest, XGBoost, redes neuronales) requiere datos de outcome para entrenamiento. Sin brotes locales, no hay outcome que aprender. Los estudios que sí usan ML para sarampión (Ru et al., Vaccine 2024) operan sobre 3.143 condados de EEUU con docenas de importaciones anuales — un universo de datos incomparable.

Tamaño de muestra insuficiente. Con $n = 19$ localidades como unidades de análisis, cualquier modelo de ML con más de 2-3 predictores incurre en sobreajuste severo. La regla empírica de 10-20 observaciones por predictor (Harrell, 2001; van der Ploeg et al., 2014) limita a un máximo de 1-2 predictores — donde un modelo paramétrico es más eficiente y transparente.

Parámetros epidemiológicos bien establecidos. El sarampión es uno de los patógenos mejor caracterizados: $R_0 = 12-18$, periodo latente 10-14 días, periodo infeccioso 6-8 días, eficacia vacunal 93-97%. Estos parámetros permiten parametrización directa desde la biología, sin necesidad de estimar relaciones empíricas desde datos (Guerra et al., Lancet ID 2017; WHO Position Paper 2017).

1.2 Marco teórico: modelos de importación y transmisión de sarampión

MIR-BOG se fundamenta en cuatro líneas de investigación convergentes:

(a) Modelos de influjo basados en flujos de viajeros. Poterek et al. (Pathogens 2021) demostraron que una variable de influjo que combina volumen de viajeros aéreos con incidencia en el origen predice importaciones de sarampión a estados de EEUU con correlación de 0,84 y AUC de 0,78. BlueDot (2025) adoptó modelos de gravedad con tamaño poblacional y tiempo de viaje para predecir extensión de brotes.

(b) Modelos de gravedad para metapoblaciones de sarampión. Xia, Bjørnstad y Grenfell (Am Nat 2004) formalizaron la interacción entre poblaciones como $\phi(i,j) = \theta \times N_i^\alpha \times N_j^\beta / d(i,j)^\gamma$, marco que Bjørnstad et al. (Nat Ecol Evol 2020) extendieron con riesgos competitivos. Este es el fundamento del Módulo 2.

(c) Modelos SEIR estocásticos para eventos raros. En régimen de eliminación (cero casos endémicos), la estocasticidad domina sobre la dinámica determinística. El algoritmo de Gillespie (J Phys Chem 1977) es el estándar para simulación exacta de procesos de Markov en tiempo continuo. Masters et al. (MMWR 2024) usaron este enfoque exacto para modelar el brote de Chicago 2024 en tiempo real.

(d) Marco zero-inflated para países en eliminación. Bracher y Held (Stat Med 2023; PMID 37439440) desarrollaron el modelo ZI-HHH que separa la probabilidad de observar cero casos del proceso generador cuando hay importación. MIR-BOG adopta esta filosofía: no calibra transmisión con datos locales inexistentes, sino que parametriza desde la biología.

2. Módulo 1: Importación internacional

2.1 Formulación matemática

Sea $I_{total}(t)$ el número esperado de casos importados de sarampión a Bogotá en el mes t . Se modela como la suma de dos componentes:

$$I_{total}(t) = I_{directa}(t) + I_{indirecta}(t) \quad (1)$$

La componente directa captura viajeros internacionales que llegan al Aeropuerto El Dorado:

$$I_{directa}(t) = \sum_k [V(k \rightarrow BOG, t) \times \lambda(k, t) \times \rho(k) \times s(k)] \quad (2)$$

La componente indirecta captura viajeros que llegan a otras ciudades colombianas y posteriormente se desplazan a Bogotá:

$$I_{indirecta}(t) = \sum_c \sum_k [V(k \rightarrow c, t) \times \lambda(k, t) \times \rho(k) \times s(k) \times \omega(c \rightarrow BOG)] \quad (3)$$

Definición rigurosa de cada variable

$V(k \rightarrow j, t)$: Volumen mensual de pasajeros aéreos internacionales desde el país k hacia el aeropuerto j en el mes t . Unidades: pasajeros/mes. Fuente: Aeronáutica Civil de Colombia, archivos Origen-Destino (OD). Se procesaron 67 archivos OD cubriendo 2018-2025 (522.485 filas consolidadas). Validación gold standard con OD junio 2019: 6/6 métricas exactas.

$\lambda(k, t)$: Tasa de incidencia de sarampión en el país k durante el mes t , expresada como probabilidad de que un viajero aleatorio del país k esté infectado. Se calcula como:

$$\lambda(k, t) = \text{casos_confirmados}(k, t) / \text{población}(k) \quad (4)$$

Fuentes: OPS/OMS actualizaciones semanales, DGE México informes diarios (59 PDFs procesados), CDC MMWR, PHAC Canadá. 188 países, 26 meses (ene 2024 — feb 2026).

