

Estado de la publicación: El preprint ha sido enviado para publicación en revista

Una Propuesta de Corrección en Tasas De Mortalidad por Diabetes Mellitus para Ciudades Colombianas con Problemas de Subregistros, 2005 – 2017

Juan de Jesus Sandoval, B. Piedad Urdinola Contreras, Paula A. Rivera Sarmiento

<https://doi.org/10.1590/SciELOPreprints.5048>

Enviado en: 2022-11-18

Postado en: 2022-11-18 (versión 1)

(AAAA-MM-DD)

Una Propuesta de Corrección en Tasas De Mortalidad por Diabetes Mellitus para Ciudades Colombianas con Problemas de Subregistros, 2005 – 2017

Autores

1. **Juan de Jesús Sandoval**, Escuela Superior de Administración Pública -ESAP-, Territorial Antioquia, Medellín, Colombia, juanj.sandoval@esap.edu.co, <https://orcid.org/0000-0003-2500-4289>
2. **B. Piedad Urdinola Contreras** Universidad Nacional de Colombia, Bogotá, bpurdinolac@unal.edu.co, <https://orcid.org/0000-0003-0273-8706>
3. **Paula A. Rivera Sarmiento**. Universidad Nacional de Colombia, pariveras@unal.edu.co, <https://orcid.org/0000-0003-1951-2295>

Resumen

El recuento de la mortalidad por enfermedades de baja prevalencia como la Diabetes Mellitus (DM), sufre de problemas serios de subregistro debido a la ausencia de datos, representados en ceros en el sistema de estadísticas vitales, lo que dificulta la necesidad de identificación de los patrones sobre todo de enfermedades que son silenciosas.

Esta investigación propone una corrección a las tasas de mortalidad para poblaciones pequeñas con alta inflación en ceros, como es el caso de la serie anual de Diabetes Mellitus (DM) para ciudades colombianas en edades de 20 a 80 años, entre los años 2005 y 2017. Se propone una modificación al modelo bayesiano jerárquico propuesto por Alexander et al. (2017), reemplazando la distribución original por una Binomial Negativa y utilizando los patrones de defunciones por la misma causa de otros países en la región, en años recientes. Con éste, se corrigieron las tasas de mortalidad por esta causa, edad, sexo y ciudad. Los resultados mostraron una apropiada corrección para las edades que oficialmente reportan ceros, manteniendo la brecha natural entre sexos, y una corrección mayor para las ciudades más pequeñas, frente a las grandes ciudades, cuando se comparó con otros modelos.

Con esta metodología se permitirá ayudar a la vigilancia epidemiológica de las enfermedades de baja prevalencia como la hipertensión, enfermedades coronarias y similares, con altos niveles de subregistro; pero que requieren de mediciones precisas a nivel subnacional para la toma de decisiones de salud pública. Cifras equivocadas solo llevarían a toma de decisiones erróneas, por ello es de vital importancia generar este tipo de correcciones.

Palabras clave: Mortalidad, Diabetes Mellitus, Subregistros, Pequeñas Áreas, Demografía

A Proposal to Correct Mortality Rates due to Diabetes Mellitus for Colombian Cities with Underreporting Problems, 2005 - 2017

Abstract

The mortality count for low-prevalence diseases such as Diabetes Mellitus (DM), demands serious problems of underreporting due to lack of data, generally represented as zeros in vital statistics systems, which makes it difficult to identify patterns, especially with diseases such as hypertension and diabetes, which are the main silent killer diseases. This paper proposes a correction to the mortality rates for small populations with high inflation in zeros, such as the annual series of Diabetes Mellitus (DM) for Colombian cities from 20 to 80 years old, between 2005 and 2017. A variation to the hierarchical Bayesian model of Alexander et al. (2017) is performed, replacing the original Poisson with a Negative Binomial distribution, but using the DM mortality patterns of other countries in the region in recent years. Therefore, DM mortality rates were corrected for age, sex, and city. The results showed a satisfactory correction for ages that officially report zeros, maintaining the natural gender gap, and a better correction for smaller cities, compared to larger ones. Because of this methodology, it will be possible to help the epidemiological surveillance of low-prevalence diseases such as hypertension, coronary heart disease and the like, with important levels of underreporting; but that require precise measurements at the subnational level for public health decision-making. Wrong data only leads to wrong decisions, so generating these types of corrections is vitally important.

Keywords: Mortality, Diabetes Mellitus, Under-registrations, Small Areas, Demography

Introducción

Existen enfermedades que impactan a las poblaciones humanas y cuya causa básica de muerte no necesariamente ocasiona mortalidad de manera directa, una de ellas es la Diabetes Mellitus (DM) (Agudelo-Botero and Dávila-Cervantes, 2015:176). La Diabetes Mellitus (DM) es un problema de salud pública que ha alcanzado niveles alarmantes en el mundo. Para el año 2019 se estimó que 463 millones de adultos de entre 20 y 79 años vivían con diabetes y, si la tendencia continúa, 700 millones de adultos tendrán diabetes para el año 2045. Además, de ellos, más del 80% viven en países con nivel de ingresos medios o bajos (International Diabetes Federation, 2019:42). De estos últimos países, la cifra puede ser aún mayor, pues el diagnóstico de la DM requiere de pruebas de laboratorio que no siempre están disponibles para todos los habitantes o ellos mismos no la demandan porque no saben o pueden auto identificar la enfermedad.

Este trabajo pretende llenar un vacío existente en el estudio de una causa de mortalidad con niveles altos en enfermedades de baja prevalencia como la Diabetes Mellitus, pero con graves problemas de subregistro en países en vía de desarrollo, con baja calidad de las estadísticas de mortalidad en áreas pequeñas, como lo son los municipios o ciudades pequeñas. La metodología propuesta se aplica al caso colombiano, pero es extensible a cualquier otro país o área pequeña que sufra de estos mismos problemas de subregistro. Esta investigación propone una modificación al modelo original de Poisson en la modelación de las tasas mortalidad (Alexander, Zagheni and Barbieri, 2017), al incorporar la modelación de la variación ocasionada por la sobre dispersión originada principalmente por presencia excesiva de ceros en la variable respuesta. De este modo, se planteó una variación funcional en la distribución de probabilidad dentro del modelo jerárquico bayesiano de Poisson, mediante una mezcla de distribución Gamma-Poisson (Agresti, 2012: 559-563), en el error, para finalmente producir un modelo Binomial Negativo (BN), en el conteo final de las muertes.

La investigación se centró en la DM por la importancia de la enfermedad en Latinoamérica, pero es extensible a otras enfermedades propias del proceso de envejecimiento que resultan de difícil identificación, como son la mayoría de las enfermedades de baja prevalencia. Una parte de este problema se debe, en gran medida, a la falta de la educación de la población respecto a estos temas de salud y sus cuidados. Por otra parte, existen al menos dos tipos de pruebas diagnósticas que son costosas y a veces confusas, lo que termina siendo excluyente (International Diabetes Federation, 2019:15). Las cifras en

la región estiman cerca de 32 millones de personas con DM y para el año 2045 se proyecta que esta cifra ascienda a 49 millones (Vargas-Uricoechea and Casas-Figueroa, 2016; International Diabetes Federation, 2019). Según la Federación Internacional de la Diabetes (FID), la prevalencia nacional de diabetes para adultos entre 20 y 79 años en América Latina varía notoriamente entre países, siendo Perú el que registra la cifra más baja (4,3%) y México (15,2%) y Puerto Rico las tasas más altas (16,8%) (Agudelo-Botero and Dávila-Cervantes, 2015; Alegre-Díaz *et al.*, 2016; International Diabetes Federation, 2019). Para Colombia (Aschner, 2010), la prevalencia se estimó alrededor de 7,2% en el año 2010 y se incrementó hasta 8,4% en 2019 (International Diabetes Federation, 2019:148).

Aunque existen trabajos realizados en Latinoamérica y el mundo que producen estimaciones de mortalidad por DM, utilizando la batería de estadísticas vitales (Laurenti, Marcondes Fonseca and Da Costa, 1982; Hamid Zargar *et al.*, 1999; Kirigia *et al.*, 2009; Arnold-Domínguez *et al.*, 2013), solo algunos estudios proponen la corrección de subregistro, ya sea a través de registro de defunción como por causas de muerte mal definidas o corrección por vecinos cercanos (Klafke *et al.*, 2015; Montoya-Betancur *et al.*, 2020). Lo cual es paradójico, pues hay evidencia amplia y suficiente del subregistro dentro de los reportes de estadísticas vitales, inclusive en países desarrollados (Ganguli *et al.*, 2005, 2006). En Latinoamérica, el subregistro en la mortalidad global a nivel nacional ha sido endémico y presenta mayores falencias en las edades extremo, aunque tiende a mejorar en el tiempo (Bay and Orellana, 2007; Palloni, Pinto and Beltran-Sanchez, 2015; Organización Panamericana de la Salud, 2017; Ribotta, Acosta and Bertone, 2019). Pero, el problema de subregistro se agrava cuando se estudian causas particulares de mortalidad y/o áreas subnacionales, como lo señalan varios autores (Lima and Queiroz, 2014; Buitelaar *et al.*, 2015; Queiroz *et al.*, 2017; Sandoval, Turra and Loschi, 2019).

El estudio de áreas menores o pequeñas poblaciones, es decir esta conjunción de mortalidad por edad, sexo, causa y área subnacional, es una necesidad para poder focalizar políticas que mejoren la calidad de vida de los habitantes sobre todos de las zonas rurales. Sin embargo, genera problemas de medición porque se gesta una sobre dispersión debido a la alta presencia de ceros en el conteo. Esta inflación en ceros representa la ausencia de casos, la ausencia de información (Rhon-Bunshi, 2016) y otras fuentes de error en la medición de las defunciones. De manera que, el aumento de porcentaje de ceros implica mayor variación estocástica en los recuentos de muertes y, por ende, problemas de sobre dispersión en las

estimaciones finales, donde la varianza de la variable original puede crecer de manera lineal conforme disminuyen (o aumentan), el número de variables a incorporar como, por ejemplo, la edad (Urdinola, Torres Avilés and Velasco, 2017).

El objetivo de esta investigación fue la de obtener un ajuste de estimación confiable de las tasas de mortalidad por DM en áreas pequeñas y para ello la metodología propuesta permitió separar las variaciones ocasionadas por los ceros debido a la ausencia de información en algunos grupos extremos de edad, en donde este porcentaje puede ser mayor conforme disminuye la densidad de población en dichas subpoblaciones. Los resultados fueron óptimos para las edades que tienen inflaciones en cero, pues las corrigió acertadamente, con intervalos bayesianos de credibilidad al 95% estrechos, mantuvo las diferencias en la mortalidad observada entre géneros. Por tanto, se propone una mayor corrección para las ciudades más pequeñas que para las más grandes, lo que está en línea con los resultados de otros estudios sobre la calidad de las cifras de vitales.

