

# Estimación de Mortalidad por departamentos: el caso de la región Pampeana (2009-2011)

Nicolás Sacco;<sup>\*</sup>Iván Williams;<sup>†</sup>Bernardo L. Queiroz<sup>‡</sup>

Agosto, 2019

## Introducción

En América Latina y el Caribe, la demanda de estimaciones epidemiológicas (y de mortalidad, específicamente) sobre heterogeneidad a nivel subnacional está creciendo, tanto como una herramienta para la aplicación de diferentes planes de desarrollo como para la asignación de recursos. La pregunta es siempre: ¿qué esconden los promedios regionales?

El problema principal es el de tratar con fenómenos con un pequeño número de experimentos, y en muchos casos desconocida cobertura. La experiencia en América Latina está liderada por Brasil, donde ya existe un desarrollo metodológico de avance sostenido (Freire (2015)).

En Argentina, se encontraron experiencias sobre estimaciones de la tendencia de mortalidad infantil (Torcida, Vega, and Velázquez (2008)) y las comunas de la Ciudad Autónoma de Buenos Aires (Grushka (2013)). Al mismo tiempo, otros estudios se centran en el desarrollo de herramientas demográficas para medir la heterogeneidad en una población (Vaupel and Missov (2013); Raalte, Sasson, and Martikainen (2018)). En este sentido, solo considerando la información de la tabla de vida, podría haber múltiples capas de análisis de desigualdad teniendo en cuenta distintos niveles espaciales.

Este trabajo como fin tiene estimar el nivel de mortalidad para áreas menores de Argentina, en este caso, de departamentos durante el período 2009-2011. Con el criterio de comenzar con la población con mayor participación en total país se seleccionó la región pampeana.

Se propone la siguiente línea de desarrollo. Se realiza un análisis inicial de calidad de datos, se repasan tres metodologías de suavizamiento, redefiniendo a su vez áreas mayores por fuera de lo estrictamente administrativo, se construyen tablas de mortalidad y se selecciona un método con el cual dar una primera aproximación al problema. En el medio se encontraron inconvenientes con la información, los cuales se trataron de destacar y dar soluciones<sup>1</sup>.

---

<sup>\*</sup>Penn State, nsacco@psu.edu

<sup>†</sup>Universidad Nacional de Luján, ivanwilliams1985@gmail.com

<sup>‡</sup>Cedeplar-UFMG, lanza@cedeplar.ufmg.br

<sup>1</sup>El material complementario, código y resultados se pueden encontrar en [https://github.com/nsacco/mortalidad\\_Argentina](https://github.com/nsacco/mortalidad_Argentina).

## Datos

Se solicitaron microdatos de defunción para los años de registro 2009, 2010 y 2011 al Departamento de Estadísticas e Información de Salud (DEIS) del Ministerio de Salud de la Nación. El registro tardío en todo el país fue de 1.05%, y dado que la idea habitual de compensación no es uniforme entre años y áreas pequeñas, se decidió procesar la base de datos y tomar los registros por año de ocurrencia (4). El porcentaje de edad desconocida fue 0.33%, y sexo desconocido 1.01%. Clasificando los eventos según provincia de residencia, la información desconocida de departamento por provincia se clasificó en 2, siendo Ciudad Autónoma de Buenos Aires la provincia en peor posición. De estas defunciones sin departamento de residencia conocido, el 71% se debe a muertes que ocurrieron en departamentos adyacentes o muy cercanos a la Ciudad Autónoma de Buenos Aires (CABA), en el Gran Buenos Aires (aglomerado en la provincia de Buenos Aires que es vecino de CABA): Tres de Febrero (20%), Vicente López (15%), La Matanza (12%), Avellaneda (7%), Lanús (5%), Morón (5%), San Isidro (4%) y Gral. San Martín (3%).<sup>2</sup> Además, debido al cambio en la definición de las unidades administrativas en CABA, a partir de 2011 en la base de datos no es posible matchear los tres años de riesgo que aquí se consideran; por ambas razones se decidió dejar de lado esta jurisdicción en esta primera estimación.

Los datos desconocidos en áreas pequeñas son un problema importante (3). Teniendo en cuenta solo las provincias pampeanas, Buenos Aires poseía los departamentos con el mayor porcentaje de edad y sexo desconocidos, con valores más grandes en sexo, siendo el líder el departamento General Pueyrredón con 7.3%. Debido a esto, adicionalmente no trataremos la desagregación por sexo. Las categorías desconocidas (en variables edad, provincia de residencia y departamento de residencia) se distribuyeron proporcionalmente debido a su peso menor.

Para la población expuesta al riesgo de cada departamento se utilizó la población estimada por el INDEC a mediados de año en 2010, y se aplicó la estructura observada en el censo 2010 (INDEC (2015)). Luego, en lugar de promediar los tres años de riesgo, se tuvo en cuenta los años-persona en que las personas hubiesen vivido en el período de tres años entre 2009 y 2011 y la fecha censal (M. R. Gonzaga and Schmertmann (2016)). Este procedimiento permitió, primero, suavizar un poco la mala declaración de edad, que podría aportar un mayor sesgo para la comparación de tasas cuando el recuento de muertes no sigue este patrón por edad; y segundo aprovechar las correcciones de omisión hechas por INDEC en el total<sup>3</sup>. Para eso se supone una distribución uniforme de la fecha de nacimiento dentro del año y que la población sea cerrada. Se utilizó una función de supervivencia única para todos los distritos, aplicando las mismas tablas de vida estándar que los autores (una media representativa de la Base de Datos de Human Mortality Database en años posteriores a 1969) pero en nuestro caso ponderada por un índice de masculinidad de 1.04 al nacer, para obtener ambos sexos. Se puede lograr una suavización similar, pero con menos interpretación demográfica, con una regresión local (procedimiento *loess* en el software R) de 3 veces el recuento censal de

---

<sup>2</sup>Camisa (2019) notó este sesgo hace muchos años: 10% de los nacimientos y muertes durante el período 1946-1948.

<sup>3</sup>No está claro la metodología aplicada para la población ajustada a mitad de año en los departamentos, y si tiene en cuenta una corrección de “residencia” (INDEC (2015))

población (James et al. (2014)) (ver 1).

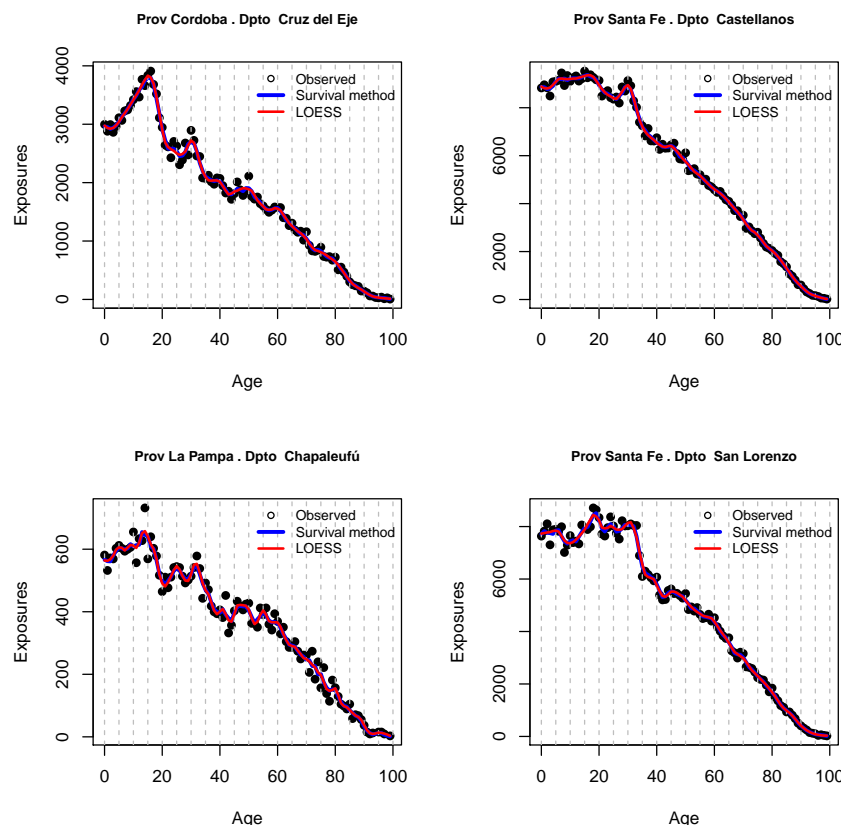


Figure 1: Ajuste de exposición 2008-2010. Cuatro casos. Fuente: elaboración propia en base a INDEC (2013)

## Algunos chequeos de consistencia

Existen cada vez más intentos por abordar el problema de cobertura de muertes en pequeñas áreas, dependiendo de la información auxiliar con que se cuente (S. Preston et al. (1980); Bennett and Horiuchi (1984); C. P. Schmertmann and Gonzaga (2018); Alexander, Zagheni, and Barbieri (2017)). La utilización de métodos demográficos indirectos de evaluación de cobertura es difícil de mantener en una población pequeña muy influenciada por la migración interna debido a su baja exposición.

