

Método de captura y recaptura: una experiencia de su aplicación para determinar la incidencia de diabetes gestacional

Vanina Pagotto^{1,✉}, Diego Giunta^{1,✉}, Pablo A. Pochettino^{2,✉} y María Lourdes Posadas-Martínez^{1,✉}

1. Área de Investigación No Patrocinada, Departamento de Investigación, Hospital Italiano de Buenos Aires. Buenos Aires, Argentina

2. Sección de Diabetes y Embarazo, Servicio de Obstetricia. Departamento de Cirugía. Hospital Italiano de San Justo Agustín Rocca. Buenos Aires, Argentina

RESUMEN

Los métodos de captura y recaptura (MCR) se emplean en la estimación de poblaciones mediante la utilización de diferentes fuentes de datos, disponibles e incompletas, que registran por separado un mismo evento. En esta metodología, las fuentes son utilizadas para extrapolar el número de individuos no registrados, usando la información recopilada sobre los individuos sí registrados. Este artículo describe todos los pasos de su aplicación práctica, a partir de un ejemplo de estimación de la incidencia de diabetes gestacional en una institución, a partir de cinco fuentes documentales.

Palabras clave: tamaño de la muestra, diabetes gestacional, captura-recaptura, recolección de datos, diseño de investigación.

Capture and recapture method: an experience of its application to estimate the incidence of gestational diabetes

ABSTRACT

Capture-recapture (CRM) methods are widely used to estimate populations by using different data sources, available and incomplete, that record the same event separately. In these methods, the available sources are used to extrapolate the number of unregistered individuals, using the information collected on the individuals that are registered. This article describes all the steps of its practical application, based on an example of estimating the incidence of gestational diabetes in an institution based on five documentary sources.

Key words: data collection, research design, sample size, gestational diabetes.

Un equipo de investigación se preguntó cuál sería la incidencia de diabetes gestacional en la población atendida en su institución. Luego de reunirse con diversos actores que conocían bien cómo se generaba la información vinculada a los procesos de atención, llegaron a la conclusión de que los casos de diabetes gestacional podían ser identificados a partir de cinco

fuentes (valores de laboratorio; problema cargado en la historia clínica; evoluciones médicas; listado de pacientes atendidas por obstetras y tratamiento farmacológico). Dichas fuentes tenían un cierto nivel de superposición (algunas pacientes se encontraban en más de una fuente). Se preguntaron, además, si era posible estimar con algún tipo de precisión otros

Autor para correspondencia: vanina.pagotto@hospitalitaliano.org.ar, Pagotto V.

Recibido: 4/02/22 Aceptado: 15/07/22 En línea: 30/09/2022

DOI: <https://doi.org/10.51987/revhospitalbaire.v42i3.207>

Cómo citar: Pagotto V, Giunta D, Pochettino PA, Posadas-Martínez ML. Método de captura y recaptura: una experiencia de su aplicación para determinar la incidencia de diabetes gestacional. *Rev. Hosp. Ital. B.Aires.* 2022;42(3):168-172.

casos realmente existentes que no fueron registrados en ninguna de esas cinco fuentes.

Introducción a los métodos de captura-recaptura

Los métodos de captura y recaptura (MCR) son empleados para estimar las poblaciones; a partir del siglo XIX comenzaron a utilizarse en poblaciones animales y luego en medicina, demografía y epidemiología¹.

En su aplicación tradicional, el método consiste en tomar una muestra de una población animal, marcar a los individuos capturados sin producir daño, luego permitir que estos animales se mezclen con la población restante, y, al cabo de un período de tiempo corto, volver a tomar una nueva muestra y contar el número de individuos recapturados.

Los cuatro supuestos básicos de este modelo son:

- La población es cerrada.
- Las muestras son independientes.

-La probabilidad de captura de todos los individuos es la misma.

-Los individuos capturados en las dos ocasiones pueden ser identificados (las marcas no se pierden).

En la figura 1 se esquematiza un ejemplo del método de captura y recaptura empleado para la estimación de una población de animales.

A partir de este planteo básico surgen modelos más complejos que consideran los casos en los que estos supuestos no se cumplen. Así, se utilizan modelos abiertos, para modelar datos de investigaciones de largo plazo en animales o en aves migratorias, y permiten estimar tasas de supervivencia, de natalidad y migratorias.