$\rho(k)$: Factor de ajuste por subreporte. Los sistemas de vigilancia no capturan el 100% de los casos. $\rho \geq 1$ escala los casos reportados hacia los casos reales estimados. Se modela como variable aleatoria con 51 escenarios distribuidos entre percentiles empíricos de la razón de subreporte reportada en la literatura (Patel et al., Lancet Global Health 2019). En la simulación Monte Carlo, ρ se muestrea uniformemente de estos escenarios.

$s(k)$: Proporción susceptible en la población viajera del país k . Se estima como:

$$s(k) = 1 - [\text{cov_MCV2}(k) \times \text{VE}_2] \quad (5)$$

Donde $\text{cov_MCV2}(k)$ es la cobertura de segunda dosis de vacuna contra sarampión reportada por el país k ante la OMS (WHO/UNICEF WUENIC estimates), y $\text{VE}_2 = 0,97$ es la eficacia vacunal de dos dosis. Para países con brotes activos y coberturas subóptimas (e.g., México con coberturas reportadas del 78% en casos confirmados), $s(k)$ captura la mayor proporción de viajeros susceptibles.

$\omega(c \rightarrow BOG)$: Probabilidad de desplazamiento posterior desde la ciudad colombiana c hacia Bogotá. Se estima como:

$$\omega(c \rightarrow BOG) = \text{flujo_doméstico}(c \rightarrow BOG, t) / \text{llegadas_internacionales}(c, t) \quad (6)$$

Donde el numerador es el volumen mensual de pasajeros en rutas aéreas domésticas $c \rightarrow BOG$ y el denominador es el total de pasajeros internacionales que llegan al aeropuerto de c . Ambos provienen de estadísticas públicas de la Aeronáutica Civil. Se calculó para 6 ciudades: Medellín (35,7%), Cali (49,6%), Barranquilla (64,8%), Cartagena (47,5%), Pereira (65,3%) y Bucaramanga (70,3%).

2.2 Modelo probabilístico

El número realizado de importaciones en el mes t sigue una distribución de Poisson:

$$X(t) \sim \text{Poisson}(I_{total}(t)) \quad (7)$$

La elección de Poisson se justifica por: (a) los eventos de importación son raros e independientes; (b) el número de "ensayos" (viajeros) es grande y la probabilidad de éxito (estar infectado) es pequeña, cumpliendo la aproximación Poisson del proceso binomial; y (c) es el modelo estándar en la literatura de importación de enfermedades (Poterek et al. 2021, CDC 2024).

Para generar intervalos de credibilidad, se ejecutan 10.000 realizaciones Monte Carlo muestreando $p(k)$ de su distribución empírica. Para cada realización r :

$$I_{total}^{(r)}(t) = \sum_k [V(k \rightarrow BOG, t) \times \lambda(k, t) \times p^{(r)}(k) \times s(k)] + \text{componente indirecta} \quad (8)$$

Los percentiles 2,5%, 50% y 97,5% de las 10.000 realizaciones definen la mediana y el intervalo de credibilidad al 95%.

2.3 Validación prospectiva

En febrero de 2026 se confirmaron 3 casos importados de sarampión en Colombia (2 en Bogotá, 1 en Santander), todos provenientes de México. El modelo, construido antes de estos eventos, estimó $I_{total} = 3,135$ casos acumulados en 26 meses. La distribución Poisson(3,135) tiene:

$$P(X = 3 / \lambda = 3,135) = e^{-(3,135)} \times 3,135^3 / 3! = 0,2234 \quad (9)$$

Esta es la probabilidad más alta de la distribución (la moda de Poisson(λ) es $[\lambda] = 3$). La probabilidad de observar exactamente 3 casos dado nuestro modelo es del 22,34% — el resultado más probable.

Nota sobre notación decimal: 3,135 es 3 COMA 135 (notación colombiana). En código y CSVs se usa punto decimal (3.135). Ambas representan el mismo valor: aproximadamente 3 casos esperados, no tres mil.

2.4 Hallazgo fundamental: λ (intensidad del brote del país de origen) domina sobre V (volumen de viajeros)

El análisis de contribución por componente reveló que la intensidad del brote en el país de origen (λ) domina sobre el volumen de viajeros (V) por órdenes de magnitud. La contribución de México al riesgo total pasó de 0,6% (pre-brote) a 82,5% (durante brote) exclusivamente por el aumento de λ , no de V . Cuantitativamente: durante un brote activo, λ se multiplica $\times 100-1.000$ (de $\sim 0,00001$ a $\sim 0,001-0,01$ casos por viajero); V se multiplica como máximo $\times 1,3$ por eventos como el Mundial. La implicación operativa es que el riesgo durante el Mundial FIFA 2026 depende de si México controla su brote, no del número adicional de turistas.

