Mortalidad en áreas pequeñas de la región Pampeana (Argentina, 2009-2011)

Agosto, 2019

# **Resumen**

Mientras aumenta la demanada sobre estimaciones epidemiológicas para áreas pequeñas, estudios recientes muestran una persistente brecha de desigualdad en la esperanza de vida al nacer en América Latina. La escasez de datos geoespaciales desafía la aplicación de diferentes métodos para estudiar estos diferenciales de salud. A menudo, los datos en áreas pequeñas o bien no existen o son escasos en la región. Los patrones espaciales son esenciales para comprender los resultados demográficos individuales relacionados con las características de un lugar, también como una herramienta para la aplicación de planes de desarrollo y para la asignación de recursos. La historia de la estimación en áreas pequeñas es escasa en Argentina, un claro ejemplo de un país con pocas fuentes de datos, pero también un caso muy interesante de la transición epidemiológica, que a menudo no se aborda en la literatura previa. Basado en la experiencia de nuevas aplicaciones y de acuerdo con la información auxiliar disponible, en este artículo aplicamos tres métodos diferentes para estimar y luego comparar los niveles de mortalidad en áreas pequeñas en la Región Pampeana de Argentina, durante el período 2009-2011. Se calculó la esperanza de vida al nacer en base a un enfoque bayesiano, un método de tabla de vida relacional, suavizando datos espaciales de acuerdo a patrones de áreas mayores vecinas, y un enfoque demográfico indirecto. Las estimaciones se compararon para calcular las tasas de esperanza de vida al nacer a partir de registros de defunción completos y luego se relacionaron con áreas espaciales, utilizando un análisis de regionalización simple. Las tasas estimadas indican que existe una gran variabilidad en la esperanza de vida al nacer y entre regiones, con una extensión de más de seis años en la Provincia de Buenos Aires. Además, encontramos evidencia sugestiva de que las tasas por edad presentan una mayor mortalidad infantil, pero también un mayor riesgo en los adultos mayores, en las provincias que muestran mayor heterogeneidad.

# **Resumo**

À medida que a demanda por estimativas epidemiológicas para pequenas áreas aumenta, estudos recentes mostram uma lacuna persistente na desigualdade na expectativa de vida ao nascer na América Latina. A escassez de dados geoespaciais desafia a aplicação de diferentes métodos para estudar esses diferenciais de saúde. Freqüentemente, dados em pequenas áreas não existem ou são escassos na região. Os padrões espaciais são essenciais para a compreensão dos resultados demográficos individuais relacionados às características do local, bem como uma ferramenta para implementar planos de desenvolvimento e alocação de recursos. A história da estimativa de pequenas áreas é escassa na Argentina, um exemplo claro de um país com poucas fontes de dados, mas também um caso muito interessante da transição epidemiológica, que muitas vezes não é abordada na literatura anterior. Com base na experiência de novas aplicações e de acordo com as informações auxiliares disponíveis, neste artigo, aplicamos três métodos diferentes para estimar e comparar os níveis de mortalidade em pequenas áreas da Região Pampeana da Argentina, durante o período 2009-2011. A expectativa de vida ao nascer foi calculada com base em uma abordagem bayesiana, um método de tabela de vida relacional, suavizando os dados espaciais de acordo com os padrões das áreas maiores vizinhas e, em terceiro lugar, uma abordagem demográfica indireta. As estimativas foram comparadas para calcular as taxas de expectativa de vida no nascimento, a partir de registros completos de óbito e, em seguida, relacionados a áreas espaciais, usando uma análise de regionalização simples. As taxas estimadas indicam que há grande variabilidade na expectativa de vida ao nascer e entre regiões, com uma extensão de mais de seis anos na província de Buenos Aires. Além disso, encontramos evidências sugestivas de que as taxas de idade mostram maior mortalidade infantil, mas também maior risco em adultos mais velhos, nas províncias que mostram mais heterogeneidade.

