

# Índice de alarma epidemiológica: un instrumento para valorar la evolución diaria de un proceso epidémico

Luis Carlos Silva Aycaguer, investigador titular

*Escuela Nacional de Salud Pública de Cuba*

## RESUMEN

*Contar con instrumentos que permitan alertar tempranamente sobre posibles agravamientos de la emergencia sanitaria asociada a una epidemia es una demanda natural para los sistemas de vigilancia. En el presente artículo se propone una herramienta concebida con ese fin, denominada “Índice de Alarma Epidemiológica” (IAE). El IAE se ha concebido como un indicador compuesto, que permita estimar mediante un número la alarma correspondiente a cualquier día epidémico que se elija, siempre que se cuente con la información sobre lo ocurrido en cierta cantidad de días epidémicos previos.*

*Los elementos más atractivos y novedosos que distinguen al IAE respecto de esfuerzos metodológicos anteriores para detectar situaciones epidemiológicamente alarmantes son los siguientes:*

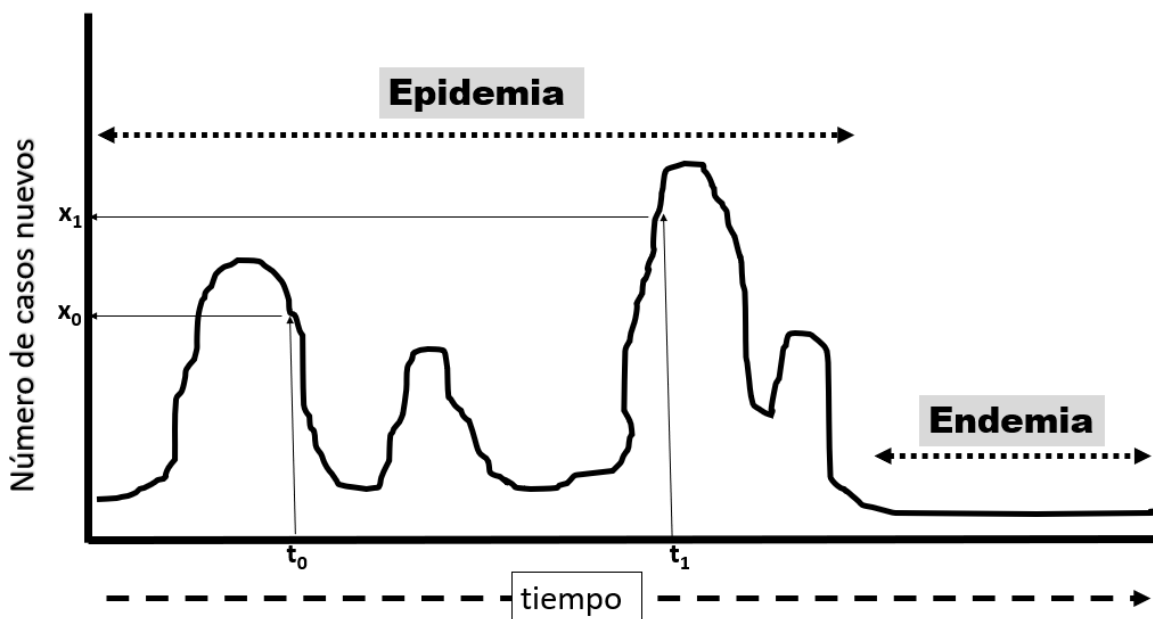
- *No se basa en una única variable, sino que permite conjugar la influencia simultánea de varias variables indicadoras de daño a la salud colectiva (variables centinela)*
- *El peso que se otorga a los valores que se registran para cada variable centinela es relativo: será mayor cuanto más se aproxime al máximo histórico (pico de la epidemia) alcanzado por dicha variable*
- *Sus valores se modifican en dependencia del signo que tenga la tendencia que exhiban los datos a lo largo de los días previos, así como de la intensidad de dicha asociación*
- *El resultado de su aplicación es un número que se halla entre 0% y 100%. Alcanza la cota inferior (0%) en el caso de que no haya motivo alguno para la alarma lo cual a su vez se produce si durante los días previos al momento que se analiza el comportamiento de las variables centinela haya sido sostenidamente nulo; y alcanza el máximo valor (100%) cuando estamos ante el peor escenario posible: para cada uno de esos días y para todas las variables se ha registrado una cifra que constituye un record en el marco de la epidemia.*

*Se da cuenta de las premisas metodológicas en que se basa y de los pasos operativos que exige su cómputo. Su fundamentación matemática, que tienen cierta complejidad algebraica, se explica detalladamente en el texto. Tanto el procedimiento como su interpretación se ilustran usando datos reales del comportamiento diario de la COVID-19 en Cuba a lo largo del período que va del 13 de marzo de 2020 al 16 de enero de 2022. Adicionalmente, se ofrece la posibilidad de aplicar el método a datos elegidos por el lector usando un programa informático elaborado en lenguaje R, que puede ser obtenido gratuitamente en Internet.*

## 1. La estimación de la alarma en un contexto epidémico

Las epidemias asociadas a dolencias infecto contagiosas suelen caracterizarse por la aparición de sucesivas “olas” de casos nuevos que marcan incrementos del daño a la salud en el seno de la sociedad que las padece. Típicamente, alcanzada una “cresta”, los valores descienden y permanecen más o menos estables hasta que comparece una nueva “ola”. En determinado momento, el fenómeno epidémico puede desaparecer por completo, aunque es mucho más frecuente que, tras un lapso de inestabilidades, pase a adquirir la categoría de “endemia”, como se muestra en el **Gráfico N°1**.

**Gráfico N°1** Comportamiento típico de las epidemias asociadas a dolencias infecciosas a lo largo del tiempo.



Pero mientras la epidemia no desaparezca ni la enfermedad se torne endémica, la sociedad toda y en particular los sistemas de salud deben monitorizar su comportamiento y, sobre todo, emprender acciones orientadas a controlar y mitigar sus efectos sobre la salud colectiva. Pueden desarrollarse acciones para, por ejemplo, evitar exigencias operativas que desborden los servicios (“aplanar” las curvas) o, al menos, para prepararlos con vistas a enfrentar las demandas que se avizoren.