3. Módulo 2: Distribución espacial intraurbana

3.1 Modelo de gravedad

El I_{total} de M1 se distribuye entre las 19 localidades urbanas de Bogotá mediante un modelo de gravedad adaptado de Xia, Bjørnstad y Grenfell (2004). Sumapaz se excluye por ser enteramente rural (4.111 habitantes, 9 hab/km²). El peso de atracción de la localidad i es:

$$w_i = N_i \times S_i / d_i^\gamma \quad (10)$$

Y la probabilidad de que un caso importado llegue a la localidad i es:

$$P_{import}(i) = w_i / \sum_{j=1}^{19} w_j \quad (11)$$

N_i : Población total de la localidad i . Fuente: proyecciones DANE post-censo 2018. Rango: 19.207 (La Candelaria) a 1.351.651 (Suba).

S_i : Fracción susceptible de la localidad i (ver sección 3.2 para construcción detallada).

d_i : Distancia geodésica (Haversine) desde el Aeropuerto El Dorado al centroide geográfico de la localidad i , en kilómetros. Rango: 2,70 km (Fontibón) a 34,46 km (Usme).

γ : Exponente de decaimiento por distancia. Valor central $\gamma = 1,0$ (decaimiento inversamente proporcional a la distancia). Análisis de sensibilidad con $\gamma \in \{0,5; 1,0; 1,5; 2,0\}$. La elección $\gamma = 1,0$ es consistente con Xia et al. (2004), quienes estimaron $\gamma \approx 0,8-1,2$ para epidemias de sarampión en Inglaterra y Gales.

3.2 Estimación de susceptibilidad: la variable más crítica del modelo

La fracción susceptible S_i es el eslabón más débil de la cadena $M1 \rightarrow M2 \rightarrow M3$ porque determina directamente el R local y, por tanto, la probabilidad de brote. Su estimación requirió resolver un problema de fuentes contradictorias.

3.2.1 Fuentes evaluadas y descartadas

BD vacunados (PAI 2.0): Base de datos nominal del programa de vacunación de la SDS. 3.753.823 registros, 1.382.881 personas únicas (deduplicadas por tipo de documento e identificación). Captura solo el 35,6% de la población DANE en cohorte 0-4 años, probablemente porque ~70% de la vacunación en Bogotá ocurre en el sector privado y la BD cubre principalmente el sector público. No se confirmó el alcance exacto (preguntas a Andrea pendientes de respuesta). Descartada como estimación central; útil solo como piso de cobertura.

Coberturas administrativas (SaluData / SIS-151): Dato oficial del sistema de salud. Cobertura = dosis aplicadas / población meta (proyección DANE). Problema fundamental: 8/19 localidades con cobertura > 100% (Bosa 175,9%, Barrios Unidos 170,7%), lo que indica errores en el denominador poblacional. Teusaquillo reporta 38,6%, un valor implausiblemente bajo para una localidad de estrato alto con acceso a vacunación privada. La correlación entre BD y coberturas administrativas es negativa: Spearman $\rho = -0,333$ ($p = 0,18$). Las dos fuentes ni siquiera coinciden en cuáles localidades están peor cubiertas. Descartadas como fuente principal.

3.2.2 Fuente adoptada: Encuesta de Cobertura Vacunal (ECV) 2025

Descripción: Encuesta domiciliaria de campo con verificación de carné de vacunación, realizada entre octubre 2025 y enero 2026. 4.318 niños encuestados en 19 localidades urbanas + Sumapaz. El encuestador visitó hogares, verificó el carné físico del niño, y registró la fecha de aplicación de cada biológico. Es el gold standard para coberturas reales según la OMS (WHO Vaccination Coverage Cluster Surveys Reference Manual, 2018).

Variable utilizada: Columna TRIPLEVIRAL_1_fecha del archivo fuente. Un niño se clasifica como vacunado con TV1 si esta columna tiene fecha (no es NaN). Se trabajó con la cohorte de 12-23 meses, que es el indicador PAI estándar para SRP1.

Resultado: Cobertura SRP1 Bogotá total = 96,3%. Rango por localidad: 84,6% (Puente Aranda, $n = 13$) a 100% (6 localidades). Cero localidades con cobertura > 100% (resuelve el problema de denominador de las coberturas administrativas).