Con el objetivo de visualizar posibles problemas de datos de calidad en los departamentos, se realizaron dos ejercicios.

Primero, se utilizó el método de Brass y Coale para la estimación indirecta de la mortalidad infantil y se mapeó con  $q_0$  a partir de los datos brutos obtenidos en la última sección (Moultrie et al. (2013)). Este no es un método preciso para poblaciones pequeñas, pero puede dar una idea sobre los problemas en las áreas más grandes (problemas relativos al numerador o

denominador). Mostramos los puntos no ponderados y ponderados poblacionalmente para dar mayor importancia a la consistencia en las áreas más pobladas, responsables del posible sesgo de suavizamiento en los procedimientos metodológicos descritos en la siguiente sección (Figura 2). Utilizando la edad promedio de la madre al nacimiento durante 2010 para cada provincia y la familia de tablas de la ONU de América Latina, se concluye que observando los puntos de mortalidad infantil no hay un sesgo claro en las áreas más grandes.

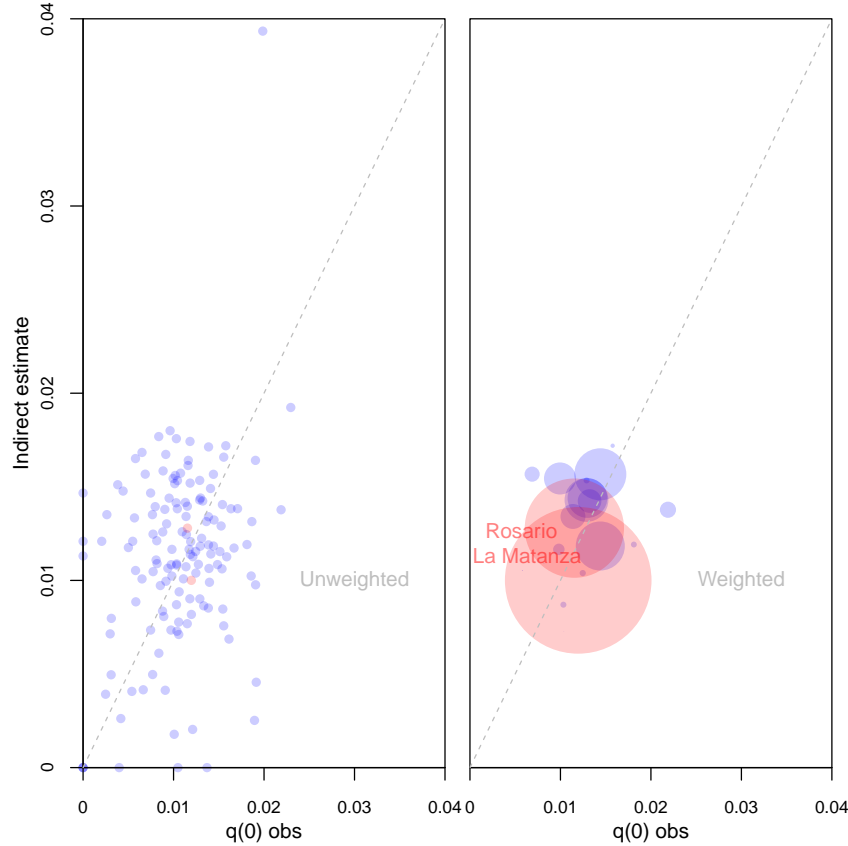


Figure 2: Estimaciones indirectas de  $q(0)$  y tasa de mortalidad  $m(0)$  observada. Departamentos de la región Pampeana (excepto CABA). Fuente: elaboración propia en base a Cesno y estadísticas vitales.

En segundo lugar, se mapeó cada departamento respecto al indicador censal de pobreza Necesidades Básicas Insatisfechas (NBI) y la tasa de mortalidad general estandarizada a partir de la estructura por edad regional, buscando una relación esperada (Kaztman (1995), S. H. Preston (1975); Grushka (2013)). En la figura 3 se muestran el NBI3 y el NBI4, que miden el porcentaje de hogares con ausencia escolar de niños y la incapacidad de subsistencia (Kaztman (1995)). En color se destaca el departamento más grande de la Región Pampeana, llamado La Matanza (que contiene al 10.7% de la provincia de Buenos Aires). El departamento de La Matanza tiene una de las tasas de mortalidad estandarizadas más bajas, pero un índice de pobreza similar (NBI3) o mayor (NBI4) al de otros. Si bien su desempeño en los índices de pobreza NBI1 (vivienda inconveniente), NBI2 (carencias sanitarias) y NBI3

(hacinamiento) no es tan llamativo como los señalados, se decidió dejarlo fuera de este estudio debido principalmente a cuestiones metodológicas, ya que las áreas más grandes son de extrema relevancia a la hora de suavizar las pequeñas y esto podría sesgar los resultados.<sup>4</sup>

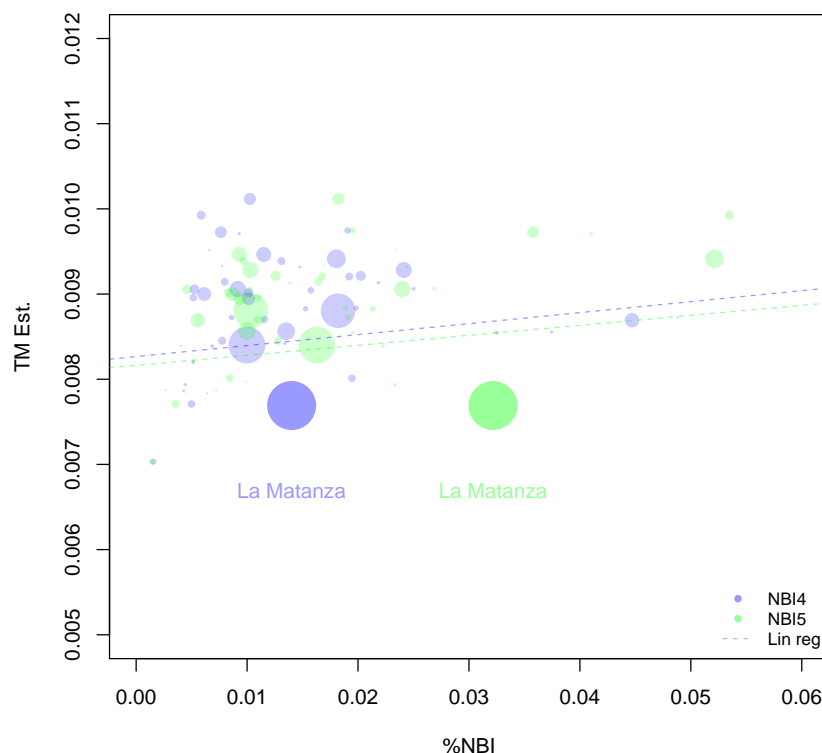


Figure 3: Standardized mortality rate and NBI. Departments in selected provinces. Source: own based in Census and DEIS

Se realizó un chequeo adicional de  $e_0$  para cada provincia a partir de los insumos anteriores y la estimación oficial (INDEC (2013))<sup>5</sup>. Los resultados dan una diferencia relativa (%) de 0,002, 0.39, 0.3, 0.98, 0.34 para las provincias Buenos Aires, Córdoba, Entre Ríos, La Pampa, Santa Fé, considerado como aceptable la aproximación dado el año de distancia en la referencia temporal y que en este ejercicio no se realizan ajustes de cobertura en la defunciones que pudieron haber sido realizados en las oficiales (tabla 5).