# **Abstract**

While the demand for epidemiological estimates of small areas is growing, recent studies show a wide gap and persistent inequalities in life expectancy at birth in Latin America. The lack of geospatial data challenge the application of different methods to study health differentials. Often data in small areas are non-existent or sparce in this region. Spatial patterns are essential to understand individual demographic outcomes related with the characteristics of a place, also as a tool for the application of development plans and for the allocation of resources. The history of estimation in small areas is scarce in Argentina, a clear example of a country with few data sources, but also a very interesting case of the epidemiological transition, often not address by the previous literature. Based on the experience of new applications according to the available auxiliary information, in this paper we apply three different methods to estimate and then compare mortality levels in small areas in the Pampean Region of Argentina, during the period 2009-2011. Life expectancy at birht was calculated based on a Bayesian approach, a relational life-table method, smoothing spatial data ‘borrowing strength’ from neighboring areas, and third a indirect demographic approach. Estimates are compared to calculate life expectancy at birth from full death registers and then related to spatial areas, using a simple regionalization analysis. Estimated rates indicate that there is a huge variability in life expectancy at birht withing and between regions, with a spread of up to more than sex years in the Province of Buenos Aires. Furthermore, we find suggestive evidence that the rates by age present a higher infant mortality but also a higher risk in older adults, in provinces that exibhit more heterogeneity.

# **Introducción**

¿Subsisten disparidades de mortalidad entre las áreas pequeñas? Investigaciones demográficas previas sobre mortalidad sostienen que los diferenciales geográficos serían gradualmente convergentes durante la transición epidemiológica, gracias a las políticas de salud pública y las mejoras en la calidad de vida. En la mayoría de los países de ingresos altos y medios, la esperanza de vida al nacer () aumentó a principios del siglo XX, debido al avance del conocimiento y la tecnología médica para combatir las enfermedades infecciosas y las mejoras en las condiciones de vida asociadas con el desarrollo socioeconómico.

Aunque los diferenciales socioeconómicos de mortalidad contribuyeron a mantener la hipotesis de la desigualdad persistente en la muerte durante y después de la transición epidemiológica, algunos autores esperan que los resultados de salud y las disparidades se extingan a medida que se desarrolla el cambio demográfico. Sin embargo, puntos de vista retrospectivos muestran que los diferenciales de mortalidad son persistentes y han crecido con el tiempo, en muchos casos de acuerdo a la condición socioeconómica. De todos modos, es útil tener en cuenta si estas disparidades continúan y/o surgen independientemente de escenarios iniciales auspiciosos.

La explicación teórica más persuasiva de estos temas argumenta que la relación entre mortalidad y posición socioeconómica se ha mantenido a lo largo del tiempo debido al acceso diferencial de clase social a la tecnología de la información y a la salud. Esta hipótesis de causa-efecto condujo a importantes desarrollos de investigación y contribuciones teóricas. Sin embargo, identificar las relaciones basadas en estos patrones y sus resultados demográficos presenta un serio desafío en países donde hay pocas fuentes de datos disponibles. La investigación de la mortalidad implica intrincadas patuas de interacciones sociales, demográficas y ambientales. Por esa razón, la demografía espacial surgió como una perspectiva significativa para responder preguntas demográficas. En ese sentido, los patrones espaciales son esenciales para comprender los resultados demográficos individuales relacionados con las características de un lugar.

Sobre estas cuestiones, aportes metodológicas recientes se han desarrollado para estimaciones de mortalidad en áreas pequeñas (Vaupel and Missov ([2013](#ref-Vaupel_Missov_2013)); Raalte, Sasson, and Martikainen ([2018](#ref-vanRaalte_Sasson_Martikainen_2018))). 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 continua pregunta es que tan representativos son los promedios regionales de la variabilidad sub-espacial. El problema principal para abordar esta problemática, 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](#ref-FreireEtAl2015))).

Bajo este contexto, este artículo tuvo como fin estimar el nivel de mortalidad para áreas menores de Argentina, en este caso, departamentos. Para ello, decidimos elaborar estimaciones de mortalidad que ejemplifiquen la relación entre el condición socioeconómica y salud, aplicando tres métodos diferentes para estimar y luego comparar los niveles de mortalidad en áreas pequeñas en la Región Pampeana argentina, durante el período 2009-2011.