3.2.3 Decisión: TV1 sola, no TV1 + refuerzo

La ECV también contiene la columna TRIPLEVIRAL_2_fecha (refuerzo, aplicado a partir de los 18 meses). Se decidió usar solo TV1 con eficacia de 1 dosis ($VE_1 = 0,93$) en vez de incorporar TV2 con eficacia de 2 dosis ($VE_2 = 0,97$). Razones:

(a) La cohorte primaria de la ECV (12-23 meses) incluye niños que aún no han alcanzado la edad del refuerzo (18 meses). Usar TV2 excluiría niños elegibles pero aún no vacunados con refuerzo, subestimando cobertura.

(b) Usar $VE_1 = 0,93$ es conservador: sobreestima ligeramente la susceptibilidad respecto a usar $VE_2 = 0,97$. Para Kennedy, la diferencia es $Re = 1,63$ con VE_1 vs $Re \approx 1,40$ con VE_2 . La narrativa ($Re > 1$, brote posible) no cambia; la magnitud sí.

(c) Es más honesto no mezclar cohortes con diferentes criterios de elegibilidad para el refuerzo.

Implicación cuantitativa: con $VE_2 = 0,97$, S_i bajaría ~18% y Re bajaría proporcionalmente. Kennedy pasaría de $Re = 1,63$ a $Re \approx 1,40$. El ranking de Hazard no cambiaría porque el efecto es uniforme.

3.2.4 Estimador shrinkage empírico de Bayes para muestras pequeñas

Varias localidades tienen muestras pequeñas en la ECV: Teusaquillo ($n = 3$), Antonio Nariño ($n = 4$), Sumapaz ($n = 4$), Chapinero ($n = 7$). La estimación puntual $p_{obs} = \text{vacunados/encuestados}$ tiene alta varianza con n pequeño. Se aplica un estimador de shrinkage empírico de Bayes (Efron & Morris, JASA 1975; Gelman & Hill, 2007):

$$\hat{p}_{shrink}(i) = w_i \times p_{obs}(i) + (1 - w_i) \times \bar{p}_{global} \quad (12)$$

$$w_i = n_i / (n_i + n_0) \quad (13)$$

Donde $p_{obs}(i)$ es la cobertura observada en la localidad i , $\bar{p}_{global} = 0,963$ es la cobertura agregada de Bogotá, n_i es el tamaño de muestra, y $n_0 = 40$ (mediana de los n_i entre localidades) es el hiperparámetro de shrinkage.

Interpretación: $w_i \in [0, 1]$ pondera entre el dato local y la media global. Localidades con muestra grande ($n_i \gg n_0$) tienen $w_i \rightarrow 1$ y conservan su estimación local. Localidades con muestra pequeña ($n_i \ll n_0$) tienen $w_i \rightarrow 0$ y se acercan a la media de Bogotá. Ejemplo: Teusaquillo ($n = 3$, $p_{obs} = 1,00$) $\rightarrow w = 3/(3+40) = 0,07 \rightarrow \hat{p}_{shrink} = 0,07 \times 1,00 + 0,93 \times 0,963 = 0,965$. El valor extremo de 100% se modera a 96,5%.

El estimador de shrinkage es equivalente a un estimador bayesiano con prior $Beta(\alpha_0, \beta_0)$ centrado en $\bar{p}_{global}$. La elección de $n_0 = mediana(n_i)$ es una heurística robusta que evita la especificación subjetiva de hiperparámetros (Carlin & Louis, *Bayesian Methods for Data Analysis*, 2009).

3.2.5 Ponderación por cohortes etarias

La ECV cubre exclusivamente la cohorte 0-5 años. Sin embargo, la susceptibilidad total de una localidad depende de todas las cohortes etarias. Se pondera:

$$S_{total}(i) = \sum_g [\pi_g \times (1 - cob_g(i) \times VE_i)] \quad (14)$$

Donde g indexa los grupos etarios, π_g es la proporción de la población en el grupo g, y $cob_g(i)$ es la cobertura vacunal estimada para el grupo g en la localidad i.

Cohorte (g)	π_g	Fuente de cob_g	Supuesto / Justificación
0-5 años	0,08	ECV 2025 (shrinkage)	Dato observado en campo, gold standard
6-14 años	0,13	0,92 (supuesto)	Coberturas pre-pandemia 2012-2019 estables al 90-95% (SaluData serie histórica)
15-39 años	0,38	0,90 (supuesto)	Generación post-vacunación masiva en Colombia; coberturas históricas nacionales ~90%
40+ años	0,41	0,97 (supuesto)	Generaciones expuestas al virus silvestre pre-eliminación; inmunidad natural alta

Transparencia sobre la composición: solo el 8% de S_{total} proviene de datos medidos (ECV). El 92% restante proviene de supuestos informados por la literatura y coberturas históricas. Un cambio de ± 5 puntos porcentuales en la cobertura asumida de la cohorte 15-39 años (la más grande, $\pi = 0,38$) impacta S_{total} en $0,38 \times 0,05 \times 0,93 = 0,018$, lo cual modifica Re en $\pm 0,27$. Este impacto es mayor que el de la variabilidad de la ECV en la cohorte 0-5 ($0,08 \times 0,10 \times 0,93 = 0,007$). La susceptibilidad total depende más de los supuestos de las cohortes mayores que de la encuesta.