<sup>4</sup>INDEC advierte en su web sobre los recuentos de población por departamento en el Censo 2010, donde al parecer Buenos Aires fue una de las provincias con dificultades (<https://bit.ly/2R7svfX>, visitado el 10/1/2020)

<sup>5</sup>Desde aquí se considera a Buenos Aires sin La Matanza)

## Metodología

Debido a la poca experiencia de estudios previos que estimen mortalidad *general* en áreas menores en Argentina, se decidió aplicar tres técnicas: una basada en la teoría bayesiana, la segunda basada en métodos relacionales de tablas de vida pero agregando técnicas estadísticas de suavizado, y tercero, un enfoque demográfico clásico, considerado el método por default debido a su simplicidad. Antes de la estimación, se realizó un procedimiento de regionalización para aprovechar la similitud espacial entre áreas pequeñas, independientemente de su pertenencia político-administrativa.

## Regionalización

La definición de una región o cluster debe explorar la similitud interna entre áreas pequeñas para poder suponer que su mortalidad es la realización de un proceso estocástico mayor. La similitud en los patrones de mortalidad suele abordarse por pertenecer a la misma provincia, donde la “distancia” entre jurisdicciones no se mide por la distancia geográfica o los atributos socioeconómicos (Data and Estimation (2019)).

Para ello, tomamos el enfoque propuesto por Assuncao et al. (2006), que trató de definir áreas grandes internamente homogéneas y con condición de contigüidad en el espacio. Primero se realizó un gráfico de conectividad entre los centroides y luego se calculó el *costo* entre ellos (distancia euclidiana en nuestro caso). Luego, un procedimiento de iteración estima el árbol de expansión mínimo, que es el árbol conectado con un costo mínimo, medido como la suma de las diferencias en todos los bordes. Finalmente, se realizó un procedimiento de partición cortando el borde que minimiza la varianza dentro de los dos grupos resultantes. Debido a que probar todas las combinaciones posibles en cada partición es un problema computacional, los autores propusieron un enfoque heurístico. Una sobreclusterización aumentaría la homogeneidad pero también aumentaría la varianza en las unidades más pequeñas debido a que no hay suficientes casos. Esa es la razón para establecer umbrales mínimos de población o áreas menores resultantes en cada pareja mayor, siendo de 20 departamentos el elegido en este caso.

Como insumos para el procedimiento de regionalización, utilizamos los archivos *shape* disponibles en línea y el índice de NBI del censo, ambos por departamento<sup>6</sup>. Se re-escaló el índice a unidades de desviación estándar y aplicó la metodología comentada anteriormente, implementada en el paquete *spdep*, mediante la función *skater* (Bivand (2019)).

Con esta segmentación se obtuvo un aumento de 14% en la varianza entre grupos y una disminución no tan importante de 1% en la varianza promedio dentro los grupo. El nuevo clúster es más distinto entre sus partes y con menos algo menos de variación relativa interna (ver 4).

---

<sup>6</sup><https://bit.ly/2sVpK9u>, visitado el 10/01/2020.

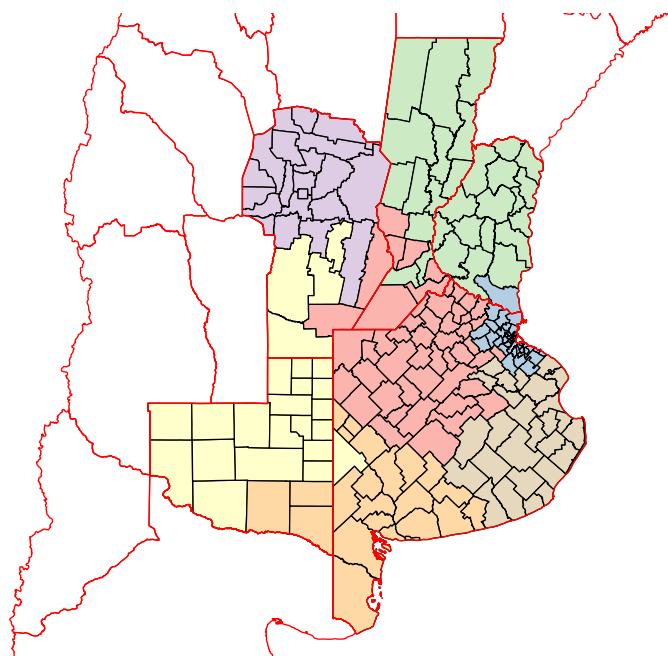


Figure 4: Regionalización de departamentos. Source: Elaboración propia en base a Censo y Estadísticas Vitales

## Métodos de estimación

Aunque Argentina generalmente se clasifica como un país con buenas estadísticas registrales de muerte (Jaspers and Orellana (1994)), se conoce que hay un porcentaje desigual de muertes infantiles no registradas por provincia (DEIS (2016)). En este sentido, dado que la fuente de datos es un registro, y pensando en la estimación de las tasas de mortalidad por edad, se podría concluir que no habría varianza, y que el sesgo (ambos componentes del error cuadrado medio de un estimador) vendría dado por el patrón de casos omitidos en cada jurisdicción. Como se mencionó anteriormente, este segundo componente del error no se abordará en este trabajo debido a que no hay información sobre su distribución por áreas pequeñas. Con respecto al primero, pese a lo mencioando, existen fenómenos con un pequeño número de “experimentos” (pocos expuestos en nuestro caso), que tienen una mayor variación en sus estimaciones, por lo que requiere un tratamiento especial para reflejar el riesgo de mortalidad subyacente (Brillinger (1986)). Para lograrlo, utilizaremos y compararemos (aunque sin conclusiones finales sobre su performance comaparada) tres métodos diferentes, los cuales se diferencian en la forma en que las áreas menores *toman prestada* información del área mayor que las contiene.

El método empírico bayesiano mejora la eficiencia estadística de los estimadores de las tasas de mortalidad por edad, disminuyendo la varianza en los casos de jurisdicciones pequeñas (Efron and Morris (1972); Marshall (1991); Data and Estimation (2019); Assunção et al. (2005)). La idea es que, suponiendo que las diferentes observaciones de cada área procedan de una distribución a priori común, cada estimación se puede mejorar utilizando la información de las otras. La distribución a priori corresponde a la distribución conjunta del vector de tasas de mortalidad por edad del área mayor. Luego, a través del comportamiento observado en cada área menor, se produce el ajuste bayesiano de la distribución de mortalidad a posteriori. La característica de “empírico” radica en que las distribuciones de los parámetros del área mayor se estiman también a partir de los datos observados, en este caso por el método de los momentos.

Veamos primero el caso univariable. Considerando un grupo de edad de cinco años, ya sea en un área  $i$ , se supone que la distribución de muertes  $d$  es un proceso de Poisson, con una media esperada de  $E(d_{x,4}^i | m_{x,4}^i) = N_{x,4}^i m_{x,4}^i$ , siendo  $N$  las exposiciones y  $m$  la tasa de mortalidad.

Primero se considera a  $\hat{m}_{x,4}^i = D_{x,4}^i / N_{x,4}^i$  como el estimador de máxima probabilidad de la tasa  $m_{x,4}^i$  en el área  $i$ , que son *iid* generados a partir de  $m_{x,4}$ . La esperanza condicionada de  $\hat{m}_{x,4}^i$  es  $E_m(E(\hat{m}_{x,4}^i) / m_{x,4}^i) = E_m(m_{x,4}^i) = m_{x,4}$  (tasa de área grande) y la varianza condicionada  $V(\hat{m}_{x,4}^i / m_{x,4}^i) = \frac{m_{x,4}^i}{N_{x,4}^i}$ .