Antecedentes Teóricos y Empíricos

La transición demográfica va de la mano de la transición epidemiológica, la primera hace referencia al paso de altas tasas de fecundidad y mortalidad a bajas tasas de fecundidad y mortalidad, mientras que la segunda explica el tipo de enfermedades que va superando la humanidad, que constituye esa primera parte de la transición demográfica que es la caída en la mortalidad (Lee, 2003; Sandoval, Turra and Loschi, 2019). Inicialmente, las enfermedades transmisibles fueron la primera causa de muerte mientras que la reducción en la mortalidad comenzó intensivamente en las primeras edades (Wong and Carvalho, 2006; Lee, 2007; Saad, 2011). A medida que avanzó la transición demográfica la mortalidad se contrajo en las edades más jóvenes, avanzó sobre la adolescencia y adultez, en la última etapa en adultos mayores, lo que ha venido desencadenando el denominado envejecimiento de la población. En paralelo, ha venido avanzando el conocimiento y tecnología médica, el avance de la salud pública y la toma de decisiones que ha mejorado la salud de las poblaciones en general y las inversiones sociales que ayudan al aumento de la educación y salud generalizada, lo que se ha venido reflejando en una mejora de las condiciones de salud de las enfermedades transmisibles y se han incrementado las defunciones por enfermedades no transmisibles y neurodegenerativas (Horiuchi, 1999 : 54). Actualmente, hay tres países en América Latina que han completado su proceso de envejecimiento: Cuba, Chile y Uruguay. La mayoría de los países están en una etapa avanzada del envejecimiento, incluyendo el acelerado proceso en Colombia (Curcio *et al.*, 2018: 1), y unos pocos en una etapa media de la transición como son

Bolivia, Perú, Guyana y Paraguay. En consecuencia, la prevalencia de enfermedades transmisibles son minoría en la región, mientras avanza la de las enfermedades no transmisibles.

Dentro de este grupo de enfermedades no transmisibles, se encuentra la causa de interés de este documento la Diabetes Mellitus (DM) que, a nivel mundial, es una de las diez principales causas de fallecimiento (International Diabetes Federation, 2019: 54). La DM afecta la forma en que el cuerpo usa la glucosa sanguínea y dependiendo del tipo de diabetes puede generar un exceso de glucosa, lo que se traduce en todos los efectos negativos en el cuerpo (Kudva, 2020). Quienes viven con diabetes están en riesgo de desarrollar un conjunto de complicaciones graves y potencialmente mortales, que conllevan una creciente necesidad de atención médica, una reducida calidad de vida y un excesivo estrés para sus familias. Si la DM y sus complicaciones no se tratan de manera adecuada, los ingresos hospitalarios pueden ser más frecuentes y la muerte, prematura. Del mismo modo, el desarrollo de la enfermedad implica costos en tiempo y financieros que crecen a medida que avanza la enfermedad e incluyen la necesidad de insulina y aparatos de medición continua de niveles de glicemia en la sangre, como cardiopatías, neuropatías, nefropatías, retinopatías, enfermedades en los pies, deterioro en la piel, la visión y hasta depresión. Estos costos usualmente los asumen los servicios de seguridad social y las familias de los pacientes.

De este modo, la prevención de la progresión de la enfermedad es bastante costosa y poco eficiente, pues requiere de un cambio de estilo de vida hacia el denominado estado de vida saludable (González, Walker and Einarson, 2009; Barquera *et al.*, 2013). Esto implica un cambio de dieta que reduzca el consumo de alimentos hiper procesados, como dulces y harinas, la reducción de grasas saturadas y la práctica regular de actividad física diaria.

Las estimaciones confiables por niveles, patrones de edad y diferencias de sexo en mortalidad en áreas pequeñas, son importantes en Salud Pública, para evaluar y focalizar estos esfuerzos tanto en la prevención de la enfermedad, como en los gastos requeridos para el tratamiento de esta.

La alta variabilidad en los eventos vitales registrados crea serios problemas de mortalidad en estimaciones por edad y sexo. La combinación de baja ocurrencia y el estudio de poblaciones pequeñas hacen que las estimaciones de mortalidad sean una tarea difícil

(Bryant and Graham, 2013:591). Bajo esas circunstancias, las tasas de exposición/eventos observados son a menudo inestables a lo largo de las edades y los tiempos, y la estimación de los patrones de mortalidad subyacentes usualmente es más compleja que en áreas mayores.

Los métodos clásicos de correcciones de subregistros en demografía usualmente no funcionan para áreas menores o poblaciones pequeñas, más aún cuando se trata de causas específicas de mortalidad, como la DM. Ésta es un área incipiente y sólo en el último lustro se han comenzado a producir modelos para estimar las tasas de mortalidad para pequeñas áreas como los modelos TOPALS y TOPALS bayesianos (Gonzaga and Schmertmann, 2016) o el método bayesiano propuesto por (Alexander, Zagheni and Barbieri, 2017), que hace una aplicación de modelos bayesianos por componentes principales. Estos métodos se ejemplifican con estimaciones sobre la mortalidad general, pero en teoría deberían funcionar para causas particulares de mortalidad. Lima and Queiroz, 2014 emplearon modelos de regresión lineal para definir los cambios en la cobertura de los registros de muertes y el perfil de edad de mortalidad, respecto a los cambios en la presentación de causas mal definidas de muerte en áreas pequeñas, con aplicación a datos brasileños usando la comparabilidad a escala meso-regional, las cuales se definieron como unidades geográficas extensas e internamente muy heterogéneas (Buitelaar *et al.*, 2015).

Un método utilizado como parte del engranaje para la construcción de otros modelos, en pronósticos de mortalidad en demografía fue el análisis de componentes principales propuesto por (Lee and Carter, 1992), que se ocupó de explicar la estructura de la matriz de varianzas y covarianzas de un conjunto usualmente grande de variables a través de algunas pocas combinaciones lineales de estas mismas. Dentro de sus objetivos generales estuvo la reducción de la dimensionalidad de los datos para facilidad en la interpretación. En teoría según Johnson and Wichern, 2007:430, para un total de p -variables se requerirá p -componentes principales para reproducir el sistema total de la variabilidad del sistema. No obstante, gran parte de esta variabilidad se puede explicar por un número menor de estas componentes principales, digamos $k \leq p$. Siendo así, la lógica de las componentes principales lo que expresa es que si hay (casi) tanta información en las k -componentes como hay en las p -variables originales, estas k -componentes principales podrían reemplazar las p -variables iniciales y se conseguiría una reducción substancial del conjunto de datos que constaba inicialmente de p -mediciones, por las k -componentes principales. Adicionalmente, un análisis de los componentes principales a menudo revelará relaciones que no fueron

previamente sospechadas y, por lo tanto, permitiría interpretaciones que normalmente sucederían con los datos originales. Los análisis de las componentes principales son más un medio para lograr un fin que un fin en sí mismos, porque con frecuencia sirven como pasos intermedios en muchas investigaciones más amplias (Johnson and Wichern, 2007).

Materiales y Métodos

Para estimar las series de tasas de mortalidad por edad y sexo se utilizaron las estimaciones del total por municipio, obtenidos desde las estadísticas vitales del Departamento Administrativo Nacional de Estadísticas (DANE), de Colombia (DANE, 2021a), por búsqueda alfanumérica a partir de las diferentes líneas primarias y secundarias obtenidas desde los certificados de muerte. De allí, mediante búsqueda algorítmica alfanumérica, se obtuvieron los microdatos extraídos de los registros de mortalidad de Colombia para los 1103 municipios de Colombia, por lugar de residencia, entre los años 2005 y 2017. La definición de muerte por DM estuvo basada en la Clasificación Internacional de Enfermedades, cap. 10 (CIE10), según los siguientes códigos: E10:E14, O240: O244, O249, excluyendo causas de muerte por R99 (otras causas mal definidas / no especificadas, pero registradas por los médicos), y las R98, o muertes sin asistencia médica. Para la obtención del denominador de las tasas fueron utilizados datos de los censos de Colombia 2005 y 2018 extraídos del DANE(DANE, 2021b), proyectados al primero de julio del respectivo año. Finalmente, para el ajuste de tasas, fue seleccionada la información de cuatro países latinoamericanos para captar las regularidades en los patrones de edad de mortalidad por DM, mediante los mismos códigos anteriores de la CIE10, entre los años 2005 y 2017. Para ello, se escogieron los países de Argentina, Brasil, México y Colombia (World Health Organization, 2018), detalles más adelante.

Mediante una modificación al modelo propuesto por Alexander, Zagheni y Barbieri, 2017, se obtuvo un modelo bayesiano para el cálculo de las tasas de mortalidad con el objetivo de obtener estimaciones más estables de estas tasas a nivel subnacional. Al igual que el modelo ya mencionado, el nuevo modelo también se basó en la utilización de patrones de mortalidad que se construyeron empleando componentes principales en un conjunto de curvas de mortalidad. Esta información sobre las tasas de mortalidad se agrupó y se suavizó en el tiempo, y permitió incorporar nuevo conocimiento demográfico a las irregularidades en las estimaciones brutas de la mortalidad por DM por edad y sexo. Se eligió implementar esta metodología por la flexibilidad que brindó el método para realizar las estimaciones en

municipios con conteos de muertes muy pequeñas, nulas o ausentes representadas por la ausencia de registros.

La modificación al modelo original generó nuevas estimaciones con una distribución de las componentes principales (PCA), con el fin de corregir mortalidad en causas específicas de muerte por edad y sexo para países que tienen problemas de subregistro en mortalidad y quieran tener una estimación para una causa de muerte particular para áreas pequeñas. La aplicación del modelo se hizo sobre algunas ciudades o municipios de Colombia pues la DM es una enfermedad que se concentra más en zonas urbanas que rurales y porque esta distribución espacial de Colombia refleja bien la del resto de Latinoamérica, en donde más de 70% de la población habita zonas urbanas. De hecho, el censo nacional de población y vivienda 2018 de Colombia estimó que cerca de 77,1% de su población residía en zonas urbanas(DANE, 2021c).

Por tanto, se consideró que una variable respuesta Binomial Negativa Bayesiana pudiera captar mejor las variaciones ocasionadas por el exceso de ceros, que correspondieron a la ausencia de información en algunos grupos extremos de edad en causas específicas de mortalidad la Diabetes Mellitus (DM) en los registros de defunción de Colombia. A continuación, se explican los detalles del modelo.

Construcción del modelo

Considérese $y_{x,a,t}$ el número de muertes de una causa específica por DM, a la edad x en una municipio a en el tiempo t . Inicialmente, se puede asumir que el conteo de muertes de DM, tienen una distribución de probabilidad de Poisson. En ese sentido, se ha comprobado que para poblaciones grandes, el modelo de Alexander, Zagheni and Barbieri, 2017, se ajusta muy bien. No obstante, en la medida en que las áreas de estudio se hacen más pequeñas, donde podrían aparecer porcentajes cada vez mayores de ceros, que hace que la distribución ordinaria de Poisson por sí sola no capte la sobre dispersión ocasionada por la presencia de ceros en los distintos grupos de edad para las ciudades(Agresti, 2012:130).