En general los indicadores son recursos útiles para la planificación y la gestión salubristas a mediano y largo plazo; aportan información con la cual optimizar el diseño de planes y programas, o corregirlos si fuera necesario.<sup>1</sup> Pero las epidemias suelen exigir reacciones mucho más ágiles, de modo que los indicadores deben a su vez responder a la dinámica propia de esos procesos.

La necesidad de desarrollar mecanismos efectivos de vigilancia epidemiológica ha promovido el empleo de herramientas capaces de anticipar tempranamente la aparición de los “picos epidémicos” o la consolidación de evoluciones más agresivas de la epidemia.<sup>2</sup> Así ha ocurrido, por ejemplo, con el dengue,<sup>3</sup> la influenza<sup>4</sup> y la malaria.<sup>5</sup>

El recurso estadístico más antiguo para identificar la irrupción de una situación alarmante data de 1932, cuando el gran estadístico y salubrista de la Universidad de Wisconsin, Selwyn De Witt Collins propuso y empleó los llamados “canales endémicos” para detectar precozmente cifras anómalamente altas de influenza.<sup>6</sup> Un canal endémico es un recurso gráfico que representa el número de casos dentro de un rango estacional normal; de modo que valores por encima de los umbrales móviles representados en ese corredor son considerados indicios de un brote epidémico.<sup>7</sup> Una notable limitación de los canales endémicos reside en que reposan solamente en los nuevos casos que se van produciendo y no en otros efectos de las situaciones epidémicas, algunas tanto o más preocupantes, tales como hospitalizaciones o muertes. Una alternativa novedosa como instrumento estadístico para identificar señales de alarma, surgida al calor de la pandemia de

COVID-19, es el llamado “Índice de volatilidad”<sup>8</sup>. Sin embargo, este indicador padece de la misma insuficiencia señalada para los canales endémicos: los casos nuevos que van apareciendo constituyen la única manifestación que puede gravitar en el nivel de alarma epidémica.

La tupida madeja de elementos que confluyen para establecer la magnitud de una amenaza epidemiológica que permita ordenar y sistematizar información útil para la toma de decisiones sugiere, sin embargo, la elaboración y posterior empleo de “indicadores compuestos”. Con ellos se procura traducir conceptos y nociones teóricas -en nuestro caso, la noción de alarma- en unidades de medida que cuantifiquen, de manera matemáticamente precisa, la realidad objeto de estudio.<sup>9</sup>

Entre las variables que usualmente se registran con mayor celo para aquilatar la situación prevaleciente en un momento dado dentro de un proceso epidémico se hallan el número de casos nuevos, la cantidad de fallecidos, así como los hospitalizados en planta y en cuidados intensivos. Un crecimiento de cualquiera de ellas transmite motivos para alarmarse. Otro indicador, relevante por apuntar a una fuente esencial para la aparición de futuros contagios, es el número prevaleciente cada día de “casos activos”: sujetos que, habiendo sido confirmados con la dolencia, aún no se han recuperado ni fallecido.

Todos estos registros son potencialmente útiles para establecer cuán alarmante es la situación en un momento dado; sin embargo, por lo general se emplean de manera desconectada entre sí. Por otra parte, ni se articulan con un análisis de tendencias, ni sus valores se enjuician teniendo en cuenta el comportamiento de dichas variables a lo largo de la propia epidemia.

## **2. La adopción de medidas por las autoridades de salud**

La importancia de poder realizar valoraciones integradas y abarcadoras reside en que, a la postre, lo que de veras interesa a la salud pública es contar con instrumentos que ayuden a optimizar sus decisiones.

Entre los elementos en que se sustentan las decisiones sobre las medidas a tomar en materia de salud en general y en el marco de una epidemia en particular, necesariamente han de contemplarse las circunstancias políticas que rodean el fenómeno epidémico. Y estas abarcan desde el grado de rechazo social que pudieran despertar las medidas que se podrían instaurar, hasta la factibilidad económica y logística de su aplicación. Además no se deben desdeñar las restricciones éticas y legales que pudieran obstaculizar, y hasta impedir su implementación.

Pero un indicador -al decir del epistemólogo Juan Samaja- puede entenderse como “un eslabón entre los hechos y los conceptos”,<sup>10</sup> de modo que, como complemento de ese marco político, concurre el componente técnico que vincula a unos y a otros. Sin un aval científico sólido, basado en el conocimiento acumulado y, sobre todo, en la interpretación objetiva y rigurosa de los datos epidemiológicos disponibles, resulta virtualmente imposible adoptar medidas de salud pública que sean racionales y efectivas.

Las medidas pueden ser de naturaleza restrictiva, tales como confinamientos, imposición de distancias interpersonales, uso obligatorio de medios de protección o normas para evitar aglomeraciones; pueden también ser de corte flexibilizador, tales como aperturas de espacios públicos o eliminación de determinadas prohibiciones.

En cualquier caso, para decidir lo que pudiera ser epidemiológicamente aconsejable, resulta crucial contar con indicadores omnicomprendivos, que contemplen a la vez la magnitud y la tendencia que exhiben los diversos datos asociados al nivel de alarma que perciben las autoridades de salud. El **Gráfico N°1** permite apreciar que ambos elementos son esenciales para aquilatar la amenaza que se vive en un momento dado. En efecto, dos razones señalan que el número de casos nuevos correspondiente al momento  $t_0$  representa una amenaza mucho menos acusada que la que se registra para el momento  $t_1$ . En primer lugar, porque  $x_0$  es mucho menor que  $x_1$  (e incluso, este último valor supera el record histórico de casos hasta ese momento); y en segundo lugar porque la tendencia de la aparición de nuevos casos es decreciente en la primera fecha y creciente en la segunda.

### 3. Premisas fundamentales para la construcción de un IAE

A partir de todas las consideraciones previas, se procedió a la construcción de un indicador al que se denominó “Índice de alarma epidemiológica” (IAE) por cuyo conducto operacionalizar el concepto de “alarma epidemiológica”. Más concretamente, se trata de lo que se conoce como un “indicador compuesto” o “variable sintética”, noción que se define formalmente como “una función de un conjunto de subindicadores o *variables intermedias*, cada una de las cuales contribuye a cuantificar algún rasgo del concepto cuya magnitud quiere sintetizarse”.<sup>11</sup>

Para la elaboración que he desarrollado, tuve en cuenta las pautas aceptadas de manera virtualmente universal para la construcción de índices compuestos, en particular las que figuran en la excelente revisión de los avances logrados en esta materia en los últimos años desarrollada por un grupo de investigadores de la Universidad de Portsmouth.<sup>12</sup> Hemos respetado en especial la enfática demanda que hacen estos autores de que la metodología seguida para la construcción del índice y para cada una de las decisiones metodológicas que se adopten se comuniquen de manera explícita y transparente. Dentro de este marco global, se establecieron cinco premisas, que se detallan a continuación.