La varianza total del estimador se puede expresar como la suma de la varianza de las medias de  $i$ 's y la esperanza de las varianzas de  $i$ 's:  $V_m(E(\hat{m}_{x,4}^i / m_{x,4}^i)) + E_m(V(\hat{m}_{x,4}^i / m_{x,4}^i)) = V_m(\hat{m}_{x,4}^i) + E_m(\frac{\hat{m}_{x,4}^i}{N_{x,4}^i}) = V_m(m_{x,4}^i) + \frac{m_{x,4}}{N_{x,4}^i}$ . Eso está relacionado con la relación jerárquica entre el hiperparámetro ( $m_{x,4}$ ), los parámetros ( $m_{x,4}^i$ ) y sus estimadores  $\hat{m}_{x,4}^i$ .

El estimador lineal bayesiano  $\hat{m}_{x,4}^i$  que minimiza el error cuadrático medio de  $m_{x,4}^i$  (e indicadores que son funciones lineales de esto) es (Robbins, 1983):

$$\hat{m}_{x,4}^i = \hat{m}_{x,4}^i + S_{x,4}^i (\bar{m}_{x,4}^i - \hat{m}_{x,4}^i)$$



Nuevamente, es empírico porque  $m_{x,4}$  se estima por método de momentos con  $\bar{m}_{x,4}$ , la media ponderada de áreas pequeñas. El factor de “contracción”  $S_{x,4}^i$  (“shrinkage” en la bibliografía) es la relación entre la expectativa de la varianza estimada en el área pequeña  $i$  y la varianza no condicionada del estimador, que es:

$$S_{x,4}^i = \frac{V_m(m_{x,4}^i)}{V_m(m_{x,4}^i) + \frac{m_{x,4}^i}{N_{x,4}^i}}$$

Visto de otra manera, esta fórmula representa la relación entre la varianza del área más pequeña con respecto a la suma de la varianza total (del área más pequeña y más grande), en sintonía con un análisis de la varianza clásica entre grupos (ANOVA). Siguiendo este razonamiento, en un contexto de extrema homogeneidad, un área menor muy pequeña podría caracterizarse a partir de la estimación del área más grande ( $S_{x,4}^i \cong 1$ ). Por otro lado, las áreas de alto peso poblacional tomarán valores cercanos a los observados ( $S_{x,4}^i \cong 0$ ). En el medio de estos extremos, la función combina linealmente la estimación del área grande con respecto al área más pequeña incluida.

Longford (1999) extendió esta idea a vectores de variables aleatorias (“contracción multi-variada”), estimando  $S_{x,4}^i$  de manera de aprovechar la correlación entre subpoblaciones. En nuestro caso, si la tasa de mortalidad del grupo de edad entre  $x$  y  $x + 4$  del área  $i$  es mayor que el área  $j$ , una correlación alta implicaría que en edades contiguas ocurriría lo mismo con mayor probabilidad. Si la covarianza fuera nula, este enfoque sería equivalente al caso univariante descrito anteriormente. Los cálculos realizados en este trabajo se realizaron para edades de 0, 1-4 y quinquenales hasta el grupo de edad abierta 80. El desarrollo se realizó siguiendo el enfoque mostrado en Assunção et al. (2005) (páginas 543 y 544), que estimó los parámetros por el método de momentos para las tasas de fecundidad en Brasil.

El otro método que se aplicará se basa en un modelo de mortalidad relacional llamado TOPALS (Tool for Projecting Age-Specific rates using Linear Splines) (Beer (2011)), que utiliza un método spline lineal para describir los ratios entre las probabilidades de muerte por edad de una población dada y un patrón. Una ventaja contra el enfoque logit clásico de Brass es que TOPALS es menos sensible al elegir el estándar. M. R. Gonzaga and Schmertmann (2016) incluyó esta idea en una regresión de Poisson en las tasas de mortalidad por edad simple, permitiendo intervalos de confianza para los resultados que tienen en cuenta la varianza por razones de baja exposición.

Específicamente, el vector de tasas de mortalidad en el área pequeña  $m^i(\alpha) = m^* * \exp^{\alpha B_x}$  es una función de los “nodos” spline  $\alpha$ , que son las edades en las que se evaluará el desvío respecto al patrón estándar, siendo  $m^*$  el vector de tasa de mortalidad estándar, y  $B_x$  es la matriz B-spline que multiplicada por  $\alpha$  brinda el desplazamiento lineal entre el logaritmo de ambas tasas.

La idea es suponer que  $D_x \sim Poi(m_x N_x)$  en cada área pequeña, construir la función de probabilidad usando las muertes y exposiciones observadas  $\log(L(m_x N_x | D_x)) = \sum_{\forall x} [-m_x N_x + D_x \ln(m_x) + D_x \ln(N_x) - \ln(D_x!)]$ , pero expresando eso en función del parámetro  $\alpha$ , agregando una penalización por distancia desde el estándar y suavizando entre edades adyacentes. Se colocan nodos en edades más determinantes en el perfil de mortalidad, para luego minimizar el log-likelihood:  $Q(\alpha) = \sum_{\forall x} [-m(\alpha)_x N_x + D_x \ln(m(\alpha)_x)] - \sum_{k=0}^5 (\alpha_k - \alpha_{k+1})^2$ .

El método empírico bayesiano es particularmente apropiado en casos con “pequeñas muestras locales, variaciones regionales sustanciales, fuertes correlaciones entre los componentes y relaciones espaciales conocidas” (Assunção et al. (2005)). En el caso de la regresión TOPALS, la aplicación es más una técnica de suavización que un modelo de variabilidad espacial, por lo que se necesitan menos supuestos sobre las relaciones entre áreas.

Finalmente, como método básico, se aplicará el método de estandarización indirecta, quizás uno de los primeros enfoques para este problema (Arriaga (2011)). Se basa en cambiar solo el nivel del área principal para replicar las muertes del área menor que se está estimando. Es el caso en el que no se tiene en cuenta ninguna información sobre la forma de la mortalidad por edad del área menor.

## Resultados

Las estimaciones se hicieron para grupos quinquenales de edad (excepto el primer grupo, separado en 0 y 1-4) con 90 y más como grupo abierto final, en todos los departamentos de la Región Pampeana (excepto La Matanza y aquellos en CABA, por motivos ya expuestos). El método TOPALS fue pensado para aplicar en edades simples, pero en este caso, debido a que no se corrigió la omisión en áreas pequeñas y para ser comparable con los demás métodos, se aplicó a edades quinquenales tomando nodos en los grupos 0, 5-9, 20-24, 40-44 y 60-64. Se muestran cuatro ejemplos de ajuste en el siguiente gráfico (ver 5).

Se considera  $e_0$  como la medida resumen de la mortalidad con motivos de comparación de nivel, considerando también que el efecto de los problemas en la edad adulta (de mayor incertidumbre en áreas pequeñas) tienen poco efecto en las estimaciones sobre  $e_0$  en poblaciones donde la mortalidad infantil y juvenil aún tienen un peso importante en el indicador.<sup>7</sup>

La correlación entre los métodos es clara: existe una gran similitud entre TOPALS y la estimación indirecta (0.97), pero menor en Bayes Empírico contra la estimación indirecta (1) y TOPALS (1) (figura 6).

Las principales diferencias se deben a que el método bayesiano tiende a tomar *siempre* alguna información sobre el patrón de edad, suponiendo una correlación entre edades contiguas dado el comportamiento global del área mayor. En los dos restantes métodos, el patrón estándar es de mayor fuerza gravitatoria. Los departamentos donde se reportan las mayores diferencias son aquellos con pocas celdas distintas de cero (ver 7).

Por otro lado, se siguió el procedimiento empleado por M. R. Gonzaga and Schmertmann (2016), relacionando las defunciones del área mayor con las obtenidas mediante la agregación de las estimaciones para las áreas menores (8). Si bien los promedios son para los métodos Indirecto, Bayesiano y Topals, se destaca un poco más de variabilidad en el segundo y tercero, y un patrón singular en el primero que sugiere un sesgo específico.