Se sabe que la presencia de grandes volúmenes de ceros produce una sobre dispersión, que es ocasionada por inflación de ceros en el sentido estricto. El número excesivo de ceros invalida el supuesto de que la media y la varianza sean aproximadamente iguales, para un modelo clásico de Poisson (Lee and Carter, 1992; Winkelmann, 2008:109; Agresti, 2012:130), debido a que la varianza del conteo $y_{x,a,t}$ es cada vez mayor que la media .

Considérese la variable aleatoria, Y_{xat} , que representa el conteo de muertes de una causa específica de mortalidad, digamos DM, donde hay un porcentaje significativo de ceros en dicha variable. Se tiene que,

$$Y_{xat} \sim \text{Poisson}(\mu_{x,a,t}) \quad (1)$$

Donde es el modelo con error multiplicativo es el siguiente,

$$\mu_{x,a,t} = P_{x,a,t} \cdot \exp(\sum_{j=1}^P \beta_{j,a,t} Y_{j,a,t}) \cdot u_{x,a,t} \quad (2)$$

En el cual, $Y_{j,a,x}$ es la j -ésima componente principal de un conjunto de curvas de mortalidad específica estándar para alguna enfermedad, digamos DM, y $u_{x,a,t}$ es el efecto aleatorio. Para un modelo tradicional $u_{x,a,t} \sim N(\mu, \sigma^2)$. No obstante, para el caso de alta presencia de ceros, el error multiplicativo, se va a suponer que tiene $u_{x,a,t} = \log(v_{x,a,t})$, en donde $v_{x,a,t} \sim \text{Gamma}(r_{x,a}, 1/r_{x,a})$ y $r_{a,t} \sim \text{Uniforme}(0, K)$, para $K \in \mathbb{Z}$.

Y donde, $r_{x,a}$ a la edad x en la ciudad a , se denominará parámetro de sobre dispersión y se cumple que el valor esperado (E), $E(u_{x,a,t}) = 1$ y $\text{Var}(u_{x,a,t}) = 1/r_{x,a}$, para todo t . Se puede demostrar que la integral marginal sobre $u_{x,a,t}$ desde 0 hasta t en (2), produce una distribución Binomial Negativa (BN), con parámetros: media $\mu_{x,a,t}$ y sobredispersión inversa $\frac{1}{r_{x,a}} = \varphi_{x,a}$,

De tal manera que linealizando (2), se obtiene,

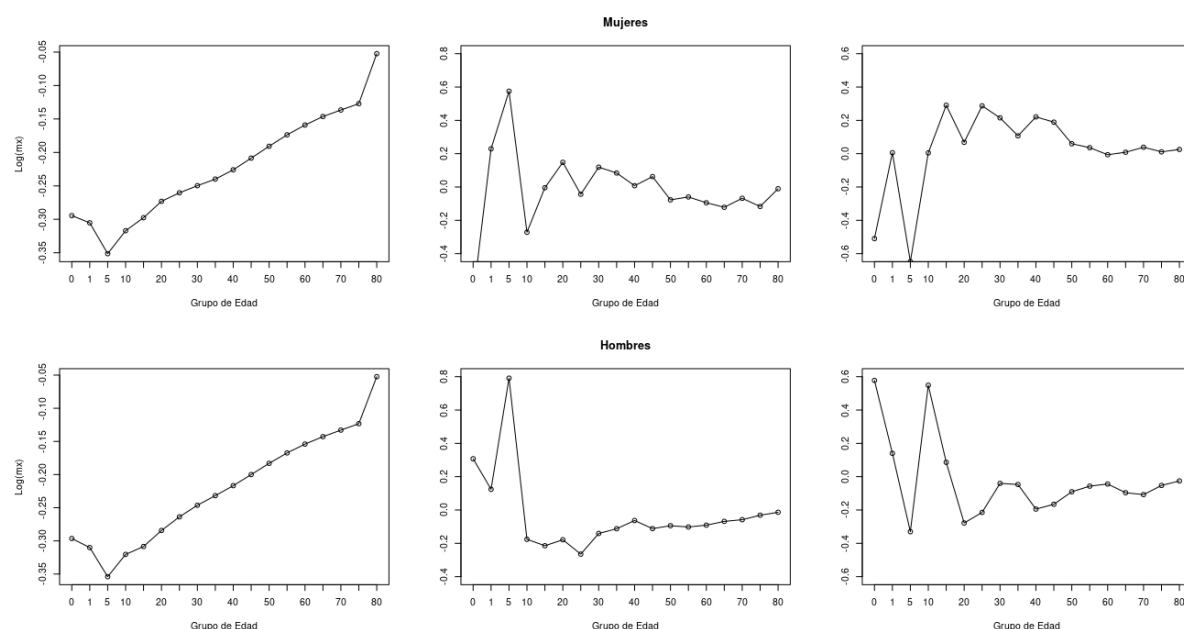
$$\log(\mu_{x,a,t}) = \log(P_{x,a,t}) + \sum_{j=1}^P \beta_{j,a,t} Y_{j,a,t} + \log(u_{x,a,t}) \quad (3)$$

Donde la expresión $\log(P_{x,a,t})$, es el offset o el log de los años persona exposición, mediante la población expuesta a mitad de periodo.

El uso de las componentes principales tuvo el mismo enfoque de Lee and Carter, 1992, que fue utilizado por Alexander, Zagheni and Barbieri, 2017. Para ello, las componentes principales fueron creadas con base en la información de cuatro países latinoamericanos captando las regularidades en los patrones de edad de mortalidad por diabetes mellitus entre los años 2005 y 2017. Se escogieron los países de Argentina, Brasil,

México y Colombia, pues se encontraron todos aproximadamente en la misma etapa de la transición demográfica, las tasas de urbanización son similares entre ellos (alrededor de 80%). Además, sustentaron una distribución poblacional por edad y sexo muy semejante y son, además, los cuatro países más poblados en América Latina. También, México sostiene una de las mayores tasas de prevalencia de la enfermedad DM en la región y en el mundo (Barquera *et al.*, 2013; Blue, Goldman and Rosero-Bixby, 2015; Alegre-Díaz *et al.*, 2016), con una consecuente altísima tasa de mortalidad, que permite mayor heterogeneidad en las tasas observadas de los otros tres países. La información se obtuvo del Observatorio Mundial en Salud de la Organización Mundial de la Salud-OMS (World Health Organization, 2018). La Figura 1 presenta las curvas estándares que se obtuvieron a partir de la conjugación de la información de estos cuatro países a partir del análisis de componentes principales y el cálculo de las tablas de vida.

Figura 1: Componentes principales (PCA), de los registros de mortalidad por DM en Argentina, Brasil, México y Colombia (logaritmo natural), para hombres y mujeres entre los años 2005–2017.



Fuente: Cálculos propios a partir de los datos del Observatorio Mundial de la Salud (OMS).

Nota: Se supone que las tasas brutas de mortalidad por DM para todo un país tienen menos subregistros. Por tanto, permiten corregir las áreas subnacionales.

Estas curvas se presentaron en escala logarítmica y describen los patrones de la mortalidad específica por DM en Argentina, Brasil, México y Colombia, combinados, por

edades 0, 1 y quinquenales a partir de esa edad, y sexo biológico para los años de 2005 a 2017. Las curvas de la primera columna corresponden a la línea de base de la mortalidad general, la segunda permite representar la mortalidad mayor en edades más tempranas, en relación con la mortalidad de adultos y la permite que la mortalidad de adultos sea mayor en relación con la mortalidad joven. Vale la pena destacar que las mujeres presentan mayores tasas de mortalidad en adultos jóvenes y levemente mayor en adultos mayores. Para nuestros análisis solo se tuvo en cuenta las PCA a partir de los 20 años.

Distribuciones *a priori*

Las distribuciones a priori de los parámetros $\beta_{p,a,t}$ para $p: 1,2,3$ en la expresión (3) para el log, se mantuvieron como están definidos en Alexander et al. (2017). No obstante, hubo una modificación en el efecto aleatorio donde $u_{x,a,t} = \log(v_{x,a,t})$, en que $v_{x,a,t} \sim \text{Gamma}(r_{x,a}, 1/r_{x,a})$ y $r_{a,t} \sim U(0, K)$. En donde K es un número grande y U es una distribución uniforme no informativa. Para simular el proceso por muestreo de Gibbs mediante cadenas de Markov de Monte Carlo (MCMC), se utilizó un algoritmo construido en Rjags (Plummer, 2013), utilizando el lenguaje computacional R (R Core Team, 2021), para los que se realizaron 30.000 interacciones, con una eliminación o quema (burning), de los primeros 10.000 para el ajuste de las estimaciones de las tasas específicas de mortalidad por edad, sexo y la causa de interés (TEM) a posteriori. El proceso de convergencia tuvo un Rhat (Gelman et al., 2004: 601-617), con una media de 1,002 IC 95% para Rhat : (1 ; 1,006082), lo que indicó una excelente convergencia para las $p = 1,2,3$ componentes, análisis independiente para hombres y para mujeres. Para más detalles ver el algoritmo en https://github.com/jjsandoval/DM_Paper1.

Comparación y Evaluación de modelos

Es importante recalcar que se realizaron comparaciones del modelo binomial negativo, con el modelo Poisson cero inflado y con el modelo tradicional de Poisson propuesto por Alexander et al, 2017. La tabla 1 muestra las estimaciones de las tasas ajustadas de mortalidad x 100.000 habitantes para los municipios Bogotá (7.181 millones de Hab), Fuente de Oro (11.599 Hab), Medellín (2.427 millones de Hab), Mocoa (58.938 Hab), Popayán (318.059 Hab), Soledad (666.109 Hab) y Valledupar (459.349 Hab), según el censo 2018, para comparar sus estimaciones con los diferentes modelos. Se observa que en la medida en que el municipio es más pequeño y más baja es el grupo de edad, las estimaciones con el modelo de Alexander et

al, 2017 al igual que con el modelo *ZIP* empeoran, pero con un poco más de estabilidad en el intervalo de credibilidad en este último. No sucede lo mismo con nuestra propuesta, el modelo Binomial Negativo, que se muestra más consistente con la presencia de ceros en los datos. Es importante destacar que en municipio más pequeño (Fuente de Oro), es donde se presentan más problemas en las estimaciones en los dos primeros modelos, sobre todo en el año 2005 y en edades de 75-79 y 40-44, donde se evidencian subregistros en el modelo *AZB* y sobre registros en el modelo *ZIP*, sobre todo en los intervalos. No pasa lo mismo con el modelo *BN*.

Tabla 1: Comparación de la estimación de las tasas de mortalidad ajustadas por tres modelos en para tres grupos de edad en mujeres entre los años 2005, 2010 y 2017. Intervalos de credibilidad al 95% para las medianas estimadas.