#### 3.1 Construir la variable sintética a partir de variables centinela

Llamaremos *variable centinela* en el contexto de una epidemia a una magnitud que puede medirse para cada uno de los días del proceso epidémico y que señale un mayor nivel de alarma cuanto mayor sea el valor que alcance.

Una premisa será que todas las *variables intermedias* del indicador compuesto tengan la condición de ser “variables centinela”, de modo que el resultado de conjugarlas a través del IAE cumpla su papel esperado: que el crecimiento de la magnitud que alcance cada día exprese una agudización del daño a la salud colectiva y, por ende, constituya un motivo para la alarma.

#### 3.2 Días de referencia inmediata y marco de referencia histórico

El cálculo del IAE se ha de realizar para cualquier día epidémico  $j$  que se elija. Si bien se han de conocer los registros que se produjeron ese día para cada una de las variables centinela involucradas, la idea no es aquilatar de manera encapsulada lo acaecido ese día específico. Se quiere tener una percepción de cuán alarmante resulta la situación atendiendo a la *dinámica epidemiológica* que ha venido presentándose hasta ese momento.

En el cómputo del IAE tendrá un peso medular la información que aporten las variables centinela para un lapso que se extiende hasta el día  $j$  y que dura  $L$  días, a los que llamaremos “días de referencia inmediata”.\*

Entrando en un marco formal, llamemos  $X$  a una variable centinela cuyos valores posibles son números naturales para cada día de la epidemia. Supongamos que han transcurrido  $N$  días desde que se comenzó a registrar el valor de esta variable y llamemos  $X_t$  al valor que alcanza  $X$  el día  $t$  para  $t: 1, \dots, N$ .

Si bien  $j$  pudiera ser cualquier día desde el comienzo de la epidemia (es decir, desde el día 1), ocasionalmente solo interesa lo ocurrido a partir de cierto día  $R$ , el cual señalaría el *marco de referencia histórico* de interés, donde  $R$  puede ser el propio día 1 o un día muy posterior.

¿Por qué pudiera no ser razonable elegir el primer día de la epidemia como la fecha de comienzo de esa referencia histórica? Algunas razones posibles se exponen a continuación.

Por una parte, el máximo valor que tomó la variable centinela pudiera haberse producido en un momento en que aún se aplicaban estrategias de enfrentamiento a la epidemia que fueron luego modificados. Otro motivo relevante para desdeñar datos antiguos puede residir en que, a partir de cierto momento, se cambiaron los protocolos de tratamiento o de diagnóstico. Finalmente, están las mutaciones típicas que experimentan algunos virus. Una cepa que circuló de manera dominante en cierta etapa puede haber dado paso a otra con una capacidad de contagio muy diferente. Si se emplean los registros correspondientes a un período en que la cepa circulante era muy contagiosa para valorar lo que ocurrió en una etapa donde la variante viral producía muchos menos contagios, se podría conformar una idea injustificada de que la epidemia ha ido remitiendo. En todos estos casos, lo más sensato sería ceñirse solo a los datos correspondientes a fechas posteriores a aquella en que se hayan producido dichos cambios.

Operativamente, al fijar cierto día de referencia histórico  $R$ , el análisis se realizará solo con los datos recogidos desde ese día en adelante, tal y como si estos registros fueran los únicos con los que se cuenta.

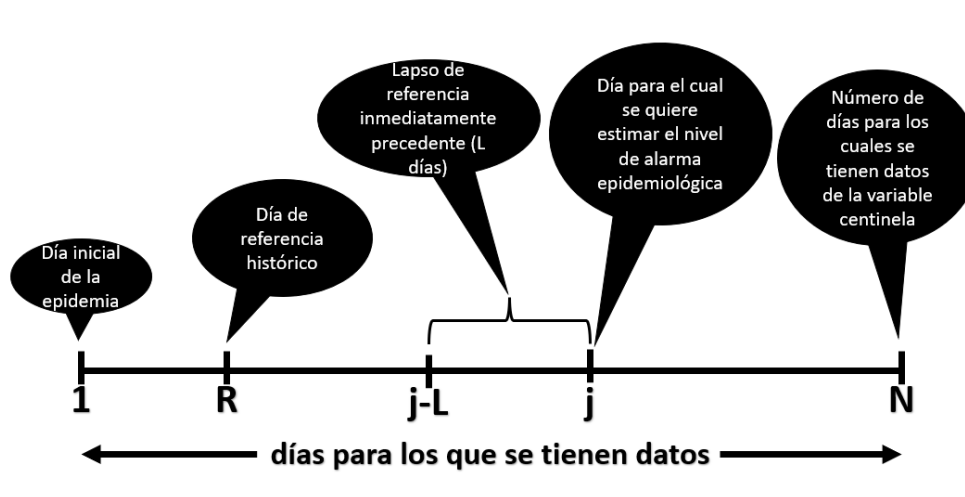
Por otra parte, es posible estimar la alarma asociada al último día para el que se tienen registros de  $X$  (en ese caso,  $j = N$ ).

En síntesis, a través del IAE se cuantificará el grado de alarma epidemiológica correspondiente a un día específico  $j$ , donde  $R + L \leq j \leq N$ . El Gráfico N°2 muestra el orden cronológico en que se ubican los días relevantes.

---

\* La magnitud de  $L$  (que por comodidad será un número par) se decide de manera parcialmente subjetiva y puede cambiar de una epidemia a otra, teniendo en cuenta, por ejemplo, el período de latencia de la infección en el caso de las epidemias asociadas a procesos infecciosos. Resulta aconsejable que  $L$  no sea ni demasiado pequeño (en cuyo caso no se tendría suficiente información contextual), ni demasiado grande (para que no influyan valores correspondientes a comportamientos epidémicos demasiado remotos).