En el siguiente gráfico se consideraron los resultados del método de Bayes Empírico debido a presentar un escenario más conservador en el rango de estimaciones (futuras líneas de

---

<sup>7</sup>En términos matemáticos,  $\frac{de_0}{d\mu_{80+}} = f(T_{80+})$  están cerca de cero dependiendo de la no rectangularidad de  $l_x$  y la corrección o cambio en la tasa. En otros términos, un cambio en estas tasas, por ejemplo debido a la corrección de sesgo, se pondera con  $l_{80}$ :  $\frac{de_0}{d\epsilon} = \int_0^{\inf} \frac{d\mu_x}{d\epsilon} e_x l_x dx$  (Wrycza and Baudisch (2012)).

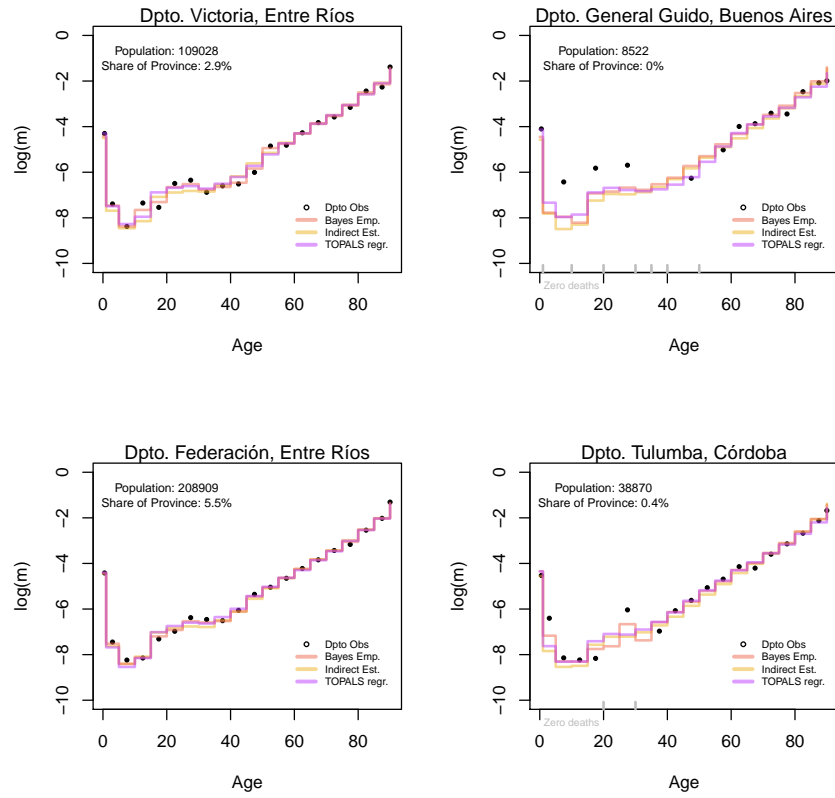


Figure 5: Mortality Estimates of Departments. Different methods. Source: own based in Census and DEIS

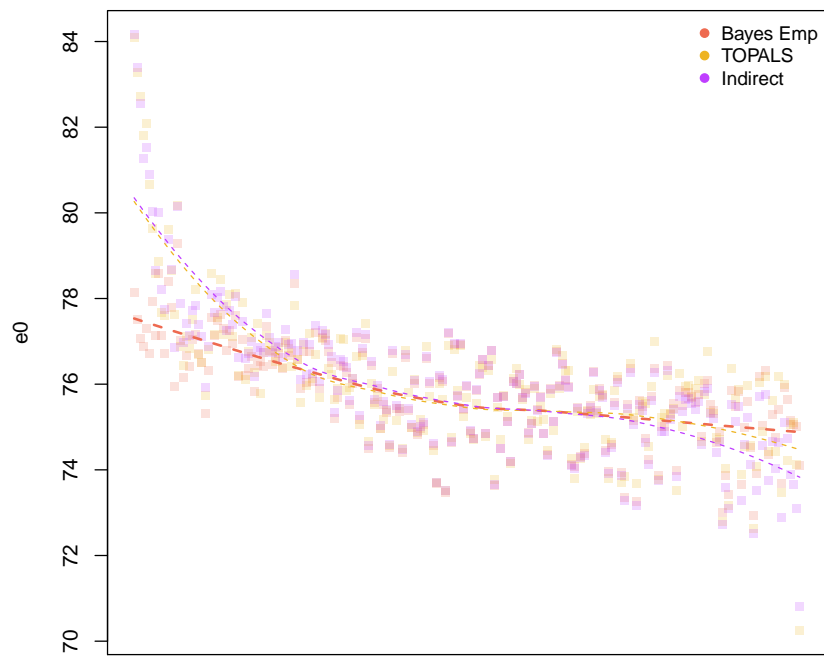


Figure 6: Estimaciones de esperanza de vida al nacer según tres metodologías. Source: elaboración propia

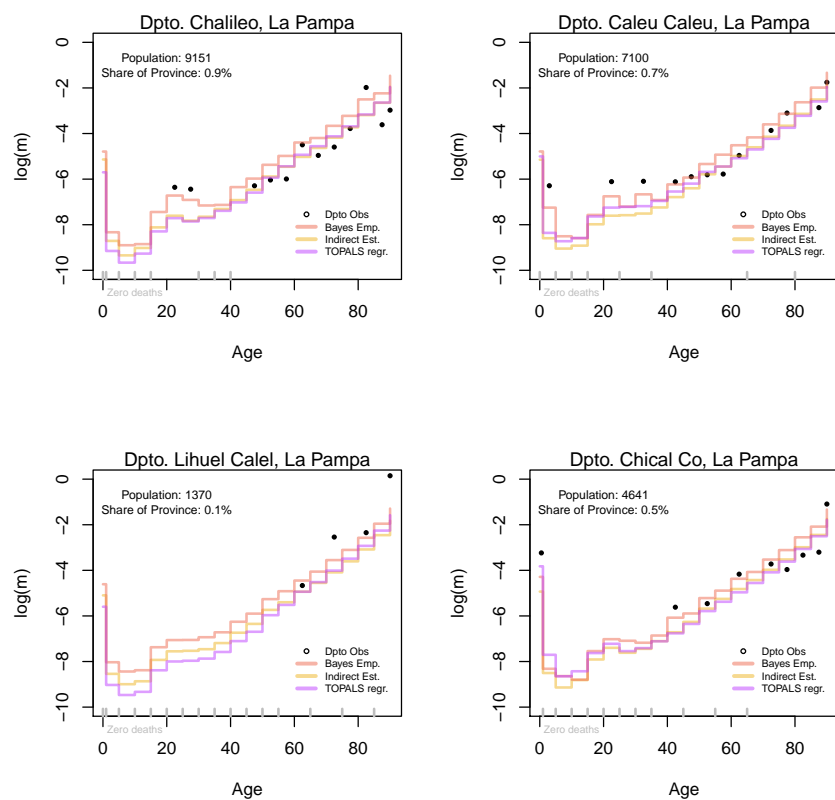


Figure 7: Estimaciones de mortalidad de los departamentos con mayores diferencias entre métodos. Fuente: elaboración propia