Limitación documentada: la estimación ideal requeriría una ECV ampliada que cubra todas las cohortes. Sin este dato, la heterogeneidad entre localidades en cohortes >5 años no se captura. Los supuestos planos (92%, 90%, 97%) se aplican uniformemente a todas las localidades, lo que reduce las diferencias de S_i entre localidades a las que aporta la ECV en la cohorte 0-5 y las diferencias demográficas (π_g varía ligeramente entre localidades).

3.2.6 Tres escenarios monotónicos

Se construyen tres escenarios de S_i garantizando monotonía estricta (sin inversiones de bounds):

OPTIMISTA: ECV shrinkage + cohortes mayores al 95%. Representa la estimación más favorable.

CENTRAL: ECV shrinkage + cohortes históricas (92%, 90%, 97%). Estimación principal.

PESIMISTA: Intervalo de confianza inferior (Clopper-Pearson 95%) de la ECV + descuento de 10 puntos porcentuales para la cohorte pandémica (nacidos 2020-2022).

Verificación: $S_{optimista} < S_{central} < S_{pesimista}$ en las 19 localidades, sin excepciones. Esto contrasta con los escenarios TECHO/CENTRAL/PISO del M2 original, que presentaban 17 inversiones de bounds.

4. Módulo 3: Transmisión local (SEIR estocástico)

4.1 Modelo compartimental SEIR

Cada localidad se modela como una población cerrada (sin acoplamiento entre localidades en la versión actual) con cuatro compartimentos:

$$S(t) + E(t) + I(t) + R(t) = N_i \quad \forall t \quad (15)$$

Las transiciones entre compartimentos son procesos estocásticos en tiempo continuo con tasas:

Transición	Tasa	Interpretación
$S \rightarrow E$	$\beta_i \times S(t) \times I(t) / N_i$	Infección: contacto entre susceptible e infeccioso
$E \rightarrow I$	$\sigma \times E(t)$	Progresión: fin del periodo latente
$I \rightarrow R$	$\gamma \times I(t)$	Recuperación: fin del periodo infeccioso

4.1.1 Tasa de transmisión local β_i

La tasa de transmisión se calibra desde el número reproductivo básico R_0 y se ajusta localmente por densidad:

$$\beta_i = R_0 \times f(d_i) \times \gamma \quad (16)$$

Donde γ es la tasa de recuperación ($1/7$ días⁻¹). Esta relación se deriva de la definición de R_0 en un modelo SIR/SEIR: $R_0 = \beta / \gamma$ en una población completamente susceptible (Anderson & May, Infectious Diseases of Humans, 1991).

4.1.2 Función de contacto $f(d_i)$

La función $f(d_i)$ modula la tasa de contacto por la densidad poblacional de la localidad:

$$f(d_i) = \log(d_i / d_{\min}) / \log(d_{\max} / d_{\min}) \in [0, 1] \quad (17)$$

Con $d_{\min} = 1.999$ hab/km² (Usme) y $d_{\max} = 30.984$ hab/km² (Bosa). La función logarítmica captura la saturación de contactos en entornos urbanos: duplicar la densidad no duplica los contactos. Esta relación es consistente con hallazgos empíricos de Hu et al. (BMC Infect Dis 2013) y con el marco teórico de contactos frequency-dependent vs density-dependent (McCallum et al., Trends Ecol Evol 2001).

Se evaluaron tres opciones: (A) lineal $f(d) = d/d_{\max}$ — sobreestima contactos en alta densidad; (B) logarítmica (adoptada) — captura saturación con mínimos supuestos; (C) matrices PREM de contacto (Prem et al. 2021) — más rigurosa pero introduce complejidad no justificada para $n = 19$ localidades sin datos de contacto locales. Las matrices PREM quedan como validación para versiones futuras.

4.1.3 Número reproductivo efectivo local

$$Re(i) = R_0 \times f(d_i) \times S_i \quad (18)$$

$Re(i)$ es el número esperado de casos secundarios generados por un caso infeccioso en la localidad i , dado el nivel actual de susceptibilidad. Si $Re > 1$, un brote sostenido es posible. Si $Re < 1$, la transmisión se extingue espontáneamente.

