



METODOLOGÍAS

Estimación del sub-reporte de casos activos debido al rezago en el reporte de casos: Metodología de Yan y Chowell

Los informes oficiales a menudo presentan recuentos agregados basados en el número de nuevos diagnosticados. Estos informes presentan un sesgo de duración en la falta de información, tal fenómeno es causado por la demora en reportar. Los retrasos en los informes se miden a niveles individuales como la brecha entre el momento del inicio de la enfermedad (o diagnóstico) y el momento en que este caso individual se informa y se ingresa en la base de datos de un registro de salud pública. Este sesgo de tiempo es más profundo cuando los datos se compilan y analizan cuando el brote aún está en curso, pero también es más importante presentar las tendencias en tiempo real durante la investigación del brote (Yan and Chowell, 2019).

Los datos sobre el retraso del informe siempre se truncan a la derecha, porque solo después del informe se puede medir retrospectivamente el retraso.

Hay dos niveles de sesgo de duración involucrados:

- En los recuentos agregados: disminución artificial de tendencia cerca del final de la serie temporal.
- Por longitud debido al truncamiento a la derecha: retrasos más cortos se sobrerrepresentan en los datos (Gráfica 1).

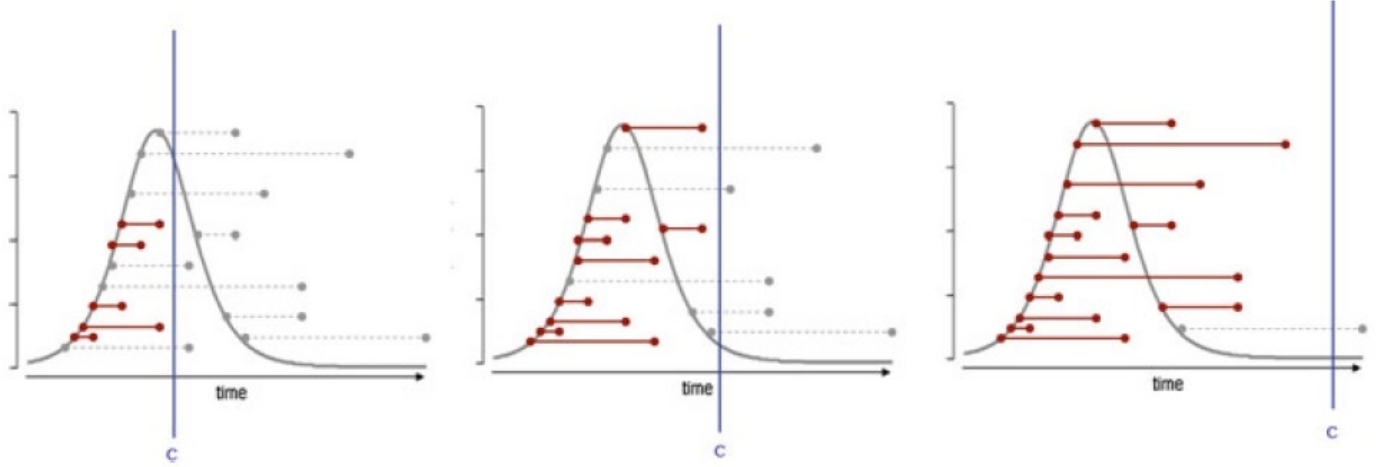


Figura 1: Ilustración esquemática de períodos de incubación comprobados retrospectivamente analizados durante, inmediatamente después y mucho después de un brote de enfermedad

Estimación

Estos recuentos se organizan en una tabla de contingencia de 2 vías en la que el triángulo inferior permanece vacío debido al truncamiento a la derecha.

- Evento iniciador: Inicio de la enfermedad.
- C: Momento más actual cuando se analizan los datos.
- t: Momento de la ocurrencia de los eventos, $t = 0, 1, \dots, C$.
- X: Demora en la presentación de informes (variable aleatoria iid entre todos los individuos). La distribución acumulativa es:

$$F(x) = F(C - t) = \Pr(X \leq C - t)$$

- $N(t)$: número de eventos que ocurrieron en el momento t.

$$N(t) = \frac{N(t; C)}{F(C - t)}$$

- n_{tx} : eventos que ocurrieron en el momento t y fue informado en $t + x$, $x \leq C - t$.
- $N(t; C)$ (recuentos agregados): número de eventos que ocurrieron en el momento t según lo informado por el tiempo C (total de filas).

$$N(t; C) = N(t; C + 1) \frac{F(C - t)}{F(|C + 1| - t)}$$

$$N(t; C) = N(t; C + 1)(1 - g(C - t + 1))$$

- n_{+x} : número de eventos que ocurrieron en el momento t (total de columnas).

$$n_{+x} = \sum_{t=0}^{C-x} n_{tx}$$

- N_{+x} : número total de casos con retraso $X \leq x$, entre eventos durante tiempos $t = 0, 1, \dots, C-x$ (suma de números en la tabla de contingencia de 2 vías dentro del área del rectángulo definida por $t = 0, \dots, C$ y $X = 0, \dots, x$.)

$$N_{+x} = \sum_{t=0}^{C-x} \sum_{j=0}^x n_{tj}$$

- $g(x)$: ratio de n_{+x}/N_{+x}

Por tanto la estimación de número de eventos que ocurrió en el momento t es:

$$\hat{N}(t) = \frac{N(t; C)}{\prod_{x=C-t+1}^C (1 - g(x))}$$

$$\hat{N}(t) = \frac{N(t; C)}{\hat{F}(C - t)}$$

Referencias

Yan, P. and Chowell, G. (2019). *Characterizing Outbreak Trajectories and the Effective Reproduction Number*, pages 273–315. Springer International Publishing, Cham.

	Reporting delay x							Row totals
	0	1	\dots	x	\dots	$C-1$	C	
0	n_{00}	n_{01}	\dots	n_{0x}	\dots	$n_{0,C-1}$	n_{0C}	$N(0; C)$
1	n_{10}	n_{11}	\dots	n_{1x}	\dots	$n_{1,C-1}$		$N(1; C)$
\vdots	\vdots	\vdots		\vdots				
t	n_{t0}	n_{t1}	\dots	n_{tx}				$N(t; C)$
\vdots	\vdots	\vdots		\vdots				
$C-x$	$n_{C-x,0}$	$n_{C-x,1}$	\dots	$n_{C-x,x}$				
\vdots	\vdots	\vdots						
$C-1$	$n_{C-1,0}$	$n_{C-1,1}$						
C	n_{C0}							$N(C; C)$
Col. totals	n_{+1}		n_{+x}			n_{+C}		

Figura 2: Tabla de contingencia para estimación del número de eventos