## ¿Por qué Argetina?¿Por qué la región Pampeana?

Argentina representa un claro ejemplo de un país con pocas fuentes de datos, pero también un caso muy interesante de la transición de la mortalidad, que a menudo no se aborda en la literatura previa. En comparación con otros países latinoamericanos, el desarrollo socioeconómico temprano de Argentina, el alto grado de urbanización y la expansión de la educación formal influyeron en la reducción de la mortalidad que tuvo lugar antes que en la mayoría de los otros países de la región. Esto se dio sobre todo debido a mejoras en las condiciones de vida asociadas con el desarrollo socioeconómico, en lugar del avance del conocimiento y la tecnología médica para combatir las enfermedades infecciosas. Aunque tuvo lugar más rápidamente y comenzó desde niveles más altos, la caída de la mortalidad en Argentina puede, en ese sentido, compararse con el patrón seguido por los países desarrollados con mayor distancia que la mayoría del resto de América Latina. Con el criterio de probar cómo performan los métodos elegidos, discriminamos la región Pampena como espacio geográfico, ya que es la región con mayor participación en total país, que además se caracateriza por una importante hetorogeneidad social, económica y demográfica.

A su vez, son pocos los estudios enfocados en el análisis de mortalidad a nivel sub-nacional en Argentina. Algunos estudios abordaron la tendencia de mortalidad infantil (Torcida, Vega, and Velázquez ([2008](#ref-torcida2008))) y las comunas de la Ciudad Autónoma de Buenos Aires (Grushka ([2013](#ref-Grushka2013))). Por ello, 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. 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 o administrativos.

Utilizamos los datos administrativos del registro de defunciones del Departamento de Estadísticas e Información de Salud del Ministerio de Salud de la Nación (DEIS), unidad oficial del gobierno a cargo de la administración de conteos de defunciones, y las estimaciones de población elaboradas por el Instituto Nacional de Estadísticas y censo (INDEC) agencia gubernamental argentina responsable de la recopilación y el procesamiento de datos estadísticos, como los censos de población. En primer lugar, aplicamos Métodos de Distribución de Muerte (DDM, por su sigla en inglés). Luego combinamos DDM con un enfoque empírico bayesiano, método ya aplicado para el caso de Brasil. El procedimiento de dos pasos nos permitió corregir la falta de registro de los recuentos de muertes y minimizar las fluctuaciones aleatorias que posibles al calcular los niveles de mortalidad en áreas pequeñas.

Lo que resta del artículo se divide de la siguiente forma: en la sección Datos se retoma un análisis inicial de calidad de datos, los problemas encontrados y las soluciones de corrección adoptadas.[[1]](#footnote-25) Luego, en Metodología se repasan las tres metodologías de suavizamiento elgidas, redefiniendo a su vez áreas mayores por fuera de lo estrictamente administrativo, y se describien los procedimientos de construcción de las tablas de mortalidad. Una vez estimada la heterogeneidad entre y dentro de los departamentos, en Resultados brevemente describimos de la relación observada entre la heterogeneidad interna y el nivel de mortalidad a nivel de cada provincia seleccionada.

# **Datos**

## Ajuste y corrección de datos\*\*

Se utilizaron los microdatos de defunción para los años de registro 2009, 2010 y 2011 provistos por el 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 (). 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 , 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%).[[2]](#footnote-29) 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.

Los datos desconocidos en áreas pequeñas son un problema importante (). Teniendo en cuenta solo las provincias pampeanas, Buenos Aires poseía los departamentos con el mayor porcentaje de edad y sexo desconocidos, con valores más elevados en sexo, siendo el líder el departamento General Pueyrredón con 7.3%. Debido a ello, adicionalmente no se realizó una 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 (Instituto Nacional de Estadística y Censos) a mediados de año en 2010, y se aplicó la estructura observada en el censo 2010 (INDEC ([2015](#ref-INDEC2015))). 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 (Gonzaga and Schmertmann ([2016](#ref-Gonzaga_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 permitó aprovechar las correcciones de omisión hechas por INDEC en el total[[3]](#footnote-30). Para eso se asumió una distribución uniforme de la fecha de nacimiento dentro del año y la población cerrada. Se utilizó una función de supervivencia única para todos los distritos, aplicando las mismas tablas de vida estándar que Gonzaga and Schmertmann ([2016](#ref-Gonzaga_Schmertmann_2016)) (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.[[4]](#footnote-31) (ver ). El gráfico 1 muestra los ajustes realizados.