Con los parámetros actuales ($R_0 = 15$, ECV central): Re varía de 0,00 (Usme, $f = 0$) a 1,69 (Bosa). 12/19 localidades tienen $Re > 1$. Ninguna tiene $Re > 2$.

4.2 Parámetros epidemiológicos

Parámetro	Símbolo	Valor	Fuente
R_0	R_0	15 (sens: 12, 18)	Consenso OMS; Guerra et al., Lancet ID 2017
Periodo latente	$1/\sigma$	11,5 días	Lessler et al., Lancet ID 2009
Periodo infeccioso	$1/\gamma$	7 días	Perry & Halsey, JID 2004
VE 1 dosis SRP	VE_1	0,93	WHO Position Paper 2017
VE 2 dosis SRP	VE_2	0,97	WHO Position Paper 2017 (no usada, ver 3.2.3)
Retardo detección	τ	5 días (sens: 3, 7)	INS Protocolo SR v06, 2024
Reducción β intervención	—	70%	Cerco + bloqueo vacunal (supuesto)
Intervalo serial	—	~11-12 días	Fine, Rev Infect Dis 1982

4.3 Algoritmo de simulación estocástica exacta (Gillespie SSA)

El algoritmo de Gillespie (J Phys Chem 1977) simula la evolución exacta del proceso de Markov en tiempo continuo definido por las tasas de transición. En cada paso:

Paso 1. Calcular las tasas de los tres eventos posibles: $r_1 = \beta_i \times S \times I / N$ (infección), $r_2 = \sigma \times E$ (progresión), $r_3 = \gamma \times I$ (recuperación).

Paso 2. Calcular la tasa total: $a_0 = r_1 + r_2 + r_3$.

Paso 3. Muestrear el tiempo hasta el siguiente evento: $\Delta t \sim \text{Exponencial}(a_0)$.

Paso 4. Seleccionar el tipo de evento con probabilidad proporcional a su tasa: $P(\text{infección}) = r_1/a_0$, $P(\text{progresión}) = r_2/a_0$, $P(\text{recuperación}) = r_3/a_0$.

Paso 5. Actualizar los compartimentos: $S-1/E+1$ (infección), $E-1/I+1$ (progresión), o $I-1/R+1$ (recuperación).

Paso 6. Repetir hasta que $E = 0$ y $I = 0$ (extinción) o $t > 365$ días.

Optimización híbrida: cuando $I > 50$, las fluctuaciones estocásticas son despreciables respecto a la dinámica determinística. El simulador cambia a resolución ODE (Runge-Kutta 4) con umbral $I = 50$, siguiendo la práctica de Keeling & Rohani (Modeling Infectious Diseases, Princeton UP 2008). Rendimiento: 500 simulaciones de Kennedy ($N = 1.045.353$) en 0,86 segundos.

4.4 Modelado de la intervención (cerco epidemiológico)

La intervención se modela como una reducción instantánea de la tasa de transmisión tras la detección:

$$\beta_{\text{post}} = \beta_i \times (1 - \varepsilon) \quad \text{para } t > \tau \quad (19)$$

Donde $\tau = 5$ días es el retardo de detección (tiempo desde inicio de síntomas hasta notificación SIVIGILA) y $\varepsilon = 0,70$ es la eficacia combinada del cerco epidemiológico + vacunación de bloqueo.

Con esta parametrización: $\text{Re}_{\text{post}} = \text{Re}(i) \times 0,30$. Para Bosa ($\text{Re}_{\text{máximo}} = 1,69$): $\text{Re}_{\text{post}} = 0,51 < 1$. Por tanto, la intervención es suficiente para contener el brote en todas las 19 localidades de Bogotá, siempre que la detección ocurra en ≤ 5 días.

La función escalón es una simplificación. En la práctica, la respuesta es gradual: identificación de contactos (24-48h), investigación de campo (48-72h), vacunación de bloqueo (72-120h). La reducción del 70% es un supuesto conservador (la respuesta completa podría reducir β en 80-90%). Se documenta como limitación.

4.5 Probabilidad de extinción: teoría vs simulación

Para un proceso de ramificación homogéneo (modelo SIR sin periodo latente), la probabilidad de extinción desde un solo caso importado es $P(\text{ext}) = 1/\text{Re}$ cuando $\text{Re} > 1$ (Bartlett, Stochastic Population Models, 1960). Para el modelo SEIR con periodo latente, $P(\text{ext})$ es menor que $1/\text{Re}$ porque el periodo latente amplía la ventana de transmisión antes de la recuperación. La relación exacta depende de la distribución del periodo latente y no tiene solución cerrada general (Nishiura, Math Biosci 2010).