Año	Grupo de edad	Municipio	Tasa Bruta	AZB			ZIP			BN		
				Mediana	LI95%	LS95%	Mediana	LI95%	LS95%	Mediana	LI95%	LS95%
2005	75-79	Bogotá, D.C.	412,1	404,0	354,6	457,1	404,0	354,6	458,2	407,0	356,2	465,8
2010	75-79	Bogotá, D.C.	439,6	414,4	369,0	466,0	414,4	366,3	464,2	420,8	374,0	473,7
2017	75-79	Bogotá, D.C.	260,6	251,6	220,9	283,0	250,0	219,7	283,9	254,3	221,1	288,9
2005	40-44	Bogotá, D.C.	4,9	1,9	1,1	3,0	1,8	1,1	2,9	4,2	3,0	5,8
2010	40-44	Bogotá, D.C.	3,6	2,2	1,3	3,3	2,2	1,3	3,3	3,8	2,7	5,2
2017	40-44	Bogotá, D.C.	4,3	2,5	1,6	3,7	2,4	1,5	3,6	3,6	2,5	5,0
2005	20-24	Bogotá, D.C.	0,6	0,2	0,1	0,5	0,2	0,1	0,5	0,5	0,3	0,7
2010	20-24	Bogotá, D.C.	0,6	0,2	0,1	0,4	0,2	0,1	0,4	0,4	0,2	0,7
2017	20-24	Bogotá, D.C.	1,0	0,1	0,1	0,3	0,1	0,1	0,3	0,4	0,2	0,7
2005	75-79	Medellín	464,5	458,5	390,8	535,9	457,0	389,7	534,3	462,8	392,5	544,3
2010	75-79	Medellín	309,4	289,9	239,9	346,5	287,4	236,6	344,4	309,9	256,9	369,9
2017	75-79	Medellín	202,8	183,5	149,3	225,3	182,2	147,6	223,0	199,5	161,2	242,1
2005	40-44	Medellín	8,6	2,2	1,1	4,1	2,2	1,0	4,1	6,6	4,6	9,2
2010	40-44	Medellín	8,1	0,6	0,2	1,4	0,6	0,2	1,4	4,4	2,9	6,6
2017	40-44	Medellín	1,3	0,1	0,0	0,3	0,1	0,0	0,3	2,9	1,8	4,6
2005	20-24	Medellín	1,1	0,3	0,1	0,7	0,3	0,1	0,7	0,8	0,4	1,3
2010	20-24	Medellín	2,0	0,3	0,1	0,8	0,3	0,1	0,8	0,5	0,3	1,1
2017	20-24	Medellín	2,0	0,2	0,0	0,6	0,2	0,1	0,7	0,5	0,2	1,1
2005	75-79	Soledad	102,8	140,1	76,2	231,4	148,1	81,8	234,3	217,8	139,3	325,9
2010	75-79	Soledad	117,5	99,9	57,3	163,7	103,1	57,6	170,6	174,4	110,5	263,1
2017	75-79	Soledad	153,9	99,2	59,9	159,2	101,0	59,9	161,7	142,8	93,6	214,7

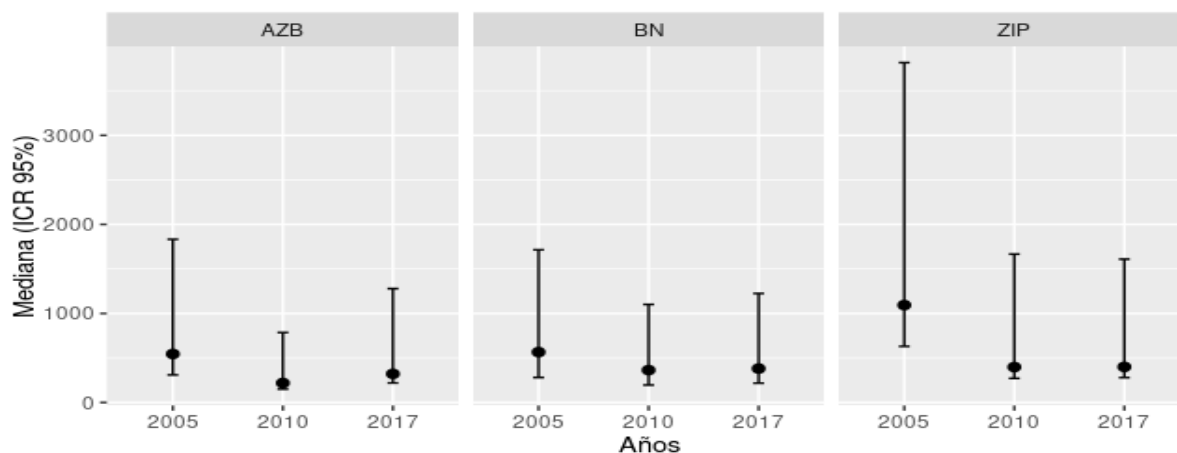
Año	Grupo de edad	Municipio	Tasa Bruta	AZB			ZIP			BN		
				Mediana	LI95%	LS95%	Mediana	LI95%	LS95%	Mediana	LI95%	LS95%
2005	40-44	Soledad	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,9	0,8	3,9
2010	40-44	Soledad	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,3	0,5	2,9
2017	40-44	Soledad	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,6	0,6	3,3
2005	20-24	Soledad	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,1	0,6
2010	20-24	Soledad	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,4
2017	20-24	Soledad	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,1	0,8
2005	75-79	Valledupar	336,2	438,8	287,6	643,6	459,5	295,4	670,8	530,1	374,3	734,3
2010	75-79	Valledupar	893,9	498,1	338,5	699,4	499,1	341,9	714,7	585,8	425,2	796,9
2017	75-79	Valledupar	610,9	480,2	340,3	651,8	480,2	339,6	660,1	503,1	373,0	684,8
2005	40-44	Valledupar	0,0	0,7	0,1	4,3	1,0	0,1	6,2	9,0	5,0	16,1
2010	40-44	Valledupar	0,0	0,1	0,0	0,6	0,1	0,0	0,6	8,7	4,8	15,0
2017	40-44	Valledupar	6,6	0,1	0,0	0,7	0,1	0,0	0,7	7,4	3,8	12,6
2005	20-24	Valledupar	0,0	0,1	0,0	0,4	0,2	0,0	0,6	1,1	0,5	2,5
2010	20-24	Valledupar	0,0	0,1	0,0	0,7	0,2	0,0	0,7	1,2	0,5	2,5
2017	20-24	Valledupar	4,7	0,1	0,0	0,6	0,1	0,0	0,7	0,9	0,3	2,4
2005	75-79	Popayán	369,6	390,8	258,8	578,7	388,9	264,5	578,3	441,7	308,1	616,5
2010	75-79	Popayán	224,2	283,8	184,5	433,7	282,5	185,6	431,5	340,0	228,8	483,1
2017	75-79	Popayán	385,2	379,8	261,2	530,4	381,7	262,9	548,9	395,3	279,9	550,5
2005	40-44	Popayán	0,0	0,0	0,0	0,0	0,0	0,0	0,0	5,7	2,9	10,3
2010	40-44	Popayán	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,7	1,7	7,4
2017	40-44	Popayán	0,0	0,0	0,0	0,3	0,0	0,0	0,3	6,4	3,2	12,6
2005	20-24	Popayán	0,0	0,0	0,0	0,1	0,0	0,0	0,1	0,7	0,3	1,5
2010	20-24	Popayán	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,2	1,0
2017	20-24	Popayán	0,0	0,0	0,0	0,3	0,1	0,0	0,3	0,8	0,3	2,3
2005	75-79	Mocoa	558,7	273,1	104,5	609,7	676,7	297,7	1449,2	410,8	210,7	767,2
2010	75-79	Mocoa	1005,0	203,8	76,2	466,1	530,4	201,0	1381,7	359,0	176,3	654,3
2017	75-79	Mocoa	0,0	238,6	95,3	529,5	461,4	157,3	1337,2	290,4	135,6	575,7
2005	40-44	Mocoa	0,0	0,0	0,0	0,5	36,8	0,4	374,7	5,3	1,6	14,3
2010	40-44	Mocoa	0,0	0,0	0,0	0,0	0,0	0,0	365,4	3,8	1,1	10,6
2017	40-44	Mocoa	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,8	1,0	11,4
2005	20-24	Mocoa	0,0	0,0	0,0	0,2	0,6	0,1	2,8	0,6	0,1	2,2
2010	20-24	Mocoa	0,0	0,0	0,0	0,0	0,1	0,0	2,0	0,4	0,1	1,6
2017	20-24	Mocoa	0,0	0,0	0,0	0,1	0,0	0,0	0,1	0,5	0,1	2,2
2005	75-79	Fuente de Oro	0,0	544,9	235,1	1287,3	1093,8	463,4	2722,6	565,9	284,6	1149,0
2010	75-79	Fuente de Oro	0,0	218,7	68,7	566,9	398,1	127,2	1268,1	363,8	167,3	738,0
2017	75-79	Fuente de Oro	769,2	321,7	101,6	957,0	400,1	120,3	1209,2	381,1	163,3	843,6
2005	40-44	Fuente de Oro	0,0	0,1	0,0	17,4	1118,1	0,1	3894970,1	8,5	2,6	27,8
2010	40-44	Fuente de Oro	0,0	0,0	0,0	0,0	0,0	0,0	182,3	4,3	1,1	13,6
2017	40-44	Fuente de Oro	0,0	0,0	0,0	0,1	0,0	0,0	0,3	5,3	1,5	18,9
2005	20-24	Fuente de	0,0	0,1	0,0	0,7	2,3	0,1	37,6	1,1	0,3	4,7

Año	Grupo de edad	Municipio	Tasa Bruta	AZB			ZIP			BN		
				Mediana	LI95%	LS95%	Mediana	LI95%	LS95%	Mediana	LI95%	LS95%
2010	20-24	Oro	0,0	0,0	0,0	0,1	0,0	0,0	1,7	0,5	0,1	2,0
2017	20-24	Fuente de Oro	0,0	0,0	0,0	0,2	0,0	0,0	0,2	0,6	0,1	3,3

Fuente: Cálculos propios a partir de los datos de las estadísticas vitales DANE, Colombia y suavizado a partir de datos del Observatorio Mundial de la Salud (OMS). **Nota:** ZIP: modelo cero inflado, BN: modelo binomial negativo, AZB: modelo Alexander et al, 2017.