En medicina, estos MCR son métodos analíticos que se emplean para realizar estimaciones cuando se cuenta con diferentes fuentes de datos que registran por separado un mismo evento. Las estimaciones por MCR consisten en aprovechar la superposición de registros presentes en fuentes disponibles, que se conoce son incompletas, para

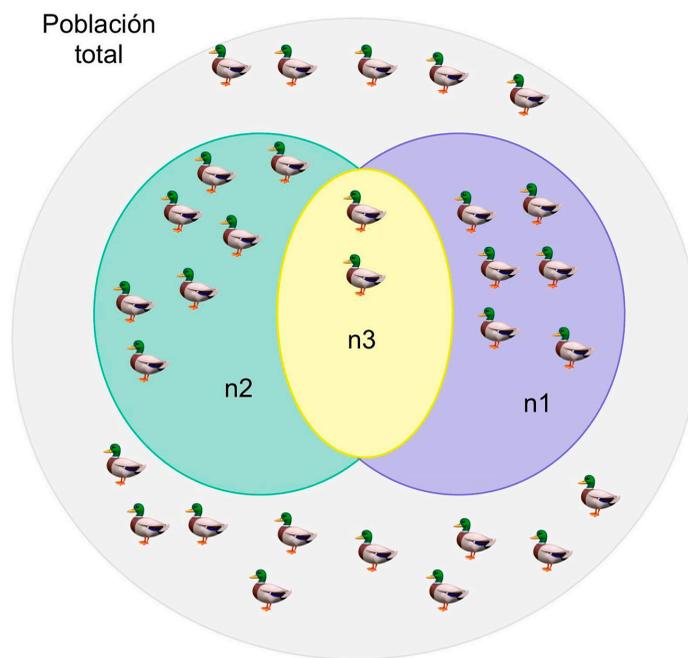


Figura 1. Esquematización del método de captura y recaptura utilizado para estimar la población de patos. El método consiste en marcar un área, luego se capturan, cuentan, marcan y liberan una selección de individuos (n_1). Después de que ha pasado suficiente tiempo para permitir que los individuos marcados se reintegren a la población, se realiza una segunda captura (n_2). En esta segunda captura se contabilizan tanto los individuos no marcados como los marcados (n_3). Sobre la base de los tres valores generados (n_1 ; n_2 ; n_3), se estima el tamaño de la población.

En gris se representa la población total.

n_1 es la muestra de la población obtenida en la primera captura.

n_2 es la muestra de la población obtenida en la segunda captura.

n_3 es la muestra recapturada.

Donde

$$\text{Población total} = \frac{\text{marcados en la primera captura} * \text{marcados en la segunda captura}}{\text{recapturados}}$$

Adaptado de Population Sampling. Capture-Mark-Release-Recapture⁹.

medir formalmente la subestimación de esas fuentes y estimar la magnitud de un evento². Las fuentes disponibles se utilizan para extrapolar el número de individuos no registrados, usando la información recopilada sobre los individuos sí registrados³.

Las fuentes de datos utilizadas pueden ser registros de enfermedades de notificación obligatoria, estadísticas hospitalarias o de otros servicios de salud, y registros de estadísticas vitales, como nacimientos y muertes, entre otros², y permiten estimar eventos de salud tan diversos como casos de sida, tuberculosis, diabetes, meningitis, discapacidades ocupacionales, enfermedades genéticas, accidentes de tránsito, etcétera¹.

En la mayoría de los trabajos epidemiológicos, el tamaño poblacional puede considerarse aproximadamente constante en un período de tiempo fijo. Para períodos de estudio relativamente cortos, este supuesto puede darse como cierto⁴.

Además, los MCR para poblaciones cerradas incorporan dos posibles fuentes de variación para las probabilidades de captura, es decir, dos fuentes de dependencia entre las muestras:

1) Heterogeneidad entre los individuos: ocasiona que las probabilidades de captura varíen entre los individuos de la población.

2) Dependencia dentro de cada individuo o dependencia de las listas: en la práctica implica que la inclusión de un individuo en una lista tiene un efecto causal directo en su inclusión o no en otras listas.

Los modelos log-lineal, como el modelo de regresión de Poisson, permiten manejar la dependencia entre las listas. Se basan en considerar los datos como parte de una tabla de contingencia incompleta de $2t$ (donde t es el número de muestras/listas) y luego ajustar varios modelos a las variables observadas en las celdas, en los que la celda sin datos es tratada como si fuera un parámetro y estimada por el método de máxima verosimilitud⁵.

Una de las principales limitaciones de la metodología de captura y recaptura en las aplicaciones epidemiológicas se relacionan con la violación de los supuestos para la validez del método. Además, es más probable que los métodos de captura y recaptura produzcan una estimación sesgada del tamaño de la población si una fuente (o combinación de fuentes) captura muy pocos casos. En esta circunstancia, la estimación del número de casos perdidos podría ser cercana a cero o muy grande.