![Ajuste de exposición 2008-2010. Cuatro casos. Fuente: elaboración propia en base a INDEC (2013)](data:application/pdf;base64,)

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

##**Chequeos de consistencia**

Existen cada vez más propuesta metodológicas por abordar el problema de cobertura de muertes en pequeñas áreas, dependiendo de la información auxiliar con que se cuente (Preston et al. ([1980](#ref-Preston1980)); Bennett and Horiuchi ([1984](#ref-Bennett_Horiuchi_1984)); Schmertmann and Gonzaga ([2018](#ref-Schmertmann2018)); Alexander, Zagheni, and Barbieri ([2017](#ref-Alexander2017))). 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 la esperanza de vida al nacer () a partir de los datos de muerte, descritos en la sección previa (Moultrie et al. ([2013](#ref-Moultrie))). Se trata de un método poco preciso para poblaciones pequeñas, pero puede dar una idea sobre los problemas en las áreas mayores (problemas relativos al numerador o denominador). En el Gráfico 2, observamos los puntos no ponderados y ponderados poblacionalmente, con el fin de otorgar mayor relevancia a la consistencia en las áreas más pobladas, responsables del posible sesgo de suavizamiento en los procedimientos metodológicos, que serán descritos en la siguiente sección (Figura ). Utilizando la edad promedio de la madre al nacimiento durante 2010 para cada provincia y la familia de tablas de la Organización de Naciones Unidas (ONU) de América Latina, de acuerdo a lo observado en el gráfico se concluyó que no hay un sesgo claro en las áreas más grandes.

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.

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) elaborado por el INDEC y la tasa de mortalidad general estadarizada a partir de la estructura por edad regional, buscando una relación esperada (Kaztman ([1995](#ref-Kaztman1995)), Preston ([1975](#ref-Preston_1975)); Grushka ([2013](#ref-Grushka2013))) (Gráfico 3). En la figura 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](#ref-Kaztman1995))). 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.[[5]](#footnote-33)

Standarized mortality rate and NBI. Departments in selected provinces. Source: own based in Census and DEIS

Standarized mortality rate and NBI. Departments in selected provinces. Source: own based in Census and DEIS

Se realizó un chequeo adicional de para cada provincia a partir de los insumos anteriores y la estimación oficial (INDEC ([2013](#ref-INDEC2013)))[[6]](#footnote-35). 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 ).

# **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](#ref-Longford2005))).

Para ello, tomamos el enfoque propuesto por Assuncao et al. ([2006](#ref-AssunCao2006)), 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 parea 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[[7]](#footnote-38). 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](#ref-Bivand2019))).

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 ).

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

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](#ref-JaspersOrellana1994))), se conoce que hay un porcentaje desigual de muertes infantiles no registradas por provincia (DEIS ([2016](#ref-DEIS2016))). 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](#ref-Brillinger1986))). 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](#ref-Efron1972)); Marshall ([1991](#ref-Marshall1991)); Data and Estimation ([2019](#ref-Longford2005)); Assunção et al. ([2005](#ref-Assuncao2005))). 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 , se supone que la distribución de muertes es un proceso de Poisson, con una media esperada de , siendo las exposiciones y la tasa de mortalidad.

Primero se considera a como el estimador de máxima probabilidad de la tasa en el área , que son generados a partir de . La esperanza condicionada de es (tasa de área grande) y la varianza condicionada .

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 : . Eso está relacionado con la relación jerárquica entre el hiperparámetro (), los parámetros () y sus estimadores .

El estimador lineal bayesiano que minimiza el error cuadrático medio de (e indicadores que son funciones lineales de esto) es (Robbins, 1983):

Nuevamente, es empírico porque se estima por método de momentos con , la media ponderada de áreas pequeñas. El factor de “contracción” (“shrinkage” en la bibliografía) es la relación entre la expectativa de la varianza estimada en el área pequeña y la varianza no condicionada del estimador, que es:

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 (). Por otro lado, las áreas de alto peso poblacional tomarán valores cercanos a los observados (). En el medio de estos extremos, la función combina linealmente la estimación del área grande con respecto al área más pequeña incluida.