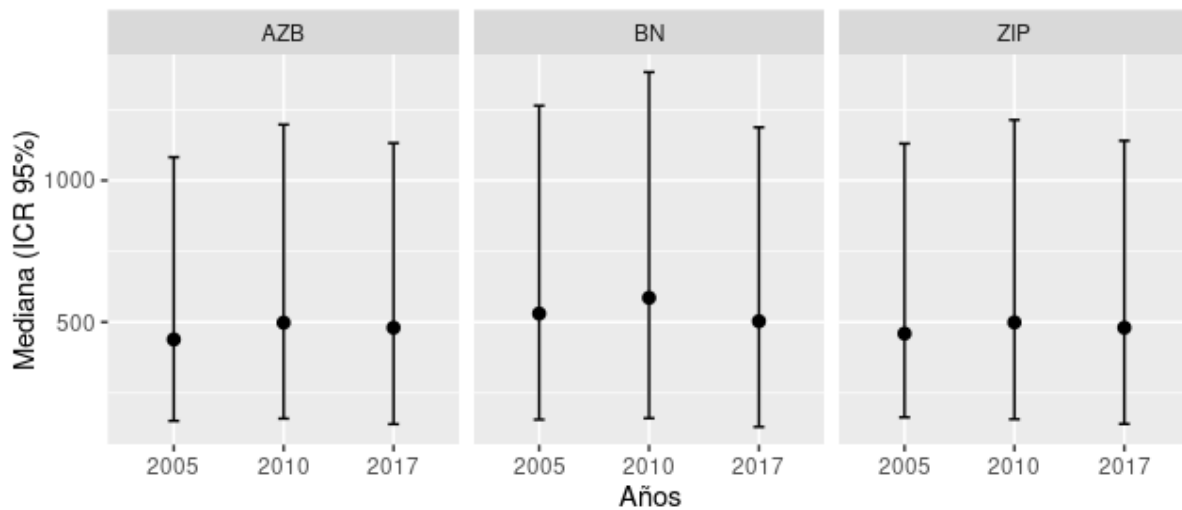
Del mismo modo las figuras 2 a 4 muestran las estimaciones de los municipios de Medellín (2.427.129 habitantes para 2018), Valledupar (490.075 habitantes para 2018) y Fuente de oro (11.599 habitantes para 2018) para las edades de 75-79 años de la DM x 100.000 habitantes. Se observa en la ciudad de Medellín (figura 2) y Valledupar (figura 3) para el grupo 75-79 años, los tres modelos coinciden aproximadamente, con solo pequeñas desviaciones, en cuanto a las estimaciones ajustadas, siendo el modelo de AZB el que produjo la estimación más baja y el modelo BN la más alta. En cuanto a mujeres de Fuente de Oro (figura 4), ya empiezan a observarse grandes diferencias, la más grande de ellas se da en el modelo ZIP. En los dos primeros modelos las diferencias son leves para este grupo de edad en el municipio más pequeño de los analizados.

Figura 2. Tasas de mortalidad por diabetes mellitus x 100.000 ajustadas por los tres modelos para mujeres Medellín, Colombia, 2005, 2010 y 2017 en edades 75-79 años. Estimación mediana *a posteriori* mediante los tres modelos con intervalos de credibilidad del 95



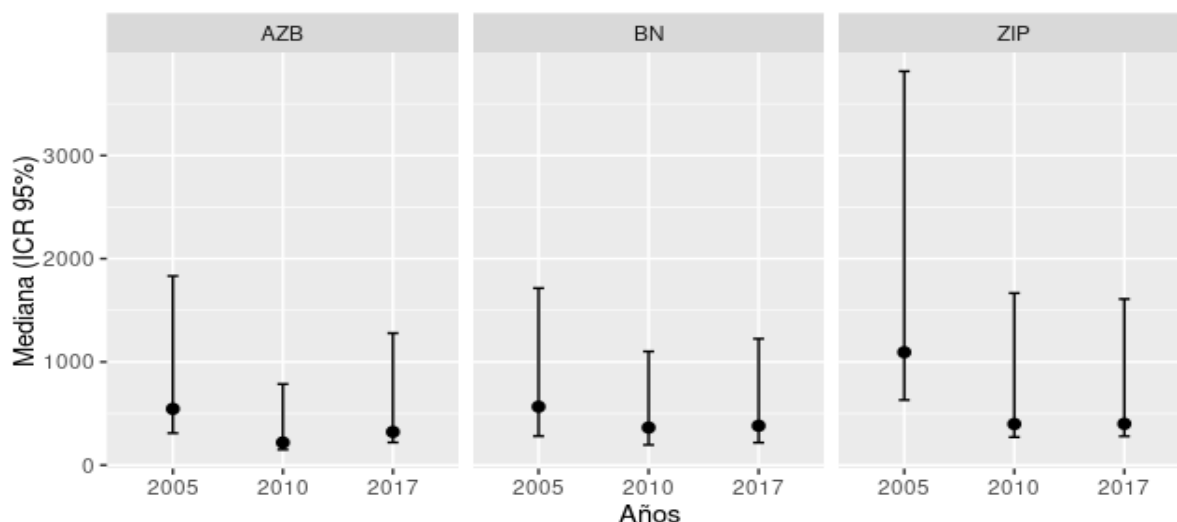
Fuente: Cálculos propios.

Figura 3. Tasas de mortalidad por diabetes mellitus x 100.000 ajustadas por los tres modelos para mujeres Valledupar, Colombia, 2005, 2010 y 2017 en edades 75-79 años. Estimación mediana *a posteriori* mediante los tres modelos con intervalos de credibilidad del 95



Fuente: Cálculos propios.

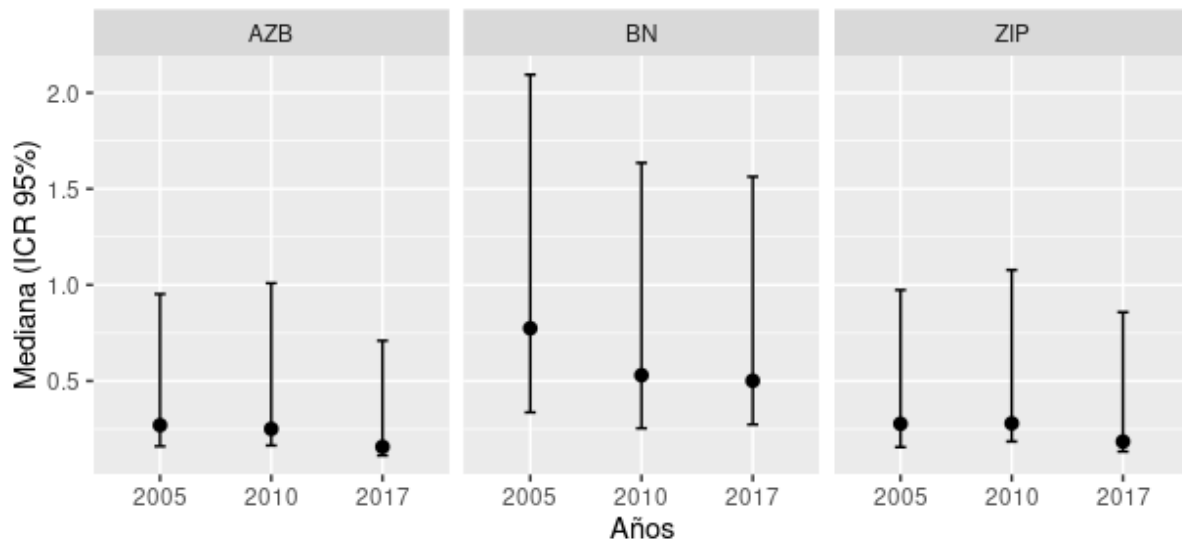
Figura 4. Tasas de mortalidad por diabetes mellitus x 100.000 ajustadas por los tres modelos para mujeres Fuente de Oro, Colombia, 2005, 2010 y 2017 en edades 75-79 años. Estimación mediana *a posteriori* mediante los tres modelos con intervalos de credibilidad del 95.



Fuente: Cálculos propios.

No obstante, las figuras desde la 5 - 7, que muestran las estimaciones de la DM para mujeres por el grupo 20-24 años que tiene más ceros, donde se muestra gran diferencia entre las estimaciones y en este caso el modelo Binomial negativo es el que produce las mejores estimaciones, sobre todo en los municipios más pequeños. En este sentido, para el caso de Medellín (figura 5), el modelo BN es el que produce las mayores estimaciones de las medianas, pero también mayor amplitud en los intervalos de credibilidad. En este sentido, el modelo AZB y ZIP produjeron resultados levemente similares.

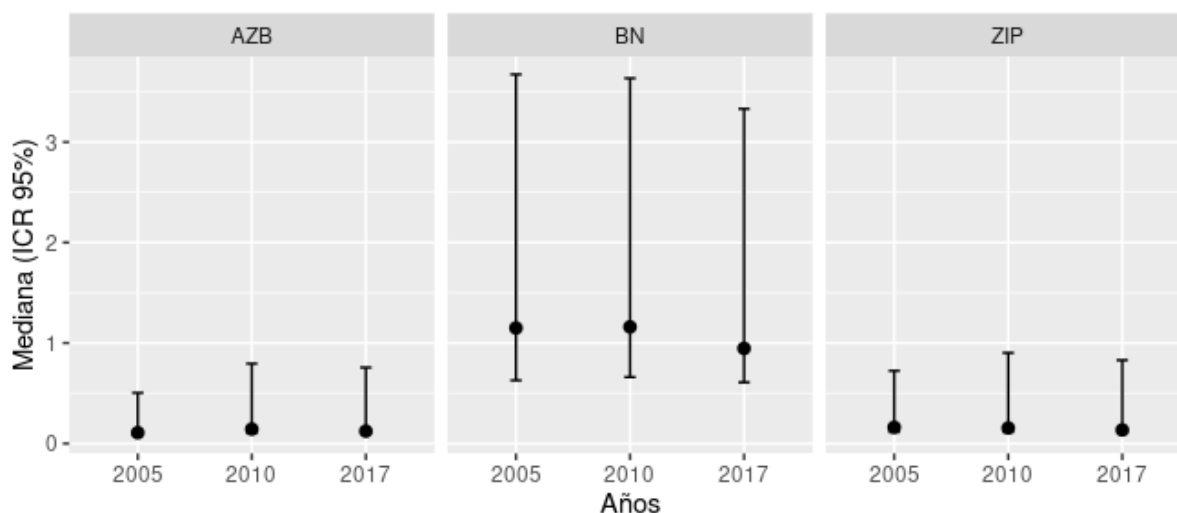
Figura 5. Tasas de mortalidad por diabetes mellitus x 100.000 ajustadas por los tres modelos para mujeres Medellín, Colombia, 2005, 2010 y 2017 en edades 20-24 años. Estimación mediana *a posteriori* mediante los tres modelos con intervalos de credibilidad del 95



Fuente: Cálculos propios.

Cuando el municipio se hace más pequeño, es el caso de ciudad de Valledupar (figura 6), se observa que las estimaciones del modelo AZB y ZIP son muy similares y caso no captan la variabilidad de las estimaciones con ceros, siendo muy cercanas a cero. De otro lado el modelo BN produjo unas tasas ajustadas, muy cercanas al 1×100.000 para este grupo de edad, considerando en este caso el modelo que mejor captó la variabilidad ocasionada por los ceros en los datos.