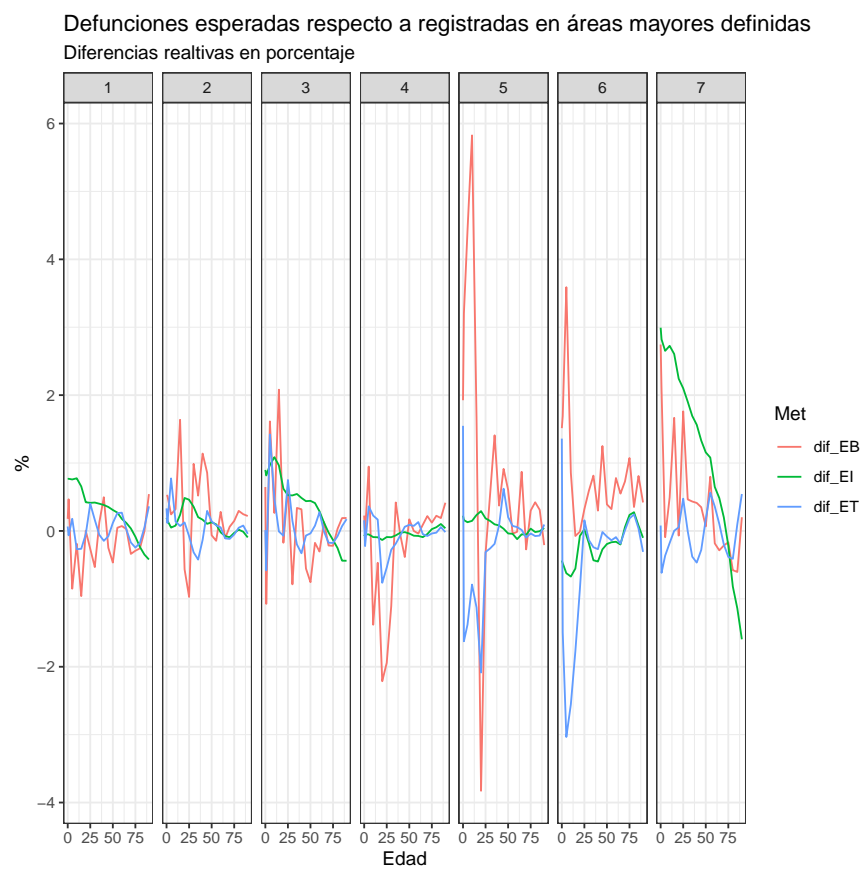


Figure 8: Diferencias relativas en la cantidad de defunciones por edad del área mayor. Fuente: elaboración propia

Table 1: Resumen de estimaciones por provincia. Esperanza de vida al nacer

Provincia	Promedio	n	Varianza	Rango
Buenos Aires	75.7	133	1.4	6.3
Córdoba	76.1	26	0.6	3.5
Entre Ríos	75.1	17	0.5	2.2
La Pampa	76.9	22	0.4	2.7
Santa Fé	75.4	19	1.1	3.4

<sup>a</sup> El promedio y varianza no están ponderados

<sup>b</sup> Fuente: elaboración propia

investigación deberían utilizar técnicas de simulación para llegar a conclusiones más sólidas, como se menciona al final del trabajo). Para tener en cuenta la aleatoriedad, se realizó un proceso bootstrap de los recuentos de muertes a partir de una distribución de Poisson del conteo de defunciones en cada grupo etario. Esto permitió contar percentiles de las funciones de la tabla de vida y específicamente de la esperanza de vida al nacer (ver 9 y 6, donde se muestran las estimaciones en el intervalo 95%).

Buenos Aires posee mayor cantidad de departamentos y a la vez una mayor dispersión, medida por el rango (distancia entre máximo y mínimo) y la varianza de los promedios (no ponderada). Conocida es su particular división entre los partidos del Gran Buenos Aires, que forman parte del aglomerado urbano más grande del país con la Ciudad Autónoma de Buenos Aires, y el resto de la provincia.

### Un análisis particular: Buenos Aires

¿Qué nos permiten decir las estimaciones? Buenos Aires es la provincia más poblada de Argentina, conteniendo 134 áreas administrativas. En el gráfico anterior se mostró su gran heterogeneidad, con un rango estimado de esperanza de vida al nacer de más de 6 años. La provincia se puede dividir entre el área del Gran Buenos Aires compuesta por 24 departamentos (área urbana que rodea a CABA) y el resto de la superficie. Para inspeccionar el significado de los resultados, podemos tomar tres jurisdicciones con exposición significativa y ubicadas a lo largo de la distribución: San Isidro, General Pueyrredón y Moreno (ver figura 10).

Si comparamos la mediana de las tasas por edad, Moreno pareció presentar una mayor mortalidad infantil pero también un mayor riesgo en adultos mayores. A priori, no hay ninguna razón para creer en una exposición con omisión diferencial en mayores de 40 años de edad, por lo que probablemente este sea un patrón de mortalidad a tener en cuenta. En el caso de San Isidro, con la mayor esperanza de vida al nacer de este grupo, presentó la curva más baja en el rango de edad típico de causas externas. Finalmente, Gral. Pueyrredón tuvo la peor posición en el rango de edad de 5 a 25 años, edades con un porcentaje importante de mortalidad por causas prevenibles. En términos estadísticos, dado el modelo empleado, los rangos de edad donde se pueden ensayar comparaciones jerárquicas entre las tres jurisdicciones tomadas son aquellos donde las áreas no se solapan: infantil y adulta mayor a 45, y

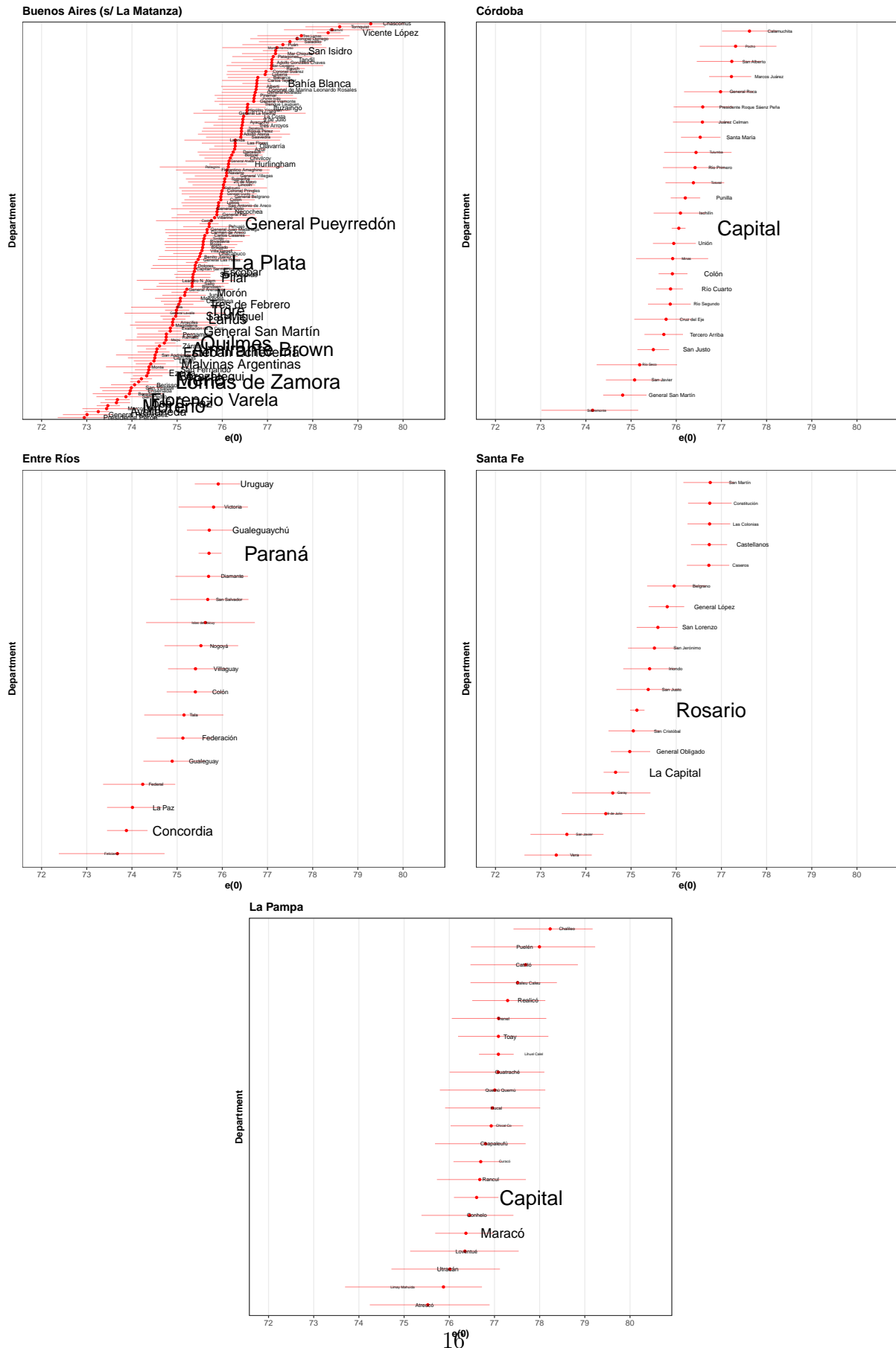


Figure 9: Esperanza de vida al nacer para las provincias pampeanas (excepto CABA, y Buenos Aires sin La Matanza). Ordenado según la media y con intervalos de confianza. El