En las simulaciones: $P(\text{ext}) \times \text{Re} \in [0,51 - 0,63]$ para localidades con $\text{Re} > 1$, consistente con la teoría para SEIR (esperado entre 0,45 y 0,65 según Jagers, Branching Processes with Biological Applications, 1975).

5. Índice Hazard Outbreak: integración M1 + M2 + M3

El índice final combina la probabilidad de que un caso importado llegue a la localidad i (producto de M2) con la probabilidad de que ese caso genere un brote significativo (producto de M3):

$$Hazard(i) = P_{import}(i) \times P(brote > 10 \mid caso \text{ importado en } i) \quad (20)$$

P_{import} proviene de la ecuación (11) del Módulo 2. $P(brote > 10)$ se estima como la proporción de las 10.000 realizaciones de Gillespie que producen un brote de más de 10 casos (excluyendo el caso índice). El umbral de 10 casos se eligió porque: (a) supera la capacidad de contención de un cerco epidemiológico tardío; (b) activa el protocolo de respuesta del INS para brote confirmado.

El Hazard se normaliza: $Hazard_norm(i) = Hazard(i) / \max(Hazard)$, resultando en un índice $[0, 1]$. Las localidades se clasifican por cuartiles: MUY ALTO ($> P75$), ALTO ($P50-P75$), MODERADO ($P25-P50$), BAJO ($< P25$).

5.1 Ranking final

Rank	Localidad	P_import	Re	P(>10)	Hazard	Nivel
1	Kennedy	0,1504	1,63	40,5%	1,00	MUY ALTO
2	Engativá	0,1306	1,51	36,5%	0,78	MUY ALTO
3	Bosa	0,0811	1,69	41,7%	0,55	MUY ALTO
4	Suba	0,1295	1,19	24,3%	0,52	MUY ALTO
5	Rafael Uribe U.	0,0624	1,63	41,0%	0,42	MUY ALTO
6	Fontibón	0,0766	1,12	22,1%	0,28	ALTO
7	Puente Aranda	0,0473	1,28	28,0%	0,22	ALTO
8	Tunjuelito	0,0368	1,40	33,2%	0,20	ALTO
9	Usaquén	0,0506	0,96	14,8%	0,12	ALTO
10	Barrios Unidos	0,0223	1,21	25,3%	0,09	ALTO

6. Verificaciones de consistencia interna

Se ejecutaron 7 pruebas de consistencia sobre el modelo completo:

Prueba	Criterio	Resultado
$\sum P_{\text{import}} = 1$	Suma de probabilidades = 1,000	PASS (1,000000)
Monotonía de escenarios	$S_{\text{opt}} < S_{\text{central}} < S_{\text{pes}} \forall i$	PASS (0 inversiones)
Re calculado exacto	$Re = R_0 \times f(d_i) \times S_i \forall i$	PASS (diff máx = 0,0000)
P(ext) consistente con teoría	$P(\text{ext}) \times Re \in [0,45 - 0,65]$ para SEIR	PASS (rango [0,51 - 0,63])
$Re < 1 \rightarrow P(>10) \approx 0$	Localidades subcríticas sin brotes grandes	PASS (máx P(>10) con $Re < 1 = 14,8\%$)
Intervención $\rightarrow P(>50) = 0$	Cerco epidemiológico contiene brotes	PASS (0% en 19/19)
Ranking invariante a R_0	Spearman ρ entre rankings $R_0=12$ vs 18	PASS ($\rho = 1,0000$)

7. Limitaciones

(a) Sin acoplamiento entre localidades. El modelo simula cada localidad independientemente. No captura propagación entre localidades vía TransMilenio/SITP. La Fase 3B incorporará una matriz de movilidad OD (pendiente de SDM).

(b) Susceptibilidad de cohortes mayores es supuesta, no medida. El 92% de S_{total} proviene de supuestos planos para cohortes >5 años. La ECV solo midió cohorte 0-5 (8% de la población). Los supuestos se aplican uniformemente, reduciendo la heterogeneidad entre localidades.

(c) Muestras pequeñas en la ECV. Teusaquillo ($n = 3$), Antonio Nariño ($n = 4$), Chapinero ($n = 7$) tienen estimaciones de baja precisión, mitigadas con shrinkage pero no resueltas. Los IC 95% de estas localidades son muy amplios (e.g., Teusaquillo: [43,8% — 100%]).

(d) Solo TV1, no TV1 + refuerzo. Se usó eficacia de 1 dosis (0,93) en vez de 2 dosis (0,97). Sobreestima susceptibilidad ~18%. El impacto en Re es ~0,23 puntos. No cambia el ranking.