Aplicación del método de captura-recaptura para la estimación de diabetes gestacional

Se aplicó el MCR para estimar la cantidad de embarazadas con diabetes no detectadas por ninguna de las cinco fuentes consideradas. Las cinco fuentes de datos de donde se detectaron los casos fueron:

1. Valores de laboratorio glucemias y pruebas de tolerancia oral a la glucosa.

2. Problema cargado en la historia clínica (diabetes/diabetes gestacional/hiperglucemia/diabetes tipo 1/diabetes tipo 2).

3. Evoluciones médicas.

4. Listado de pacientes atendidas por obstetras.

5. Tratamiento (insulina/metformina).

Se cumplían los supuestos:

– La población era cerrada, dado que todas las embarazadas incluidas realizaban su seguimiento en el Hospital Italiano de Buenos Aires, donde se utiliza una historia clínica electrónica (HCE) centralizada, en la cual todo el proceso de atención quedaba registrado (atención médica ambulatoria, internación, procedimientos, laboratorios, fármacos)⁶.

– Las muestras eran independientes: se consideró en el caso de las mujeres con más de un embarazo en el período 2015-2018, el embarazo con diagnóstico de diabetes y en las mujeres sin diabetes el último embarazo. En el caso de las mujeres con más de un embarazo con diabetes en el período 2015-2018 se consideró el último.

– La probabilidad de captura de todas las mujeres era la misma dadas las características de la HCE⁷.

– Las mujeres capturadas en las dos ocasiones podían ser identificadas, es decir, las marcas no se perdían, eran las mismas dadas las características de la HCE⁷.

Luego, para estimar la cantidad de embarazadas con diabetes no detectadas por ninguna de las cinco fuentes consideradas, se usó un modelo de regresión de Poisson. Todos los modelos utilizaron todas las fuentes disponibles. El modelo reducido sin interacciones representaba la independencia entre las fuentes, mientras que el agregado al modelo de interacciones entre fuentes consideraba la dependencia entre los componentes de la interacción, situación que podría haber ocurrido en este caso, por las características de interdependencia de las fuentes.

Se generaron modelos reducidos y con distintas cantidades de interacciones significativas entre fuentes diferentes^{5,8}. Fue seleccionado el modelo con un menor valor de Akaike (AIC) y se calcularon los intervalos de confianza del 95% (IC 95%) para el número estimado de casos, a través de una aproximación normal. Se presentaron las incidencias: aquella con los numeradores observados en la cohorte y la incidencia con el número agregado mediante el valor estimado por el modelo (los casos observados más los estimados que no habían sido recuperados por ninguna fuente).

RESULTADOS

Se obtuvo un listado de 12 705 mujeres con un embarazo en el período de tiempo comprendido entre 01/06/2015 y el 31/06/2018. Se detectaron por validación manual de un total de 7711 embarazadas, 617 casos nuevos de diabetes y embarazo (incidencia 8,0% IC 95%: 7,4-8,6).

Luego de aplicar el método captura-recaptura, la celda faltante global estimada por Poisson arrojó 3,2 casos (IC 95%: 2,3-4,2). La estimación de los casos de diabetes y embarazo aumentó de 617 a 620, con un mínimo de 2,3 y un máximo de 4,2 casos, con una incidencia estimada de 8,04% (IC 95%: 7,4-8,7). En la tabla 1 se observa cómo se distribuyeron los casos en las distintas fuentes, en el tabla 2 se muestran los modelos de Poisson con interacción y

Tabla 1. Distribución de los casos en cada una de las fuentes

Laboratorio	Tratamiento	Evolución	Problema	Listado de obstetras	Ncasos
	X				2
	X	X		X	2
	X	X	X		1
X					199
X				X	112
X			X		2
X		X		X	16
X		X	X		3
X	X				114
X	X			X	26
X	X		X		1
X	X	X		X	127
X	X	X	X		11
X	X	X	X	X	1

En el cuadro se observan, en las primeras 5 columnas, las diferentes fuentes de datos que registraban por separado a los casos. Las X indican en qué fuente se observaron los casos. La sexta columna es el número de casos que se encontraban en una o más fuentes; por ejemplo, 16 casos se encontraron en las fuentes laboratorio, evolución y listado de obstetras. La celda faltante es la estimada por el modelo del método captura-recaptura.