Gráfico N°2. Días relevantes y notaciones correspondientes para la conformación del IAE.



### 3.3 Pesos absoluto y relativo de los valores de las variables centinela

El IAE se basará en los valores que se hayan registrado para la variable  $X$  entre los días  $R$  y  $j$ . Pero particular influencia se concederá a lo que haya ocurrido durante el lapso de referencia inmediata. Una vez fijado  $L$ , el IAE se calcula en principio para un día específico, para una variable concreta y para un marco de referencia histórico prefijado, lo cual se pone de manifiesto la notación elegida para aludir a él:  $IAE_R^X(j)$ .

Por definición sabemos que cuanto más elevados sean los  $X_{j-L+t}$ , ( $t: 1, \dots, L$ ) mayor será el aporte que hagan al nivel de alarma que quepa atribuir al día  $j$ . Por ejemplo, el promedio de esos  $L$  números sería una posible medida de dicho aporte. Sin embargo, mucho más expresivo resulta ese promedio si, en lugar de calcularse usando los valores absolutos de  $X$ , se computara usando valores relativos: los que miden cuán cerca se encuentra  $X_{j-L+t}$  del máximo valor que la propia variable haya alcanzado en el pasado. Es muy intuitiva la pertinencia de considerar que hay más motivos de alarma en la medida que los valores de la variable en los días de referencia inmediata se acerquen, o incluso superen, a máximos históricos dentro de la propia epidemia.

Repárese en que esta idea responde a un principio metodológico esencial: para conformar un juicio que califique cierta realidad, concerniente a cualquier disciplina, e incluso estético o moral, ha de contarse con un parámetro, un paradigma o un elemento de referencia -elegido con acuerdo a un posicionamiento teórico o a una experiencia empírica previa- en función del cual el analista habrá de pronunciarse.

Por ejemplo, consideremos el caso de la esperanza de vida al nacer (EVN) en muchos países de África, la cual se ubica entre 53 y 65 años. A partir de esos valores podríamos calificar la situación de salud de estos países como “angustiosa”, “injusta” o “precaria”. Pero tales calificativos serían legítimos siempre que implícita o explícitamente se esté reparando en la existencia de naciones con EVN notoriamente superiores, como ocurre con casi todas las naciones europeas (con EVN superiores de los 80 años). Sin embargo, una EVN de 60 años, pongamos por caso, pudiera también

considerarse “alentadora” o “positiva” si, aun siendo baja comparada con la de otras naciones, supera sustantivamente las cifras que ese propio país exhibía en épocas relativamente recientes.

Esto quiere decir que el peso que tendrá un valor ubicado dentro de los  $L$  días de referencia inmediata sobre el nivel de alarma que finalmente se considere apropiado para el día  $j$ , debería depender también del grado en que dicho valor se aproxime a lo que suele llamarse “el pico” epidémico (es decir, el máximo valor que haya alcanzado esa variable entre los días  $R$  y  $j$ ).

### 3.4 Influencia de la tendencia de los datos en el nivel de alarma

Se hayan o no relativizado los valores  $X_{j-L+t}$ , el aporte de estos registros a la alarma global (si solo lo incorporamos al IAE a través de su suma o su promedio) no depende del orden en que ellos hayan ido apareciendo en el tiempo. Sin embargo, es obvio que el nivel de alarma no debería ser independiente de la tendencia con se han ido presentando los valores de  $X$ . Por ejemplo, si dichos valores se fueran incrementando en la medida que avanzan los días, la alarma al final de ese lapso debe ser más acusada que si los  $L$  valores (aun siendo los mismos) hubieran ido disminuyendo. Consecuentemente, la tendencia que se esté manifestando en los datos correspondientes al lapso de referencia inmediata se debe cuantificar objetivamente y, sobre todo, dicha cuantificación debe gravitar adecuadamente sobre el valor que asuma el IAE que a la postre se proponga.

### 3.5 Cotas epidemiológicamente interpretables

Tal y como ocurre con otros esfuerzos de similar naturaleza (por ejemplo, los índices creados por el Fondo de Población de NNUU para cuantificar el Desarrollo Humano de un país, e instrumentos conexos<sup>13</sup> así como muchos otros que intentan operacionalizar nociones sumamente complejas y polémicas, como el llamado *Índice de Percepción de Corrupción* que calcula anualmente para todos los países la ONG [Transparency International](#)), lo ideal es que un indicador de este tipo tenga tanto una cota inferior como una cota superior, que ambas cotas sean alcanzables y cualitativamente interpretables. Más concretamente, en nuestro caso, cabe aspirar a que la construcción ideada consiga que la cota inferior se alcance en el escenario más alentador posible y que en ese caso se tenga  $IAE_R^X(j)=0$ , así como que el mayor valor posible se produzca ante el escenario epidemiológico más sombrío posible, y que en tal caso  $IAE_R^X(j)=100$ .

## 4. Construcción del Índice de Alarma Epidemiológica

Una vez inidentificados los elementos que se consideran relevantes y que serán condensados a través del indicador compuesto, es menester apelar a tres maniobras metodológicas ineludibles: la estandarización de los subindicadores, el peso que se le atribuirá a cada uno de ellos y la manera en que se producirá la “agregación”.<sup>14 15 16</sup>

Teniendo en cuenta las indicaciones metodológicas generales,<sup>12 17</sup> atendidos a las 5 premisas o demandas enumeradas en el epígrafe precedente y respetando la notación ya introducida, a continuación, se expone el algoritmo que da cuenta de la propuesta del IAE.

Un repaso a los términos ya convenidos nos permite recordar que para comenzar construiremos un índice que exprese cuán alarmante fue o es la situación epidemiológica prevaleciente cierto día  $j$

partiendo de los datos de una única variable centinela  $X$ , registrados entre el día  $R$  y el propio día  $j$ , donde  $R$  es un número seleccionado que cumple la restricción  $1 \leq R \leq j - L$  y donde  $L$  es la extensión de un lapso previo al día  $j$  sobre el cual hemos decidido concentrar nuestro interés.

Sea  $M_R^X(j - L + t)$  el mayor valor de  $X$  registrado entre los días  $R$  y  $j-L+t$ . La razón

$$P_{j-L+t}^R = \frac{X_{j-L+t}}{M_R^X(j-L+t)} \text{ para } t: 1, \dots, L \quad [3.1]$$

es un número que se hallará necesariamente entre 0 y 1. El “índice de alarma epidemiológica” correspondiente al día  $j$ , que denotaremos como  $IAE_R^X(j)$ , se calculará usando exclusivamente estos  $L$  valores.