(e) Perfil de viajero no diferenciado. El modelo asume distribución proporcional a población. Turistas extranjeros se concentran en Chapinero/Usaquén/La Candelaria; colombianos retornando en estratos medios-altos. Esto probablemente sobreestima riesgo en localidades populares (Bosa, Rafael Uribe Uribe) y subestima en localidades turísticas.

(f) Distancia geodésica, no tiempo de viaje en red vial. d_i es distancia Haversine, no tiempo de viaje real. La red vial y el TransMilenio hacen que Fontibón (2,7 km) sea mucho más accesible que Usme (34,5 km) — pero la relación no es lineal con la distancia geodésica.

(g) Intervención como función escalón. La reducción de β es instantánea al día τ . En realidad, el cerco epidemiológico es gradual (48-120h). La reducción del 70% es un supuesto no calibrado.

(h) No se implementó inferencia bayesiana ni Random Forest. La propuesta original los contemplaba. Se descartaron por insuficiencia de datos (ver sección 1.1). La cuantificación de incertidumbre se logra vía Monte Carlo (M1) y simulación estocástica (M3).

8. Referencias bibliográficas

- Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, 1991.
- Bartlett MS. Stochastic Population Models in Ecology and Epidemiology. Methuen, 1960.
- Bjørnstad ON et al. A competing-risks model explains hierarchical spatial coupling of measles epidemics en route to national elimination. *Nat Ecol Evol.* 2020;4:934-939.
- Bracher J, Held L. A zero-inflated endemic-epidemic model with application to measles in Germany. *Stat Med.* 2023;42(21):3750-3768. PMID: 37439440.
- Carlin BP, Louis TA. Bayesian Methods for Data Analysis. 3rd ed. Chapman & Hall/CRC, 2009.
- CDC Center for Forecasting and Outbreak Analytics. Behind the Model: Interactive Measles Outbreak Simulator. Agosto 2025.
- Efron B, Morris CN. Data analysis using Stein's estimator and its generalizations. *JASA.* 1975;70(350):311-319.
- Fine PEM. The interval between successive cases of an infectious disease. *Am J Epidemiol.* 1982;116(2):356.
- Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge UP, 2007.
- Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem.* 1977;81(25):2340-2361.
- Guerra FM et al. The basic reproduction number (R_0) of measles: a systematic review. *Lancet Infect Dis.* 2017;17(12):e420-e428.
- Harrell FE. Regression Modeling Strategies. Springer, 2001.
- Hu H et al. Scaling laws of the urban-rural contact structure. *BMC Infect Dis.* 2013;13:405.
- Jagers P. Branching Processes with Biological Applications. Wiley, 1975.
- Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton UP, 2008.
- Lessler J et al. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis.* 2009;9(5):291-300.
- Masters NB et al. Real-time use of a dynamic model — Chicago, 2024. *MMWR.* 2024;73(19):430-434.
- McCallum H et al. How should pathogen transmission be modelled? *Trends Ecol Evol.* 2001;16(6):295-300.
- Nishiura H. Correcting the actual reproduction number. *J Appl Math Comput.* 2010;34:17-30.
- OPS/OMS. Alerta Epidemiológica: Sarampión en la Región de las Américas. 3 de febrero de 2026.
- Patel MK et al. Progress toward regional measles elimination — worldwide, 2000–2019. *Lancet Global Health.* 2019.
- Perry RT, Halsey NA. The clinical significance of measles: a review. *J Infect Dis.* 2004;189(Suppl 1):S4-S16.
- Poterek ML et al. Air passenger travel and international surveillance data predict spatiotemporal variation in measles importations to the United States. *Pathogens.* 2021;10(2):155.
- Prem K et al. Projecting contact matrices in 177 geographical regions. *PLOS Comp Biol.* 2021;17(7):e1009098.
- Ru Y et al. Prediction of measles cases in US counties: a machine learning approach. *Vaccine.* 2024;42(25):126096.
- SDS Bogotá. Encuesta de Cobertura Vacunal 2025 — Matriz programada revisión ECV.
- van der Ploeg T et al. Modern modelling techniques are data hungry. *BMC Med Res Methodol.* 2014;14:137.
- WHO. Measles vaccines: WHO position paper — April 2017. *Wkly Epidemiol Rec.* 2017;92(17):205-228.
- WHO. Vaccination Coverage Cluster Surveys: Reference Manual. WHO/IVB/18.09, 2018.
- Xia Y, Bjørnstad ON, Grenfell BT. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am Nat.* 2004;164(2):267-281.