Tabla 2. Modelos de Poisson empleados en el método captura-recaptura

A. Modelo de Poisson sin interacción						
	RTI	ES	Z	p	IC 95%	
Laboratorio	42,9	19,5	8,26	< 0,001	17,6-104,6	
Tratamiento	0,8	0,3	-0,47	0,639	0,4-1,7	
Problema HCE	0,2	0,01	-6,26	< 0,001	0,0-0,8	
Evolución	1,1	0,8	0,17	0,876	0,3-4,4	
Obstetras	0,4	0,2	-2,04	0,042	0,2-0,9	
B. Modelo de Poisson con interacción						
Laboratorio	52,6	20,9	9,99	< 0,001	24,2-114,4	
Problema HCE	0,3	0,2	-1,81	0,07	0,9-1,1	
Evolución	1,1	0,8	0,16	0,872	2,8-4,4	
Tratamiento	0,8	0,3	-0,47	0,639	0,4-1,7	
Obstetras	0,4	0,3	-2,04	0,042	0,9-0,9	
Interacción HCE-laboratorio	0,7	0,0	-4,42	< 0,001	0,20,3	

ES: error estándar; HCE: historia clínica electrónica; IC 95%: intervalo de confianza del 95%; p: p valor; RTI: razón de tasa de incidencia; z: estimado.

Tabla 3. Casos estimados en cada modelo con sus criterios de Akaike y criterios de información bayesiano

Modelo de Poisson	AIC	BIC	Casos estimados (IC 95%)
A. Sin interacción	261,16	264,26	3,96 (2,89-5,03)
B. Con interacción	257,33	261,16	3,23 (2,26-4,20)

AIC: criterio de información de Akaike; BIC: criterio de información bayesiano; IC 95%: intervalo de confianza del 95%.

sin ella y, en el tabla 3, se comparan los casos detectados por cada modelo y los criterios bayesiano y de Akaike. El modelo final seleccionado fue el B, que correspondía al modelo con interacción entre la base de laboratorio y el problema cargado en la historia clínica, porque se consideró que esas dos bases no son independientes entre sí.

El listado de pacientes obtenido fue validado de acuerdo con los criterios de inclusión (mujeres mayores de 18 años con una gesta única) y se clasificaron en grupos de diabetes mellitus pregestacional (es decir, presente antes de la gestación) y diabetes mellitus gestacional.

CONCLUSIONES

El método de captura y recaptura resultó preciso ya que, al comparar la aproximación por él con la cantidad total de mujeres con diabetes en el embarazo, el MCRP estimó un 0,5% más de casos, sin diferencias en las incidencias estimadas por el método de captura- recaptura y la validación manual.

Para que el método sea exitoso, necesariamente se deben tener en cuenta las consideraciones expuestas en el análisis. El método de captura y recaptura es una forma eficaz para calcular el tamaño poblacional.

REFERENCIAS

1. Chao A, Tsay PK, Lin SH, et al. The applications of capture-recapture models to epidemiological data. *Stat Med.* 2001;20(20):3123-3157. <https://doi.org/10.1002/sim.996>.
2. Schmid B, Silva NN. Estimation of live birth underreporting with a capture-recapture method, Sergipe, Northeastern Brazil. *Rev Saude Publica.* 2011;45(6):1088-1098. <https://doi.org/10.1590/s0034-89102011000600011>.
3. Rivest LP, Baillargeon S. Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics.* 2007;63(4):999-1006. <https://doi.org/10.1111/j.1541-0420.2007.00779.x>.
4. Baillargeon S, Rivest LP. Rcapture: loglinear models for capture-recapture in R. *J Statistical Software.* 2007;19(5):1-31. <https://doi.org/10.18637/jss.v019.i05>.
5. Forbes RB, Swingle RJ. Estimating the prevalence of multiple sclerosis in the United Kingdom by using capture-recapture methodology. *Am J Epidemiol.* 1999;149(11):1016-1024. <https://doi.org/10.1093/oxfordjournals.aje.a009746>.
6. do Valle JB, Silva JC, Oliveira DS, et al. Use of a clinical-laboratory score to guide treatment of gestational diabetes. *Int J Gynaecol Obstet.* 2018;140(1):47-52. <https://doi.org/10.1002/ijgo.12326>.
7. Centro Nacional de Terminología en Salud. SNOMED Argentina [Internet]. Buenos Aires: Ministerio de Salud; 2019 [citado 2020 sep 20]. Disponible en: <https://www.argentina.gob.ar/salud/snomed>.
8. Hook EB, Regal RR. Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. *Am J Epidemiol.* 1995;142(9 Suppl):S48-52. https://doi.org/10.1093/aje/142.supplement_9.s48.
9. Bioninja. Population sampling [Internet]. [2016-] [citado 2022 feb 4]. Disponible en: <https://ib.bioninja.com.au/options/option-c-ecology-and-conser/c5-population-ecology/population-sampling.html>.