Inicialmente, se calcula el promedio

$$P_R^X(j) = \frac{1}{L} \sum_{t=1}^L P_{j-L+t}^R \quad [3.2]$$

Es fácil ver que este promedio alcanzará el máximo valor posible ( $P_R^X(j) = 1$ ) solo en caso de que todos los sumandos de [3.2] sean iguales a la unidad, lo que equivale a que, para cada uno de los  $L$  días de referencia inmediata, el valor de  $X$  haya constituido un record a partir del día  $R$ ;  $P_R(j)$  será igual a 0 solamente cuando  $X_{j-L+t} = 0$  para todo  $t: 1, \dots, L$ . Es obvio que  $P_R^X(j)$  representa el aporte cuantitativo que ha hecho  $X$  a lo largo de los últimos  $L$  días a la situación de alarma prevaleciente el día  $j$ ; consecuentemente, es natural que sea un “ingrediente” de  $IAE_R^X(j)$ .

Pero el nivel de alarma no debería ser independiente de la tendencia con se han ido produciendo los valores de  $P_{j-L+t}^R$ . Consideremos la pendiente  $b_R^*(j)$  de la recta de regresión lineal ajustada usando los  $L$  pares  $(t, P_{j-L+t}^R)$  donde  $t: 1, \dots, L$ . Si ella fuera positiva, tendríamos un indicio de que la situación ha ido empeorando durante los días de referencia inmediata y se tendrían, por ende, más motivos para la alarma que si, por ejemplo,  $b_R^*(j) = 0$ . Inversamente, una pendiente negativa estaría indicando que la situación ha ido mejorando a lo largo de los días de referencia inmediata. De modo que la cuantificación de la pendiente sería el segundo “ingrediente” para conformar el  $IAE_R^X(j)$ . Pero para conjugar este componente con  $P_R^X(j)$ , lo ideal sería tener la pendiente “normalizada”; es decir, expresada como una fracción del máximo valor que pudiera alcanzar.

En general la pendiente de una recta de regresión lineal ajustada no está ni inferior ni superiormente acotada, pero en las condiciones de nuestro problema (donde las ordenadas son  $L$  números naturales consecutivos y las abscisas son números reales entre 0 y 1), afortunadamente, si lo está (véase la demostración en el ANEXO). En particular, cuando  $L$  es par, los valores posibles de  $b_R^*(j)$  se ubican en el intervalo  $[-\omega_L, \omega_L]$  donde

$$\omega_L = \frac{3L}{2(L^2-1)} \quad [3.3]$$

Consecuentemente, la tendencia se puede “normalizar” del modo siguiente:

$$b_R(j) = b_R^*(j)/\omega_L \quad [3.4]$$



de modo que se cumple que  $-1 \leq b_R(j) \leq 1$

Si para conjugar los dos componentes de la alarma (magnitud de  $X$  y tendencia) decidimos conceder 4 veces más peso al aporte de  $X$  que a la tendencia, llegaríamos a la propuesta siguiente:

$$IAE_R^X(j) = P_R^X(j) + (0,25)b_R(j) \quad [3.5]$$

Es fácil corroborar que  $0 \leq IAE_R^X(j) \leq 1$ .

Ahora supongamos que tenemos  $k$  variables centinela  $X_i$  ( $i=1, \dots, k$ ). Siguiendo para cada una de ellas el mismo procedimiento que se ha descrito para la que hemos llamado genéricamente  $X$ , tendríamos  $k$  índices de alarma para el día  $j$ . Llamémosle  $IAE_R^{X_i}(j)$ . La forma natural de conjugarlos es, nuevamente, conformando una media ponderada para arribar a la forma final del IAE para el día  $j$ :

$$IAE_R(j) = \sum_{i=1}^k \lambda_i \cdot IAE_R^{X_i}(j) \quad [3.6]$$

donde los factores  $\lambda_i$  son números positivos que cumplen la condición  $\sum_{i=1}^k \lambda_i = 1$  y se definen con acuerdo al peso que se decida conferir a cada variable centinela en cuanto a la alarma que agregan.

## 5. Cómputo de un intervalo de confianza para el IAE

El valor del índice depende de números diarios correspondientes a las variables centinela. Esos números son “registros” que no necesariamente coincidan con lo que realmente ha ocurrido cada día, sino que se verán afectados por distorsiones que pudieran deberse a fenómenos tales como alteraciones accidentales en los registros estadísticos, a inestabilidades en el trabajo de los laboratorios o a deficiencias coyunturales en los sistemas de flujo informativo. Consecuentemente, tiene sentido pensar en el cómputo de un intervalo de confianza que contemple estas fluctuaciones. Para concretar esta idea se procede del modo siguiente.

Llamemos  $X_{uj}$  ( $u=1, \dots, L$ ) a los  $L$  valores de la variable  $X$  registrados para los días de referencia inmediata asociados a un día  $j$ . Sea  $s(j)$  la desviación estándar de dichos valores. Aplicando la fórmula  $X_{uj}^* = X_{uj} + e_{uj}$ , donde  $e_{uj}$  es una realización de una distribución normal con media igual a  $0$  y desviación estándar igual a  $s(j)/10$ , se obtienen respectivos datos simulados.

Ahora se procede a computar el  $IAE_R^X(j)$  tal y como se hizo antes, pero usando los  $X_{uj}^*$  en lugar de los  $X_{uj}$ . Este mismo procedimiento se lleva adelante para cada una de las  $k$  variables centinelas consideradas, de modo que con esos  $k$  resultados se podrá calcular  $IAE_R(j)$  empleando la fórmula [3.6] Esta generación de  $L$  valores se realiza en  $\eta$  ocasiones (por ejemplo, se repite  $\eta = 500$  veces). Con cada una de esas  $\eta$  simulaciones se calcula el índice  $IAE_R(j)$ . La media de estos  $\eta$  resultados

podiera tomarse como una estimación puntual del **IAE**, en tanto que los percentiles 2,5% y 97,5% de este conjunto de  $\eta$  magnitudes simuladas, serían los extremos de un intervalo de confianza al 95% para el indicador.