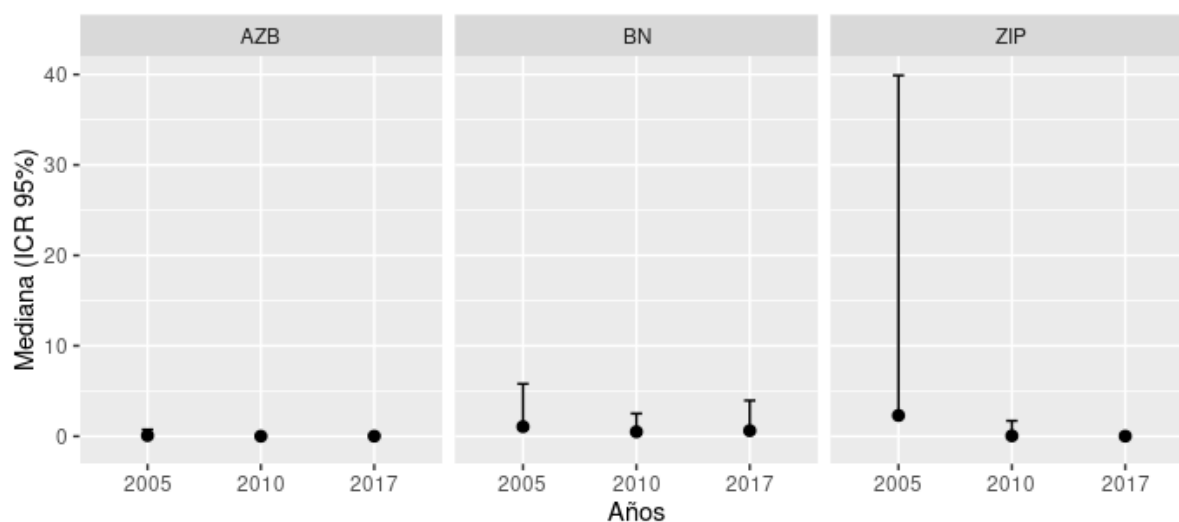
Figura 6. Tasas de mortalidad por diabetes mellitus x 100.000 ajustadas por los tres modelos para mujeres Valledupar, Colombia, 2005, 2010 y 2017 en edades 20-24 años. Estimación mediana *a posteriori* mediante los tres modelos con intervalos de credibilidad del 95.



Fuente: Cálculos propios.

Finalmente, la figura 7, muestra las estimaciones de uno de los municipios más pequeños considerados, en un grupo poblacional, también muy escaso en cuanto a la DM. Se observa que el modelo AZB, no capta la variabilidad ocasionada por los ceros, el modelo ZIP produce cambios muy bruscos en 2005 y no capta en 2010 y 2017. No obstante, el modelo BN produce estimaciones mayores, donde se evidencia que captó mejor la alta variabilidad ocasionada por los ceros, produciendo una estimación mediana de 6 casos de DM por cada 1.000.000 de habitantes con un intervalo de credibilidad entre 1 caso y 33 por cada millón de habitantes.

Figura 7. Tasas de mortalidad por diabetes mellitus x 100.000 ajustadas por los tres modelos para mujeres Fuente de Oro, Colombia, 2005, 2010 y 2017 en edades 20-24 años. Estimación mediana *a posteriori* mediante los tres modelos con intervalos de credibilidad del 95



Fuente: Cálculos propios

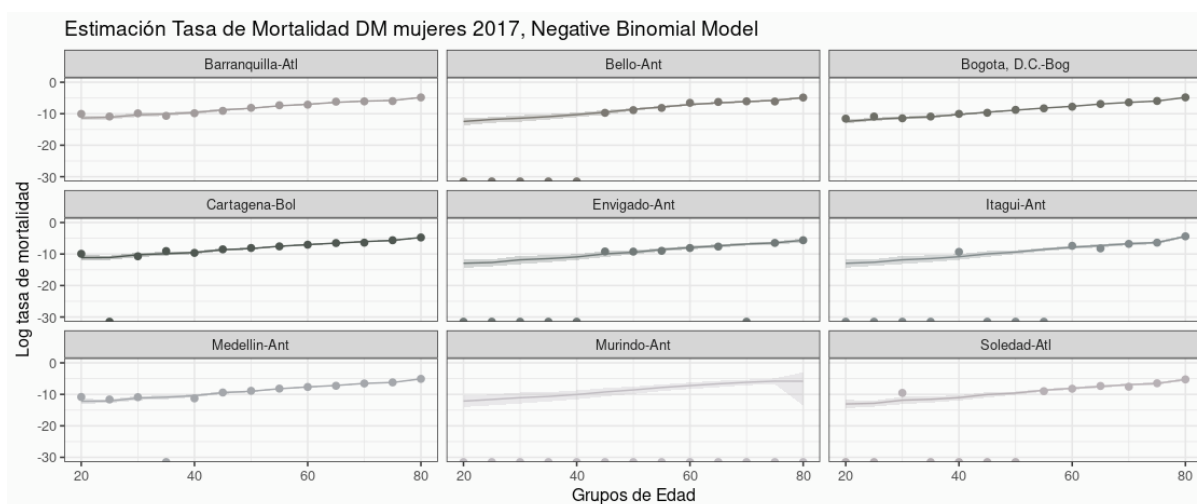
Análisis de Resultados

A continuación, mostrarán los resultados de análisis de las tendencias por los diferentes grupos de edad para el modelo final de Regresión Binomial Negativa Bayesiano (BN). Se muestran las gráficas del logaritmo de las tasas específicas de mortalidad observadas o reportadas por los registros oficiales de defunciones (los puntos) y las ajustadas representadas por las líneas, que sigue el modelo propuesto modelo BN, en mujeres para algunos municipios colombianos, con los respectivos intervalos de credibilidad al 95%, representados en el área sombreada. Se presentan solo algunos años por cuestiones de espacio. Todos los demás años no presentados en este documento producen similares resultados.

La figura 8,9, 10 y 11 muestran los resultados para las mujeres en 2017 para grupos de 9 ciudades, que se escogieron por los diferentes comportamientos que caracterizan los resultados obtenidos. Se observa en las ciudades más grandes, como Medellín y Bogotá, que el ajuste es casi perfecto entre los datos observados y estimados por el modelo BN. Mientras en las ciudades intermedias, que son de menor tamaño poblacional y por consiguiente con mayor presencia de ceros, se observa que en las edades más jóvenes, cuando hay valores de cero se corrige hacia el patrón modelado a partir del análisis de componentes principales y es precisamente para esos valores donde se genera una corrección por el modelo BN Bayesiano, de acuerdo a la información a priori aportada.

De manera que, el modelo funciona bien en la captura del patrón general de la mortalidad en estas áreas pequeñas, pero es bastante susceptible a valores extremos, un defecto que requiere de un desarrollo metodológico posterior. Por eso, la sugerencia para la utilización de este tipo de modelos es solo aplicar la corrección para los datos que poseen valores inflados en ceros. Es curioso como municipios como Murindó (Figura 8, posición 8), con 4911 habitantes para 2018 según el Censo 2018, muestra estimaciones mediante el modelo, solo que éste se inclina mucho más a la información a priori del modelo bayesiano.

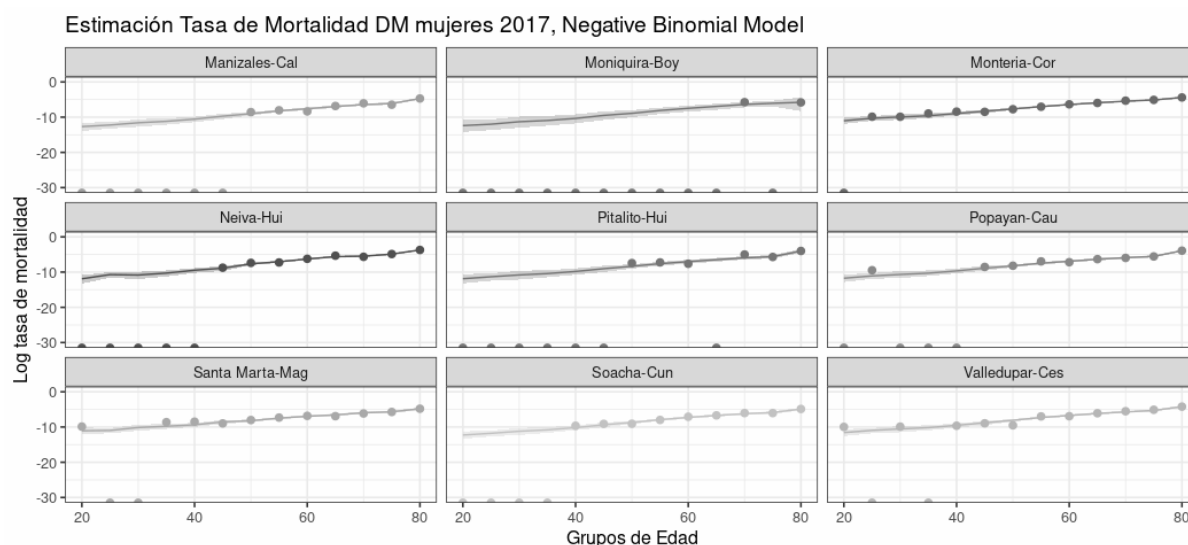
Figura 8. Tasas de mortalidad observadas y estimadas por diabetes mellitus (en logaritmo) para mujeres en nueve ciudades de Colombia, 2017. Estimación mediana *a posteriori* del modelo BN con intervalos de credibilidad del 95%.



Fuente: Cálculos propios. **Nota.** Los puntos sobre la línea horizontal inferior de las gráficas representan el $\log(0) = -\infty$. El logaritmo de la ausencia de un registro da un negativo valor muy pequeño.

Análogamente se podría decir que para los municipios de Moniquirá y Pitalito (Fig 9, pos 2 y 5), con 20848 y 135711 habitantes respectivamente para 2018, también son de los presentan estimaciones de las tasas ajustadas por DM, inclinándose más a la información a priori del modelo Bayesiano.

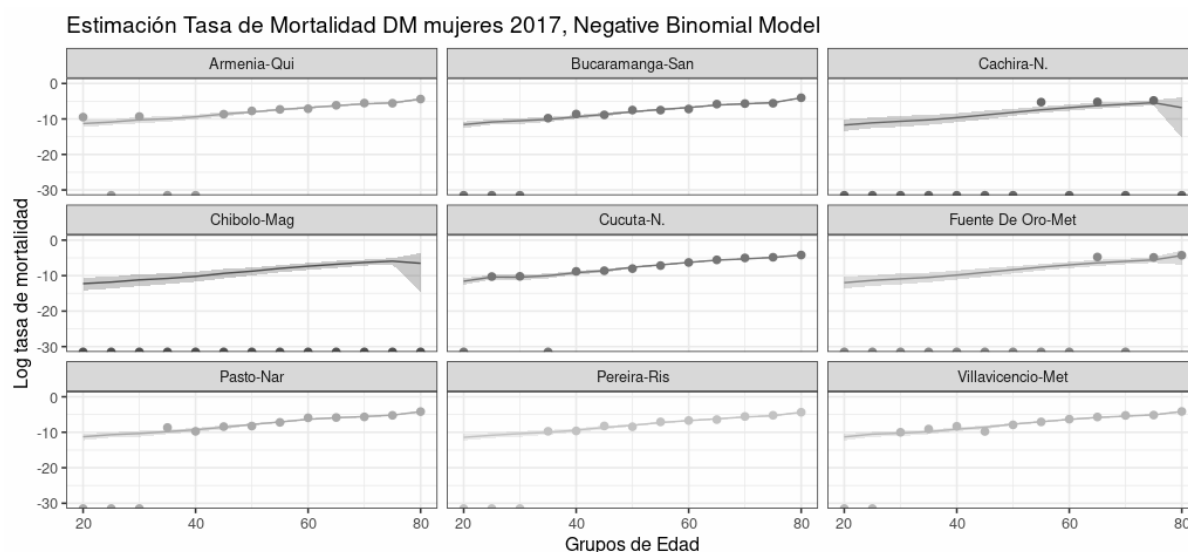
Figura 9. Tasas de mortalidad observadas y estimadas por diabetes mellitus (en logaritmo) para mujeres en nueve ciudades de Colombia, 2017.