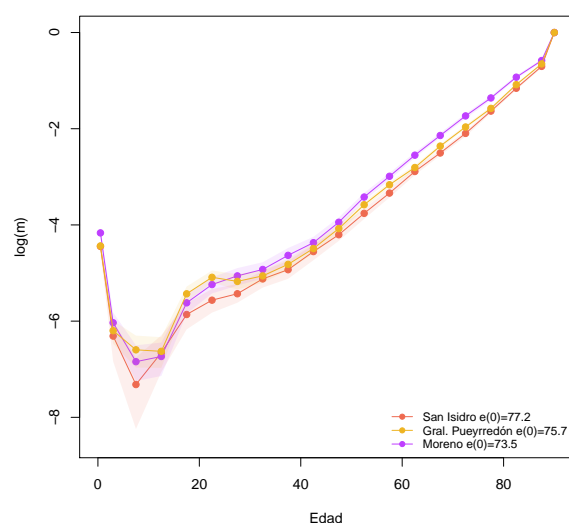


Figure 10: Mortality rates for selected areas in Buenos Aires. Source: own based in Census and DEIS

15-24 entre San Isidro y Gral. Pueyrredón.

## Limitaciones y Trabajo futuro

Se desconoce el nivel de cobertura de las áreas menores. Se realizó un análisis de datos desconocidos en el registro de defunciones y algunas comprobaciones visuales sobre la consistencia entre un indicador censal socioeconómico (NBI) y estimaciones indirectas de mortalidad infantil con el fin de detectar posibles anomalías, pero solo enfocándose en los departamentos de gran volumen debido a las propiedades estadísticas de los métodos. El costo fue grande: se dejó fuera de la estimación al departamento más grande del país, a la Ciudad Autónoma de Buenos Aires, y no se desagregó por sexo el ejercicio. Correcciones sobre los datos deben realizarse a partir de información externa si se deciden incorporar en futuras investigaciones, al menos en el período considerado.

Próximos pasos se desprenden necesariamente. Para realizar una comparación metodológica robusta podrían simularse perfiles de mortalidad según tablas modelo, en diferentes escalas y patrones de omisión, incorporando otros desarrollos recientes al análisis (por ejemplo Alexander, Zagheni, and Barbieri (2017)). Por otro lado, la desigualdad espacial puede estudiarse en capas, como una mamushka, una mortalidad fractal mandelbrotiana, con relaciones jerárquicas que pueden mostrar patrones de nivel y dispersión. La búsqueda de los mismos para Argentina y América Latina puede aportar puntos de vista interesantes.

## Conclusiones

La investigación demográfica, como otras ciencias sociales, está limitada por la fuente de información disponible y su calidad. En general, cuando esta no es de buena o simplemente es muy costosa de recopilar, el campo de la demografía se caracterizó por haber construido una rica historia en la aplicación de métodos indirectos para comenzar a acercarse a los valores reales de los fenómenos. El trabajo presente va en esta dirección. Los antecedentes de estimación en áreas pequeñas son escasos en Argentina. Decidimos comenzar con la Región Pampeana debido a su participación poblacional del país. Aplicamos tres métodos para estimar la estructura y el nivel de mortalidad, y realizamos verificaciones de consistencia previas para descartar grandes problemas. Las principales diferencias entre los métodos se deben a que el método bayesiano empírico tiende a tomar siempre alguna información sobre el patrón de edad del área menor, pero a su vez con un poco de menor consistencia en lo agregado. Se ensayó un análisis comparativo para el caso de Buenos Aires, caracterizando 3 departamentos y cuantificando diferentes perfiles de mortalidad, aunque con reparos estadísticos.

En la búsqueda de información sobre heterogeneidad intraprovincial en la mortalidad hay decisiones que tomar en el numerador y denominador de las tasas por edad. Hay un límite en lo que podemos decir, pero podemos empezar a decir algo que guíe la atención hacia aquellas áreas en desventaja. Este puede ser un punto de partida para dar prioridades en el diseño de políticas de salud a nivel local y dar pie a futuras investigaciones en el ámbito demográfico argentino que avancen tanto en la calidad de datos como metodológico en áreas menores.

## References

### Tables

Alexander, Monica, Emilio Zagheni, and Magali Barbieri. 2017. “A Flexible Bayesian Model for Estimating Subnational Mortality.” *Demography* 54 (6). NIH Public Access: 2025. doi:10.1007/s13524-017-0618-7.

Arriaga, E. 2011. *Análisis Demográfico de La Mortalidad*. Universidad Nacional de Córdoba.

Assuncao, R. M., M. C. Neves, G. Camara, and C. Da Costa Freitas. 2006. “Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees.” *International Journal of Geographical Information Science* 20 (7). Taylor & Francis: 797–811. doi:10.1080/13658810600665111.

Assunção, Renato M., Carl P. Schmertmann, Joseph E. Potter, and Suzana M. Cavenaghi. 2005. “Empirical Bayes Estimation of Demographic Schedules for Small Areas.” *Demography* 42 (3). Springer: 537–58. <http://www.jstor.org/stable/4147361>.

Beer, Joop de. 2011. “A new relational method for smoothing and projecting age-specific fertility rates: TOPALS.” *Demographic Research* 24 (March). Demographic Research: 409–54. <https://www.demographic-research.org/volumes/vol24/18/default.htm>.

Bennett, Neil G, and Shiro Horiuchi. 1984. “Mortality Estimation from Registered Deaths

Table 2: Provinces with unknown residence department

Province	Unknown %
Ciudad Autónoma de Buenos Aires	8.6
Buenos Aires	0.9
Catamarca	0.7
Córdoba	0.3
Corrientes	1.0
Chaco	0.8
Chubut	1.5
Entre Ríos	0.7
Formosa	0.9
Jujuy	3.1
La Pampa	1.6
La Rioja	0.7
Mendoza	0.4
Misiones	0.8
Neuquén	0.7
Río Negro	1.7
Salta	0.6
San Juan	0.7
San Luis	1.4
Santa Cruz	2.4
Santa Fe	0.4
Santiago del Estero	1.2
Tucumán	1.7
Tierra del Fuego, Antártida e Islas del Atlántico Sur	2.5

<sup>a</sup> Source: Own calculations based in DEIS registers

Table 3: Dptos sin dato

Prov_Nombre	Dpto_Nombre	PorcEdad	Prov_Nombre	Dpto_Nombre	PorcSexo
Buenos Aires	General Alvear	2.2	Buenos Aires	General Pueyrredón	7.3
Buenos Aires	Leandro N. Alem	1.9	Buenos Aires	Vicente López	5.6
Buenos Aires	General La Madrid	1.8	Buenos Aires	Quilmes	3.8
Buenos Aires	General Pinto	1.6	Buenos Aires	Coronel Dorrego	3.7
Buenos Aires	Las Flores	1.6	Buenos Aires	Ituzaingó	3.1
Buenos Aires	Maipú	1.6	Buenos Aires	San Andrés de Giles	2.5
Buenos Aires	Florentino Ameghino	1.4	Buenos Aires	Bahía Blanca	2.4
Buenos Aires	Salliqueló	1.4	Buenos Aires	General San Martín	2.3
Buenos Aires	Castelli	1.2	Buenos Aires	San Miguel	2.2
Buenos Aires	Pellegrini	1.2	Buenos Aires	La Plata	2.1

<sup>a</sup> Source: own calculations based in DEIS registers

Table 4: Distribution of Registered and Ocurrred deaths. Argentina

	Ocurrred year					
	2008	2009	2010	2011	2012	2013
2009	0.94	99.06	0.00	0.00	0.00	0.00
2010	0.00	0.79	99.21	0.00	0.00	0.00
2011	0.00	0.00	1.12	98.88	0.00	0.00
2012	0.00	0.00	0.00	1.13	98.87	0.00
2013	0.00	0.00	0.00	0.00	1.29	98.71

<sup>a</sup> The year register is in rows, and occurred in columns.