## 6. Traducción de los valores del IAE a una escala cualitativa

Una vez que se cuenta con el valor numérico que procura cuantificar el grado de alarma, se sugiere calificarlo según la siguiente regla:

- $IAE_R(j) < 20\%$  .... Situación no alarmante
- $20\% \leq IAE_R(j) < 40\%$  .... Situación de alarma reducida
- $40\% \leq IAE_R(j) < 60\%$  .... Situación de alarma notoria
- $60\% \leq IAE_R(j) < 80\%$  .... Situación de alarma elevada
- $IAE_R(j) \geq 80\%$  .... Situación de alarma apremiante

## 7. Ilustración de la aplicación del índice de alarma epidemiológica usando datos reales

La construcción del índice que se ha expuesto depende de un conjunto de parámetros que hay que fijar al momento de implementarlo y aplicarlo en una situación real. Seleccionado un día  $j$  para el análisis, hay que elegir:

- un conjunto de  $k$  variables centinela
- un lapso  $L$  de días de referencia inmediata
- un día de referencia histórico  $R$
- los coeficientes de ponderación  $\lambda_i$  ( $i=1, \dots, k$ )

El índice puede aplicarse en cualquier espacio geográfico que se elija: podría ser un país, una provincia, una ciudad o región, e incluso su aplicación podría ceñirse a un grupo poblacional prefijado (e.g. mujeres mayores de 65 años).

Para ilustrar el proceso, usaré datos registrados oficialmente<sup>18</sup> durante la epidemia de COVID-19 en la República de Cuba a lo largo de los  $N = 677$  días transcurridos entre el 11 de marzo de 2020 y el 16 de enero de 2022.

Manejaremos  $k = 3$  variables centinela:  $X_1$ =número de “casos activos”,  $X_2$ =número de “fallecidos” y  $X_3$ =número de “casos nuevos” registrados para cada uno de los 677 días. Aunque para la construcción final del índice se operará con estas tres variables, combinando sus resultados (según se ha explicado en el epígrafe anterior y como se verá más adelante) para la ilustración que sigue nos ceñiremos inicialmente a la medición del **IAE** usando solo la primera de ellas.

En el marco de esta epidemia concreta, es razonable fijar el valor  $L = 14$  días para marcar un contexto ni muy lejano ni muy cercano al día epidémico en el que habrá de centrarse el análisis. Ese número, además, es el que se ha estimado como el período de incubación en el caso del virus SARS-COV-2, que produce la enfermedad.<sup>19</sup>

Imaginemos que queremos analizar la situación que prevalecía el 9 de septiembre de 2020. Este fue el día epidémico **183** (es decir, se fija  $j = 183$ ). En primer lugar, se identifican los valores de  $X$  (casos activos) correspondientes a los 14 días consecutivos que comienzan el día 170 tal y como se recoge en La **Tabla Nº1**.

**Tabla Nº1. Número de casos activos para el día epidémico  $j=183$  y los 13 días precedentes. Epidemia de COVID-19. Cuba.**

$t$	$j - 14 + t$	$X_{j-14+t}$
1	170	550
2	171	556
3	172	550
4	173	558
5	174	573
6	175	568
7	176	638
8	177	677
9	178	626
10	179	616
11	180	606
12	181	571
13	182	626
14	183	664

Para operar con datos correspondientes a una etapa en que los procedimientos de diagnóstico y de notificación estadística ya estaban consolidados, se desdeña la información correspondiente al primer mes de la epidemia. Eso equivale a fijar el 9 de abril de 2020 ( $R=30$ ) como día de referencia histórico. Ahora, para cada uno de los 14 días se identifica el mayor valor de  $X$  que se haya registrado desde el día 30 hasta entonces. Una revisión de los datos permite constatar que el mayor número de casos activos a partir del trigésimo día de la epidemia se produjo el día **45** (24 de abril de 2020), cuando se contabilizaron 847 enfermos activos. Como ninguno de los valores de la tercera columna de la **Tabla Nº1** supera a ese máximo, tenemos que  $M_R^X(j - L + t) = 847$  cualquiera sea  $t$  entre **1** y **L**. La **Tabla Nº2** muestra el resultado de usar este máximo como denominador para el cálculo de  $P_{j-L+t}^R$  con acuerdo a la fórmula [3.1].\*

---

\* Nótese que, en este caso, el valor que se ubica en la penúltima columna es el mismo en todas las filas; pero si algún valor de  $X$  entre los días de referencia inmediata hubiese sido mayor que 847, a partir de dicho día los denominadores usados para conformar la última columna, hubieran cambiado. Por ejemplo, si se hubiese producido que  $X_{179} = 903$ , con el resto de los valores de  $X$  sin modificación, entonces  $P_{179}^{30}$  hubiera sido igual a 1, y los siguientes 4 valores de la quinta columna hubieran resultado de dividir  $X_{j-14+t}$  por el nuevo record.

**Tabla N°2. Casos activos registrados los días del 170 al 183 y fracción que estos registros representan respecto del mayor número de casos activos registrados desde el día R=30 hasta cada uno de esos días. Epidemia de COVID-19. Cuba.**

$t$	$j - 14 + t$	$X_{j-14+t}$	$M_{j-14+t}^{30}$	$P_{j-14+t}^{30}$
1	170	550	847	0,649
2	171	556	847	0,656
3	172	550	847	0,649
4	173	558	847	0,659
5	174	573	847	0,677
6	175	568	847	0,671
7	176	638	847	0,753
8	177	677	847	0,799
9	178	626	847	0,739
10	179	616	847	0,727
11	180	606	847	0,715
12	181	571	847	0,674
13	182	626	847	0,739
14	183	664	847	0,784

Ahora, según [3.2] se obtiene  $P_R^X(183) = \frac{9,893}{14} = 0,7066$  y con los  $L=14$  pares  $(t, P_{169+t}^{30})$ , donde  $t: 1, \dots, 14$ , que se conforman usando la primera y la quinta columna de la **Tabla N°2**, se calcula la pendiente de la recta ajustada:  $b_{30}^*(183) = 0,0083$ . De acuerdo a la fórmula [3.3], sabemos que el máximo valor que podría haber alcanzado esta pendiente es:

$$\omega_{14} = \frac{3 \cdot 14}{2(14^2 - 1)} = \frac{21}{195} = 0,1077$$

de modo que la pendiente “normalizada”, asciende a:

$$b_{30}(183) = 0,0083/0,1077 = 0,0771$$

En síntesis, la pareja  $P_R^X(j)$  y  $b_R(j)$  cuantifican respectivamente la magnitud de la alarma que aporta el total de los casos acumulados en los últimos 14 días y la tendencia que exhiben dichos registros a lo largo de las dos semanas.