Fuente: Cálculos propios. **Nota.** Los puntos sobre la línea horizontal inferior de las gráficas representan el $\log(0) = -\infty$. El logaritmo de la ausencia de un registro da un negativo valor muy pequeño.

Análogamente, en la figura 10, para el municipio de Chibolo (18208 Hab en 2018), se presenta el caso similar para municipios pequeños, donde se hace un ajuste por el modelo inclinándose mas hacia la información a priori dada previamente por el modelo Bayesiano. En este caso también hay una variación mayor en el grupo de 80 y más

Figura 10. Tasas de mortalidad observadas y estimadas por diabetes mellitus (en logaritmo) para mujeres en nueve ciudades de Colombia, 2017.

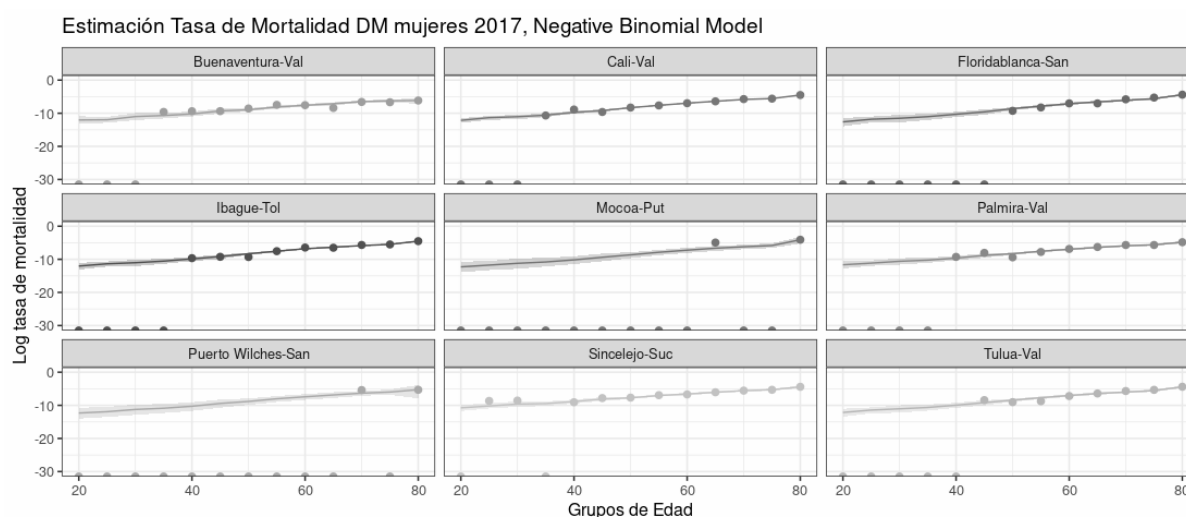


Fuente: Cálculos propios. **Nota.** Los puntos sobre la línea horizontal inferior de las gráficas representan el $\log(0) = -\infty$. El logaritmo de la ausencia de un registro da un negativo valor

muy pequeño.

Finalmente es importante destacar, en la figura 11, para el municipio de Puerto Wilches (31698 Hab en 2018), se presenta el mismo caso similar para municipios pequeños que ya se había mencionado, donde la información a priori del modelo es la que prima para el ajuste.

Figura 11. Tasas de mortalidad observadas y estimadas por diabetes mellitus (en logaritmo) para mujeres en nueve ciudades de Colombia, 2017.



Fuente: Cálculos propios. **Nota.** Los puntos sobre la línea horizontal inferior de las gráficas representan el $\log(0) = -\infty$. El logaritmo de la ausencia de un registro da un negativo valor muy pequeño.

Conclusiones

El estudio de las tasas específicas de mortalidad para una causa específica de mortalidad, aun cuando es de baja prevalencia como la diabetes mellitus, por edad y sexo en áreas subnacionales es una necesidad para los países en vía de desarrollo que han alcanzado un estado avanzado en el proceso de envejecimiento poblacional, con la consecuente transición epidemiológica hacia enfermedades no transmisibles específicamente para la tercera edad. Sin embargo, hay dos grandes problemas al momento de hacer este tipo de vigilancia en mortalidad: las cifras de los registros de defunciones para áreas o poblaciones pequeñas sufren de problemas de subregistro, por tanto, se reporta una gran cantidad de ceros en todos los grupos de edad, sobre todo en las edades más jóvenes, esto último es esperado

debido a que esta es una enfermedad crónica que afecta en mayor proporción a las personas adultas. Pero lo más grave es que en la mayoría de los países en vía de desarrollo, no es claro evidenciar si estos ceros son ausencias de registros o de información en las bases de datos de estadísticas vitales.

De allí, que se hace importante contar con un modelo de estimación para las tasas de mortalidad específicas para enfermedades de baja prevalencia, es decir con gran presencia de ceros y por ello se generó una transformación de la propuesta del modelo jerárquico bayesiano de Alexander, Zagheni and Barbieri, 2017. Se ha evidenciado, que este último, no resuelve completamente el problema de capturar la variabilidad ocasionada por el exceso de ceros. Por esta razón, se proponen en esta investigación ajustes a la variación mediante dos modelos de regresión: el modelo Poisson cero inflado y el modelo de regresión Binomial Negativo. Esta última alternativa no es la única solución a este problema, existen otras alternativas más que utilizaron información georreferenciada y aprovechan la información de vecindades, modelos de regresión entre otras (Lima and Queiroz, 2014; Gonzaga and Schmertmann, 2016; Schmertmann and Gonzaga, 2018; Montoya-Betancur *et al.*, 2020), no obstante, estas alternativas no son una opción metodológica desde la propuesta de Alexander, Zagheni and Barbieri, 2017.

Nuestra propuesta utiliza como base un sistema de componentes principales, para el cual se utilizaron las tasas específicas de mortalidad por diabetes mellitus nacionales, por edad y sexo de Argentina, Brasil, Colombia y México y se aplicaron a 36 ciudades de Colombia, usando la información de los microdatos de los registros de defunciones oficiales. El modelo presentó una limitación particular y es que el patrón encontrado permite corregir solo los valores nulos o muy cercanos al cero. A pesar de esta limitación el ajuste final mostrado por el modelo resultó ser una buena opción para la estimación del subregistro de mortalidad, en causas específicas de mortalidad por edad y sexo para grandes pequeñas áreas. En efecto, bajo este enfoque de interpretación, el modelo no solo permite una alternativa de medición para estos puntos con tan bajos niveles de reporte, sino que también evidencia que las grandes ciudades tienen menos subregistro que las ciudades intermedias o pequeñas y que este subregistro mejora en el tiempo.

El modelo también identificó visiblemente las diferencias en los patrones de mortalidad por DM entre hombres y mujeres, siendo superior para los primeros, como es lo

esperado(Ares Blanco *et al.*, 2020). Adicionalmente, se observó que en la mayoría de las ciudades estudiadas hubo una disminución de la mortalidad por DM conforme aumentaban los años. Sin embargo, esta disminución no fue tan clara para las pequeñas ciudades, donde primó la información a priori del modelo BN.

Las estimaciones aquí presentadas podrán servir como una base para poder focalizar esfuerzos que ayuden a generar políticas de intervención con por ejemplo la disminución del consumo de bebidas azucaradas o alimentos ultra procesados que permitan la disminución de la mortalidad por DM(Wang *et al.*, 2015; Ortún Rubio, Lopez Valcarcel and Pinilla, 2016), sobre todo en hombres, en países en vía de desarrollo, con problemas de subregistro en la mortalidad. El modelo mostró algunas limitaciones en cuanto a la estimación de la variabilidad, en los extremos de las edades, sobre todo para pequeños municipios, posiblemente por la categorización de la edad. Adicionalmente el modelo BN, es conocido por presentar mayores intervalos en las estimaciones de los parámetros. No obstante, nuestro trabajo ha sido una buena aproximación a corrección de subregistros para pequeñas áreas, en una área donde aún falta mucho por explorar, sobre todo con métodos contemporáneos de estadística, y se esperan futuros esfuerzos que permitan generar mejores estimaciones para los procesos de vigilancia epidemiológica efectiva en enfermedades de baja prevalencia sobre todo en adultos mayores, que propenda por la mejora de la calidad de vida de los habitantes de ciudades grandes y pequeñas de los países en vía de desarrollo de las Américas.

CONFLICTOS DE INTERÉS

Todos los autores declaran no tener conflictos de interés en relación con esta investigación.

DECLARACIÓN DE CONTRIBUCIÓN DE LOS AUTORES

B.P.U.C y J.J.S contribuyeron con las ideas, formulación o evolución de las metas y objetivos generales de la investigación y llevaron a cabo un proceso de investigación, específicamente realizando los experimentos o la recopilación de datos/evidencia y contribuyeron a los análisis e interpretaciones. J.J.S y P.A.R.S contribuyeron con la aplicación de técnicas estadísticas, matemáticas, computacionales u otras técnicas formales para analizar o sintetizar datos del estudio.

REFERENCIAS BIBLIOGRÁFICAS

Agresti, A. (2012) *Categorical data analysis*. 2nd edn. New York: Wiley (Probability and statistics, pag. 559-563).

Agudelo-Botero, M. and Dávila-Cervantes, C. A. (2015) ‘Carga de la mortalidad por diabetes mellitus en América Latina 2000-2011: los casos de Argentina, Chile, Colombia y México’, *Gaceta Sanitaria*, 29(3), pp. 172–177. doi: 10.1016/j.gaceta.2015.01.015.

Alegre-Díaz, J. *et al.* (2016) ‘Diabetes and Cause-Specific Mortality in Mexico City’, *New England Journal of Medicine*, 375(20), pp. 1961–1971. doi: 10.1056/nejmoa1605368.

Alexander, M., Zagheni, E. and Barbieri, M. (2017) ‘A Flexible Bayesian model for estimating subnational mortality’, *Demography*, 54(6), pp. 2025–2041. doi: 10.1007/s13524-017-0618-7.

Ares Blanco, J. *et al.* (2020) ‘Diferencias de género en la mortalidad de personas con diabetes tipo 2: Estudio Asturias 2018’, *Gaceta Sanitaria*, 34(5), pp. 442–448. doi: 10.1016/j.gaceta.2019.02.014.

Arnold-Domínguez, Y. *et al.* (2013) ‘Mortalidad por causa básica de diabetes mellitus en Cuba, 2000-2009’, *Rev. peru. epidemiol. (Online)*, 17, pp. 1–6. Available at: <https://www.redalyc.org/articulo.oa?id=203128542004>.

Aschner, P. (2010) ‘Epidemiología de la diabetes en Colombia’, *Avances en diabetología*, 26(2), pp. 95–100. doi: [https://doi.org/10.1016/S1134-3230\(10\)62005-4](https://doi.org/10.1016/S1134-3230(10)62005-4).

Barquera, S. *et al.* (2013) ‘Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy’, *Globalization and Health*, 9(3), pp. 1–9. doi: <https://doi.org/10.1186/1744-8603-9-3>.