Source: own calculations based in DEIS registers

Table 5: Differences between life expectancy calculated with non adjusted data and official estimates

Province	Own	INDEC	Realtive Diference
Buenos Aires	75.18	75.18	0.00
Cordoba	76.05	75.75	0.39
Entre Rios	75.20	74.98	0.30
La Pampa	76.95	76.20	0.98
Santa Fe	75.36	75.10	0.34

<sup>a</sup> Source: Own calculations and INDEC (2013)

Table 6: Life expectancy at birth estimates. Cordoba 2008-2010

Province	Department	Mean	p97.5	p2.5
Córdoba	Calamuchita	77.62	78.24	77.02
Córdoba	Capital	76.06	76.21	75.90
Córdoba	Colón	75.91	76.25	75.61
Córdoba	Cruz del Eje	75.78	76.34	75.07
Córdoba	General Roca	76.98	77.71	76.17
Córdoba	General San Martín	74.81	75.34	74.38
Córdoba	Ischilín	76.09	76.65	75.50
Córdoba	Juárez Celman	76.58	77.20	75.93
Córdoba	Marcos Juárez	77.22	77.66	76.73
Córdoba	Minas	75.92	76.71	75.11
Córdoba	Pocho	77.31	78.22	76.52
Córdoba	Presidente Roque Sáenz Peña	76.59	77.26	75.94
Córdoba	Punilla	76.20	76.53	75.88
Córdoba	Río Cuarto	75.87	76.15	75.55
Córdoba	Río Primero	76.42	77.08	75.71
Córdoba	Río Seco	75.19	76.02	74.24
Córdoba	Río Segundo	75.87	76.32	75.38
Córdoba	San Alberto	77.23	77.88	76.45
Córdoba	San Javier	75.08	75.69	74.44
Córdoba	San Justo	75.49	75.84	75.14
Córdoba	Santa María	76.53	76.98	76.10
Córdoba	Sobremonte	74.15	75.16	73.02
Córdoba	Tercero Arriba	75.73	76.15	75.30
Córdoba	Totoral	76.38	77.06	75.76
Córdoba	Tulumba	76.44	77.22	75.73
Córdoba	Unión	75.94	76.43	75.48

<sup>a</sup> Source: Own calculations

- in Less Developed Countries.” *Demography* 21 (2). Springer: 217–33. doi:10.2307/2061041.
- Bivand, Roger. 2019. “Spatial Dependence: Weighting Schemes, Statistics and Models [R package spdep version 1.1-2].” Comprehensive R Archive Network (CRAN). <https://cran.r-project.org/web/packages/spdep/index.html>.
- Brillinger, D. R. 1986. “The natural variability of vital rates and associated statistics.” *PubMed. Comprises. More. Than. 29 Million. Citations. for. Biomedical. Literature. from. MEDLINE, Life. Science. Journals., and. Online. Books.* 42 (4). Wiley: 693–734. <https://www.ncbi.nlm.nih.gov/pubmed/3814721>.
- Camisa, Zulma C. 2019. “Tabla Abreviada de Mortalidad de La Region Pampeana de La Republica Argentina 1946-1948; Precedida de Un Analisis Critico de Las Estadisticas Basicas.” *Repec.org*. <https://econpapers.repec.org/paper/ecrcol048/8246.htm>.
- Data, Missing, and Small-Area Estimation. 2019. “Missing Data and Small-Area Estimation - Modern Analytical Equipment for the Survey Statistician | Nicholas T. Longford | Springer.” *Springer.com*. <https://www.springer.com/gp/book/9781852337605>.
- DEIS. 2016. Ministerio de Salud de la Nación. <http://www.deis.msal.gov.ar/wp-content/uploads/2016/09/Estadisticasvtales2016.pdf>.
- Efron, Bradley, and Carl Morris. 1972. “Empirical Bayes on Vector Observations: An Extension of Stein’s Method.” *Biometrika* 59 (2). [Oxford University Press, Biometrika Trust]: 335–47. <http://www.jstor.org/stable/2334578>.
- Freire, Queiroz, F. H. M. d. A. 2015. “Mortality Estimates and Construction of Life Tables for Small Areas in Brazil, 2010.”
- Gonzaga, Marcos Roberto, and Carl Paul Schmertmann. 2016. “Estimating Age- and Sex-Specific Mortality Rates for Small Areas with Topals Regression: An Application to Brazil in 2010.” *Revista Brasileira de Estudos de População* 33 (3): 629–52. doi:10.20947/s0102-30982016c0009.
- Grushka, Baum, C. 2013. “Vivir Y Morir En Las Comunas de La Ciudad de Buenos Aires: Un Estudio de Diferenciales.” Población de Buenos Aires.
- INDEC. 2013. “Tablas Abreviadas de Mortalidad Por Sexo Y Edad 2008-2010: Total Del País Y Provincias.”
- . 2015. “Estimaciones de Población Por Sexo, Departamento Y Año Calendario 2010-2025.”
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.
- Jaspers, D., and H. Orellana. 1994. “Evaluación Del Uso de Las Estadísticas Vitales Para Estudios de Causas de Muerte En América Latina.” *Notas de Población* 60 (CELADE): 47–77.
- Kaztman, Rubén. 1995. “La Medición de Las Necesidades Básicas Insatisfechas En Los

Censos de Población.” Centro Latinoamericano de Demografía.

Longford, N. T. 1999. “Multivariate Shrinkage Estimation of Small Area Means and Proportions.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162 (2): 227–45. doi:10.1111/1467-985X.00132.

Marshall, Roger J. 1991. “Mapping Disease and Mortality Rates Using Empirical Bayes Estimators.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 40 (2). [Wiley, Royal Statistical Society]: 283–94. <http://www.jstor.org/stable/2347593>.

Moultrie, Tom, RE Dorrington, A G Hill, K H Hill, Ian Timaeus, and Basia Zaba. 2013. *Tools for Demographic Estimation*.

Preston, Samuel H. 1975. “The Changing Relation Between Mortality and Level of Economic Development.” *Population Studies* 29 (2). [Population Investigation Committee, Taylor; Francis, Ltd.]: 231–48. doi:10.2307/2173509.

Preston, Samuel, Ansley J. Coale, James Trussell, and Maxine Weinstein. 1980. “Estimating the Completeness of Reporting of Adult Deaths in Populations That Are Approximately Stable.” *Population Index* 46 (February): 179–202. doi:10.2307/2736122.

Raalte, Alyson A. van, Isaac Sasson, and Pekka Martikainen. 2018. “The Case for Monitoring Life-Span Inequality.” *Science* 362 (6418): 1002–4. doi:10.1126/science.aau5811.

Schmertmann, Carl P., and Marcos R. Gonzaga. 2018. “Bayesian Estimation of Age-Specific Mortality and Life Expectancy for Small Areas with Defective Vital Records.” *Demography* 55 (4): 1363–88. doi:10.1007/s13524-018-0695-2.

Torcida, Sebastián, Andrea L Vega, and Guillermo A Velázquez. 2008. “Análisis de La Evolución de La Tasa de Mortalidad Infantil En Los Departamentos de Argentina. 1994-20.”

Vaupel, James W, and Trifon I Missov. 2013. “Unobserved Population Heterogeneity: A Review of Formal Relationships.” *Demographic Research* 31. Demographic Research: 659–86. <https://www.demographic-research.org/volumes/vol31/22/default.htm>.

Wrycza, Tomasz, and Annette Baudisch. 2012. “How life expectancy varies with perturbations in age-specific mortality.” *Demographic Research* 27 (September). Demographic Research: 365–76. <https://www.demographic-research.org/volumes/vol27/13/default.htm>.