Aplicando la fórmula [3.5], para la variable centinela  $X_1$  con que hemos venido operando (“casos activos”, en este ejemplo) obtenemos el valor del Índice de Alarma Epidemiológica:

$$IAE_{30}^{X_1}(183) = P_R(j) + (0,25) b_R(j) = 0,7066 + (0,0771)/4 = 0,7259$$

Este proceso se ha desarrollado en detalle para  $X_1$ =número de “casos activos”; cuando se reproduce para las otras dos variables centinela ( $X_2$ =número de “fallecidos” y  $X_3$ =número de “casos nuevos”), da lugar a:

$$IAE_{30}^{X_2}(183) = 0,5786 \quad \text{y} \quad IAE_{30}^{X_3}(183) = 0,2730$$

Ahora, toca calcular el Índice de Alarma Epidemiológica final conjugando

Tomando  $\lambda_1 = \lambda_2 = 0,3$  y  $\lambda_3 = 0,4$  (es decir, eligiendo un peso ligeramente mayor para los fallecidos que para los nuevos contagios y el número de activos) se conjugaron los índices correspondientes a las tres variables centinela a través de la fórmula [3.6]:

$$IAE_R(j) = \sum_{i=1}^3 \lambda_i \cdot IAE_R^{X_i}(j)$$

$$= (0,3) IAE_{30}^{X_1}(183) + (0,3) IAE_{30}^{X_2}(183) + (0,4) IAE_{30}^{X_3}(183) = 0,5008$$

Llevado a porcentaje, el valor del índice resulta ser:

$$IAE_{30}(183) = 50,08\%$$

En la escala cualitativa, este resultado conduce a calificar la situación prevaleciente el día 183 (9 de septiembre de 2020) como de “alarma notoria”.

Al calcular un intervalo de confianza, el resultado fue el siguiente: *IAE estimado=50,05%, IdeC(95%): 49,24-50,82*. Nótese que, puesto que estos valores se derivan de un proceso de simulación, los resultados que se obtengan cambiarán ligeramente si dicho proceso se repitiera.

Nota complementaria: Si en un enclave dado no se contara con la información diaria de los “casos activos” sino solo de los “casos nuevos” y de los “fallecidos”, el índice debe adaptarse. Todo sería igual salvo la ponderación final, que solo involucraría a los dos índices parciales asociados. Se sugiere en tal caso el empleo de la fórmula

$$IAE_R(j) = (0,4)IAE_R^{X_2}(j) + (0,6)IAE_R^{X_3}(j)$$

Si en el ejemplo se opera solo con esas dos variables centinela, se obtendría:

$$IAE_{30}(183) = (0,4)(0,5786) + (0,6)(0,2730) = 0,3952$$

## 8. Aplicación informática

Para el caso particular en que se quiere operar con las  $k=3$  variables centinelas empleadas en el ejemplo del epígrafe precedente, se ha confeccionado un programa informático en lenguaje R altamente amigable que puede explotarse en línea accediendo a la dirección:

[https://paufarmart.shinyapps.io/IAE\\_covid/?\\_ga=2.60102958.2054292187.1634898404-1154226650.1634757898](https://paufarmart.shinyapps.io/IAE_covid/?_ga=2.60102958.2054292187.1634898404-1154226650.1634757898)

También puede obtenerse el programa descargando una carpeta desde el sitio <http://lcsilva.sbhac.net/Otros/Aportes%20covid/covid.htm> la cual contiene un archivo .bat que permite explotar el programa informático desde un navegador sin necesidad de estar conectado a la red.

La entrada al programa consiste en un archivo en Excel que contenga los datos necesarios, así como la indicación del día para el cual se quiere computar el Índice de Alarma Epidemiológica y el día inicial de referencia (**R**).

El programa tiene una pestaña donde figuran las instrucciones para su empleo y otra por conducto de la cual se puede descargar los archivos de prueba que contienen los datos empleados para el ejemplo desarrollado (datos de COVID-19 para Cuba) y que ilustra cómo deben aparecer los datos en la hoja del archivo de Excel para explotar el programa.

## ANEXO

**TEOREMA:** Sea  $y = a + bx$  la ecuación de regresión lineal entre las variables  $x$  e  $y$ . Para  $n$  ( $n \neq 1$ ) pares ordenados  $(x_i, y_i)$  se sabe que:

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Si los  $x_i$  son números enteros sucesivos y se cumple que  $0 \leq y_i \leq 1$ , entonces se cumple que  $-\Omega \leq b \leq \Omega$ , donde  $\Omega = \frac{3n}{2(n^2-1)}$  si  $n$  es par y  $\Omega = \frac{3}{2n}$  si  $n$  es impar.

### Demostración

Llamemos  $N$  al numerador de  $b$ ,  $D$  al denominador y  $k$  a la parte entera de  $\frac{n}{2}$ . Una vez fijado  $x_1$ , queda establecido el resto de los  $x_i$  ya que  $x_i = x_{i-1} + 1$  para  $i = 2, \dots, n$ . El valor de  $b$  es el mismo cualquiera sea  $x_1$ , debido a que diferentes valores de  $x_1$  producirían diferentes rectas, pero todas paralelas entre sí. De modo que si establecemos que  $x_1 = -k$ , ya queda determinado el valor de  $D$ , que no depende de los  $y_i$ .

Demostraremos que ese máximo se alcanza cuando se cumple la siguiente condición:

$$y_i = 0 \text{ para } i: 1, \dots, k \text{ y } y_i = 1 \text{ para } i: k + 1, \dots, n$$

Llamemos  $V_M$  al vector  $(0, 0, \dots, 0, 1, 1 \dots 1)$  donde el primer 1 se ubica en el lugar  $k + 1$ . En tal caso  $\sum_{i=1}^n y_i = k^*$ , donde  $k^*$  es  $k$  o  $k + 1$ , dependiendo de que  $n$  sea par o impar respectivamente; de modo que si llamamos  $S_x = \sum_{i=1}^n x_i$ , tendremos  $N = n \sum_{i=1}^n x_i y_i - S_x k^*$ . Puesto que  $S_x k^*$  no depende de los  $y_i$ , solo resta probar que  $\sum_{i=1}^n x_i y_i$  como función de los  $y_i$  alcanza el máximo cuando se evalúa en  $V_M$ .