Bay, G. and Orellana, H. (2007) *La calidad de las estadísticas vitales en la América Latina (versión preliminar para discusión)*. Santiago de Chile. Available at: <https://repositorio.cepal.org/bitstream/handle/11362/21394/S2007053.pdf?sequence=1>.

Blue, L., Goldman, N. and Rosero-Bixby, L. (2015) ‘Enfermedad y pérdida de peso: un estudio prospectivo de adultos y adultos mayores en Costa Rica e Inglaterra’, *Salud Pública de México*, 57(4), pp. 312–319. Available at: http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S0036-36342015000400009&lng=es&nrm=iso&tlng=en.

Bryant, J. R. and Graham, P. J. (2013) 'Bayesian demographic accounts: Subnational population estimation using multiple data sources', *Bayesian Analysis*, 8(3), pp. 591–622. Available at: <http://projecteuclid.org/euclid.ba/1378729921>.

Buitelaar, R. M. *et al.* (2015) *La comparabilidad de las estadísticas territoriales en América Latina*. 1st edn. Edited by CEPAL. Santiago de Chile: Publicación de las Naciones Unidas. Available at: https://repositorio.cepal.org/bitstream/handle/11362/37793/1/S1421051_es.pdf.

Curcio, C.-L. *et al.* (2018) 'Successful Aging in Colombia: The Role of Disease', *Gerontology and Geriatric Medicine*, 4, p. 2333721418804052. doi: 10.1177/2333721418804052.

DANE, C. (2021a) *Archivo Nacional de Datos, Estadísticas Vitales - EEVV - 1998 -2019*. Available at: https://microdatos.dane.gov.co/catalog/MICRODATOS/about_collection/22 (Accessed: 20 July 2021).

DANE, C. (2021b) *Archivo Nacional de Datos, Censo General de Población*. Available at: https://microdatos.dane.gov.co/catalog/MICRODATOS/about_collection/2 (Accessed: 20 July 2021).

DANE, C. (2021c) *Resultados Censo Nacional de Población y Vivienda 2018, ¿Dónde estamos?* Available at: <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivienda-2018/donde-estamos> (Accessed: 20 August 2021).

Ganguli, M. *et al.* (2005) 'Alzheimer disease and mortality: a 15-year epidemiological study', *Archives of Neurology*, 62(5), pp. 779–784. doi: 10.1001/archneur.62.5.779.

Ganguli, M. *et al.* (2006) 'Discrepancies in information provided to primary care physicians by patients with and without dementia: the Steel Valley Seniors Survey', *The American journal of geriatric psychiatry*, 14(5), pp. 446–455. doi: 10.1097/01.JGP.0000199340.17808.77.

Gelman, A. *et al.* (2013) *Bayesian data analysis*. 3rd editio, *Chapman & Hall/CRC Texts in Statistical Science*. 3rd editio. USA: Taylor & Francis. Available at: <http://www.stat.columbia.edu/~gelman/book/BDA3.pdf>.

Gonzaga, M. R. and Schmertmann, C. P. (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), pp. 629–652. doi:

<http://dx.doi.org/10.20947/S0102-30982016c0009>.

González, J. C., Walker, J. H. and Einarson, T. R. (2009) ‘Cost-of-illness study of type 2 diabetes mellitus in Colombia’, *Revista Panamericana de Salud Pública*, 26(1), pp. 165–174.

doi: 10.1590/S1020-49892009000700009.

Hamid Zargar, A. *et al.* (1999) ‘Mortality in diabetes mellitus data from a developing region of the world’, *Diabetes Research and Clinical Practice*, 43(1), pp. 67–74. doi:

10.1016/S0168-8227(98)00112-0.

Horiuchi, S. (1999) ‘Epidemiological Transitions in Human History’, in Affairs, D. of E. and S. and Division, P. (eds) *Health and Mortality: Issues of Global Concern, Proceedings of the Symposium on Health and Mortality, Brussels, 19-22 November 1997*. 1st edn. New York:

United Nations publication, pp. 54–71. Available at:

<https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/d>

[documents/2020/Jan/un_1999_health_and_mortality_issues_of_global_concern.pdf](https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/d).

International Diabetes Federation (2019) *IDF DIABETES ATLAS, Edición (9th ed.)*.

Available at: <https://diabetesatlas.org/atlas/ninth-edition/> (Accessed: 23 August 2020).

Johnson, R. and Wichern, D. (2007) *Applied multivariate statistical analysis*. 6th edn,

Pearson. 6th edn. New York: Upper Saddle River.

Kirigia, J. M. *et al.* (2009) ‘Economic burden of diabetes mellitus in the WHO African region’, *BMC International Health and Human Rights*, 9(1). doi: 10.1186/1472-698X-9-6.

Klafke, A. *et al.* (2015) ‘The decline in mortality due to acute complications of diabetes mellitus in Brazil, 1991-2010’, *BMC Public Health*, 15(1). doi: 10.1186/s12889-015-2123-5.

Kudva, Y. C. (2020) *Diabetes - Síntomas y causas, Mayo Clinic: Definición de la Diabetes*.

Available at: <https://www.mayoclinic.org/es-es/diseases-conditions/diabetes/symptoms-causes/syc-20371444> (Accessed: 18 May 2021).

Laurenti, R., Marcondes Fonseca, L. A. and Da Costa, M. L. (1982) ‘Mortality due to diabetes mellitus in the city of Sao Paulo (Brazil). Its evolution over a period of 79 years (1900-1978) and analysis of some associated causes’, *Revista de Saude Publica*, 16(2), pp.

77–91. doi: 10.1590/s0034-89101982000200003.

Lee, R. (2003) ‘The Demographic Transition: Three Centuries of Fundamental Change’, *The*

Journal of Economic Perspectives, 17(4), pp. 167–190. doi: 10.1257/089533003772034943.

Lee, R. D. (2007) *Global population aging and its economic consequences*. First Ed. Washington DC: AEI Press.

Lee, R. D. and Carter, L. R. (1992) ‘Modeling and forecasting U.S. mortality’, *Journal of the American Statistical Association*, 87(419), pp. 659–671. doi: 10.1080/01621459.1992.10475265.

Lima, E. C. and Queiroz, B. L. (2014) ‘Evolution of the deaths registry system in Brazil: associations with changes in the mortality profile, under-registration of death counts, and ill-defined causes of death’, *Cadernos de Saúde Pública*, 30(8), pp. 1721–1730. doi: 10.1590/0102-311X00131113.

Montoya-Betancur, K. V. *et al.* (2020) ‘Exploratory spatial analysis of diabetes mortality and its relationship with the socioeconomic conditions of Colombian municipalities’, *Cadernos de Saúde Pública*, 36(6). doi: 10.1590/0102-311x00101219.

Organización Panamericana de la Salud (2017) *Lineamientos básicos para el análisis de la mortalidad*. Washington, D.C.: OPS. Available at: <https://iris.paho.org/bitstream/handle/10665.2/34492/9789275319819-spa.pdf?sequence=7>.

Ortún Rubio, V., Lopez Valcarcel, B. G. and Pinilla, J. (2016) ‘El impuesto sobre bebidas azucaradas en España’, *Revista española de salud pública, ISSN-e 1135-5727, N° 90, 2016*, 90(90), p. 7. Available at: <https://dialnet.unirioja.es/servlet/articulo?codigo=7020803&info=resumen&idioma=ENG> (Accessed: 24 August 2022).

Palloni, A., Pinto, G. and Beltran-Sanchez, H. (2015) *Two Centuries of Mortality Decline in Latin America : From Hunger to Longevity*.

Queiroz, B. L. *et al.* (2017) ‘Completeness of death-count coverage and adult mortality (45q15) for Brazilian states from 1980 to 2010’, *Revista Brasileira de Epidemiologia*, 20, pp. 21–33. doi: 10.1590/1980-5497201700050003.

R Core Team (2021) *R: A language and environment for statistical computing.*, R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.r-project.org/> (Accessed: 10 March 2022).

Rhon-Bunshi, M. del R. (2016) *Análisis de los errores y omisiones de registro existentes en*

los formularios de defunción general que determinan sub-registro de muerte materna correspondientes al periodo 2013 – 2015. Universidad San Francisco de Quito. Available at: <https://repositorio.usfq.edu.ec/bitstream/23000/6011/1/129431.pdf> (Accessed: 18 May 2021).

Ribotta, B. S., Acosta, L. M. S. and Bertone, C. L. (2019) ‘Evaluaciones subnacionales de la cobertura de las estadísticas vitales. Estudios recientes en América Latina’, *Gerencia y Políticas de Salud*, 18(36), pp. 1–59. doi: <https://doi.org/10.11144/Javeriana.rgps18-36.esce>.

Saad, P. M. (2011) ‘Demographic trends in Latin America and the Caribbean’, in Cotlear, D. (ed.) *Population Aging Is Latin America Ready? Is Latin America Ready?*. The World. Santiago de Chile: World Bank Publications, pp. 43–75. doi: 10.1596/978-0-8213-8487-9.

Sandoval, J. de J., Turra, C. M. and Loschi, R. H. (2019) ‘Tasas corregidas de mortalidad atribuible a la demencia por la enfermedad de Alzheimer, Brasil, 2009-2013’, *Cadernos de saude publica*, 35(6), p. e00091918. doi: 10.1590/0102-311X00091918.

Schmertmann, C. P. and Gonzaga, M. R. (2018) ‘Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records’, *Demography*, pp. 1–26. Available at: <https://link.springer.com/article/10.1007/s13524-018-0695-2>.

Urdinola, B. P., Torres Avilés, F. and Velasco, J. A. (2017) ‘The homicide atlas in Colombia: contagion and under-registration for small areas’, *Cuadernos de Geografía: Revista Colombiana de Geografía*, 26(1), pp. 101–118. doi: <https://doi.org/10.15446/rcdg.v26n1.55429>.

Vargas-Uricoechea, H. and Casas-Figueroa, L. Á. (2016) ‘Epidemiología de la diabetes mellitus en Sudamérica: la experiencia de Colombia’, *Clínica e Investigación en Arteriosclerosis*, 28(5), pp. 245–256. doi: <https://doi.org/10.1016/j.arteri.2015.12.002>.

Wang, M. *et al.* (2015) ‘Association between sugar-sweetened beverages and type 2 diabetes: A meta-analysis’, *Journal of Diabetes Investigation*, 6(3), p. 360. doi: 10.1111/JDI.12309.

Winkelmann, R. (2008) *Econometric analysis of count data*. 1st edn. Springer Science & Business Media. doi: <https://doi.org/10.1007/978-3-540-78389-3>.

Wong, L. L. R. and Carvalho, J. A. (2006) ‘The rapid process of aging in Brazil: serious challenges for public policies’, *Revista Brasileira de Estudos de População*, 23(1), pp. 5–26. doi: <https://doi.org/10.1590/S0102-30982006000100002>.

World Health Organization (2018) *Mortality and global health estimates*, *Global Health*

Estimates: Life expectancy and leading causes of death and disability. Available at:
<https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates> (Accessed:
15 June 2018).

Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.