Longford ([1999](#ref-Longford1999)) extendió esta idea a vectores de variables aleatorias (“contracción multivariada”), estimando de manera de aprovechar la correlación entre subpoblaciones. En nuestro caso, si la tasa de mortalidad del grupo de edad entre y del área es mayor que el área , 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](#ref-Assuncao2005)) (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](#ref-deBeer2011))), 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. Gonzaga and Schmertmann ([2016](#ref-Gonzaga_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 es una función de los “nodos” spline , que son las edades en las que se evaluará el desvío respecto al patrón estándar, siendo el vector de tasa de mortalidad estándar, y es la matriz B-spline que multiplicada por brinda el desplazamiento lineal entre el logaritmo de ambas tasas.

La idea es suponer que en cada área pequeña, construir la función de probabilidad usando las muertes y exposiciones observadas , pero expresando eso en función del parámetro , agregando una penalización por distancia desde el estándar y suavizando entre edades adyacentes. Se colocan nodos en edades más determinantes en el perfil de mrotalidad, para luego minimizar el log-likelihood: .

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](#ref-Assuncao2005))). 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](#ref-Arriaga2011))). 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 ajsute en el siguiente gráfico (ver ).

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

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

Se considera 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 en poblaciones donde la mortalidad infantil y juvenil aún tienen un peso importante en el indicador.[[8]](#footnote-42)

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 ).

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

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

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 stándard es de mayor fuerza gravitatoria. Los departamentos donde se reportan las mayores diferencias son aquellos con pocas celdas distintas de cero (ver ).

Estimaciones de mortalidad de los departamentos con mayores diferencias entre métodos. Fuente: elaboración porpia

Estimaciones de mortalidad de los departamentos con mayores diferencias entre métodos. Fuente: elaboración porpia

Por otro lado, se siguió el procedimiento empleado por Gonzaga and Schmertmann ([2016](#ref-Gonzaga_Schmertmann_2016)), relacionando las defunciones del área mayor con las obtenidas mediante la agregación de las estimaciones para las áreas menores (). 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.

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

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

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 investigación deberían utilizar técncias 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 y , donde se muestran las estimaciones en el intervalo 95%).

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 tamaño de los nombres sigue al tamaño relativo de la población en esa provincia. Fuente: elaboración propiaEsperanza 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 tamaño de los nombres sigue al tamaño relativo de la población en esa provincia. Fuente: elaboración propiaEsperanza 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 tamaño de los nombres sigue al tamaño relativo de la población en esa provincia. Fuente: elaboración propiaEsperanza 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 tamaño de los nombres sigue al tamaño relativo de la población en esa provincia. Fuente: elaboración propiaEsperanza 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 tamaño de los nombres sigue al tamaño relativo de la población en esa provincia. Fuente: elaboración propia

# **Conclusiones**

Alexander, Monica, Emilio Zagheni, and Magali Barbieri. 2017. “A Flexible Bayesian Model for Estimating Subnational Mortality.” *Demography* 54 (6). NIH Public Access: 2025. <https://doi.org/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. <https://doi.org/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 in Less Developed Countries.” *Demography* 21 (2). Springer: 217–33. <https://doi.org/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 Slaud de la Nación. <http://www.deis.msal.gov.ar/wp-content/uploads/2016/09/Estadisticasvitales2016.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. <https://doi.org/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 Calendario2010-2025.”

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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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>.

1. El material complementario, código y resultados se pueden encontrar en <https://github.com/nsacco/mortalidad_Argentina>. [↑](#footnote-ref-25)
2. Camisa ([2019](#ref-Camisa_2019)) notó este sesgo hace muchos años: 10% de los nacimientos y muertes durante el período 1946-1948. [↑](#footnote-ref-29)
3. 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](#ref-INDEC2015))) [↑](#footnote-ref-30)
4. 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 población (James et al. ([2014](#ref-James2014))) [↑](#footnote-ref-31)
5. 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) [↑](#footnote-ref-33)
6. A partir de lo que resta del artículo, se considera a Buenos Aires sin La Matanza) [↑](#footnote-ref-35)
7. <https://bit.ly/2sVpK9u>, visitado el 10/01/2020. [↑](#footnote-ref-38)
8. En términos matemáticos, están cerca de cero dependiendo de la no rectangularidad de 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 : (Wrycza and Baudisch ([2012](#ref-Wrycza2012))). [↑](#footnote-ref-42)