Repárese en que  $\sum_{i=1}^n x_i y_i = S_1 + S_2$  donde  $S_1 = \sum_{i=1}^k x_i y_i$  y  $S_2 = \sum_{i=k+1}^n x_i y_i$ . Como todos los  $x_i$  que figuran en  $S_1$  son negativos, si alguno

de los respectivos  $y_i$  fuera superior a 0,  $S_1$  disminuiría respecto del caso en que dichos  $y_i$  son nulos. Análogamente, puesto que ninguno de los  $x_i$  que figuran en  $S_2$  son negativos, si alguno de los  $y_i$  presentes en  $S_2$  fuera menor que 1, dicho segundo sumando nunca podría incrementarse, con lo cual queda demostrado que  $V_M$  maximiza a  $\sum_{i=1}^n x_i y_i$ , y por tanto también a  $D$  y a  $b$ .

Ahora podemos calcular el valor de  $\Omega$ . Tras unas pocas maniobras algebraicas se arriba a que  $D = \frac{n^2(n^2-1)}{12}$  cualquiera sea la paridad de  $n$ . El numerador  $N$ , en cambio, depende de que  $n$  sea par o impar; es fácil corroborar que, en el primer caso, se tiene  $N = \frac{n^3}{8}$ , mientras que, si  $n$  es impar, entonces  $N = \frac{n(n^2-1)}{8}$ .

Haciendo las divisiones  $\frac{N}{D}$  respectivas, tendríamos finalmente que  $\Omega = \frac{3n}{2(n^2-1)}$  cuando  $n$  es par y  $\Omega = \frac{3}{2n}$  si  $n$  es impar.

Por último, siguiendo el mismo proceso se demuestra que el mínimo de  $b$  sería igual a  $-\Omega$  si suplimos los ceros por unos y viceversa en el vector  $V_M$ .

+++++

Nota: resulta interesante que el vector que maximiza a  $b$  como función de los  $y_i$  es único si  $n$  es par y coincide con  $V_M$ . Sin embargo, cuando  $n$  es impar, el máximo se alcanza para cualquier vector de la forma  $(0, 0, \dots, 0, y_{k+1}, 1, 1 \dots 1)$  cualquiera sea el valor de  $y_{k+1}$ . Obviamente,  $V_M$  es uno de ellos.



## Bibliografía

---

- <sup>1</sup> Rojas F, López C, Silva LC. (1993) Indicadores de salud y bienestar en municipios saludables (Doc. HPP/HPS/94.30), OPS/OMS, Washington, D. C., 1994.
- <sup>2</sup> Brett TS, Rohani P. (2020) Dynamical footprints enable detection of disease emergence. *PLoS Biology*. 18, e3000697.
- <sup>3</sup> Chang F-S, Tseng Y-T, Hsu P-S y col. (2015) Re-assess vector indices threshold as an early warning tool for predicting dengue epidemic in a dengue non-endemic country. *PLoS Negl Trop Dis* 9(9): e0004043. <https://doi.org/10.1371/journal.pntd.0004043>
- <sup>4</sup> Vega T, Lozano JE, Meerhoff T y col. (2013) Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. *Influenza and Other Respiratory Viruses* 7, 546–558.
- <sup>5</sup> Abeku TA, Hay SI, Ochola S y col. (2004) Malaria epidemic early warning and detection in African highlands. *Trends in Parasitology* 20: 400–405.
- <sup>6</sup> Collins SD. (1932) Excess mortality from causes other than influenza and pneumonia during influenza epidemics. *Public Health Reports* 47 (46):2159-2179.
- <sup>7</sup> WHO. (2017) Operational guide: Early Warning and Response System (EWARS) for dengue outbreaks. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
- <sup>8</sup> Kostoulas P, Meletis E, Pateras K y col. (2021) The epidemic volatility index, a novel early warning tool for identifying new waves in an epidemic. *Science Reports* 11, 23775.
- <sup>9</sup> Gutiérrez D. 2009. La Construcción de Indicadores como Problema Epistemológico. *Cinta Moebio*. 34:16-36.
- <sup>10</sup> Samaja J. (1993). *Epistemología y metodología*. Buenos Aires: Eudeba.
- <sup>11</sup> Silva LC. (1997). *Cultura estadística e investigación científica en el campo de la salud: Una mirada crítica*. Madrid: Díaz de Santos. (epígrafe 4.1)
- <sup>12</sup> Greco S, Ishizaka A, Tasiou M, Torrisi G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141:61–94 .
- <sup>13</sup> Wikipedia. Índice de desarrollo humano. [https://es.wikipedia.org/wiki/%C3%8Dndice\\_de\\_desarrollo\\_humano](https://es.wikipedia.org/wiki/%C3%8Dndice_de_desarrollo_humano)

---

<sup>14</sup> Saturno PJ. (2004) La invasión de los indicadores compuestos. Riesgos y beneficios para la gestión de la calidad. *Revista de Calidad Asistencial* 19(6):407-415.

<sup>15</sup> Schuschny AR, Soto H. (2009) Guía metodológica: diseño de indicadores compuestos de desarrollo sostenible. Documento de proyecto. Santiago de Chile: CEPAL.

<sup>16</sup> Camacho M, Horta R. (2020). Metodologías para la construcción de índices compuestos. Working Paper 1/2020. Departamento de Administración y Finanzas. Universidad Católica del Uruguay. ISSN: 2697-2875.

<sup>17</sup> Becker W, Saisana M, Paruolo P y col. (2017). Weights and importance in composite indicators: closing the gap. *Ecological Indicators*, 80:12-22.

<sup>18</sup> Informaciones del MINSAP. Periódico Granma. República de Cuba.  
<https://www.granma.cu/informacion-minsap>

<sup>19</sup> Lauer SA, Grantz KH, Bi Q, y col (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine* 172(9):577-582.