Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL

(FOR EVALUATION OF THE ACT DOCTORAL THESIS)

Año académico (*academic year):* 2019/20

DOCTORANDO *(candidate PHD)*: **SÁNCHEZ DÍAZ, GERMÁN**
D.N.I./PASAPORTE (*Id.Passport*): ******6027v**
PROGRAMA DE DOCTORADO *(Academic Committee of the Programme)*: **D444-TECNOLOGÍAS DE LA INFORMACIÓN GEOGRÁFICA**
DPTO. COORDINADOR DEL PROGRAMA *(Department)*: **GEOLOGÍA, GEOGRAFÍA Y MEDIO AMBIENTE**
TITULACIÓN DE DOCTOR EN *(Phd title)*: **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 27/11/19, reunido el tribunal de evaluación, constituido por los miembros que suscriben el presente Acta, el aspirante defendió su Tesis Doctoral **con Mención Internacional** (*In today assessment met the court, consisting of the members who signed this Act, the candidate defended his doctoral thesis with mention as International Doctorate),* elaborada bajo la dirección de (*prepared under the direction of)* FRANCISCO JAVIER ESCOBAR MARTÍNEZ // VERÓNICA ALONSO FERREIRA**.**

Sobre el siguiente tema *(Title of the doctoral thesis)*: *AN EPIDEMIOLOGICAL AND GIS-BASED ANALYSIS OF MORTALITY IN SELECTED RARE DISEASES*

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL[1] de **(no apto, aprobado, notable** y **sobresaliente)** *(After the defense and defense of the thesis, the court agreed to grant the GLOBAL RATING (fail, pass, good and excellent)*: SOBRESALIENTE

Alcalá de Henares, a 27 de Noviembre de 2019

Fdo. *(Signed)*: JAVIER SOLAS

Fdo. *(Signed)*: DIANA GÓMEZ

Fdo. *(Signed)*: LUISA A. BORRELL

FIRMA DEL ALUMNO (candidate's signature),

Fdo. *(Signed)*: GERMÁN SÁNCHEZ

Con fecha 16 de diciembre de 2019 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

☒ Conceder la Mención de "Cum Laude"
☐ No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada

[1] La calificación podrá ser **"no apto" "aprobado" "notable"** y **"sobresaliente"**. El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad. (*The grade may be* ***"fail"*** *"****pass****"* *"****good****"* *or "****excellent****". The panel may confer the distinction of "cum laude" if the overall grade is "Excellent" and has been awarded unanimously as such after secret voting.).*

INCIDENCIAS / OBSERVACIONES:
(Incidents / Comments)

Universidad de Alcalá

ESCUELA DE DOCTORADO
Servicio de Estudios Oficiales de Posgrado

DILIGENCIA DE DEPÓSITO DE TESIS.

Comprobado que el expediente académico de D./Dª GERMÁN SÁNCHEZ DÍAZ reúne los requisitos exigidos para la presentación de la Tesis, de acuerdo a la normativa vigente, y habiendo presentado la misma en formato: ☒ soporte electrónico ☐ impreso en papel, para el depósito de la misma, en el Servicio de Estudios Oficiales de Posgrado, con el nº de páginas: 202 se procede, con fecha de hoy a registrar el depósito de la tesis.

Alcalá de Henares a 18 de JULIO de 2019

UNIVERSIDAD DE ALCALÁ SERVICIO DE ESTUDIOS OFICIALES DE POSGRADO

PILAR DE LA VEGA GALLEGO

Fdo. El Funcionario

Programa de Doctorado en

Tecnologías de la Información Geográfica

An epidemiological and GIS-based analysis of mortality in selected Rare Diseases

Tesis Doctoral presentada por

GERMÁN SÁNCHEZ DÍAZ

2019

Universidad de Alcalá

UNIVERSIDAD DE ALCALÁ, PATRIMONIO DE LA HUMANIDAD

D/Dª Francisco ESCOBAR MARTÍNEZ, Coordinador de la Comisión Académica del Programa de Doctorado en Tecnologías de la Información Geográfica

HAGO CONSTAR que la Tesis Doctoral titulada AN EPIDEMIOLOGICAL AND GIS-BASED ANALYSIS OF MORTALITY IN SELECTED RARE DISEASES, presentada por D/Dª **Germán SÁNCHEZ DÍAZ**, bajo la dirección del / de la Dr/a. Francisco Escobar Martínez y Verónica Alonso Ferreira, ha sido realizada por compendio de artículos, reuniendo los requisitos exigidos a este tipo de tesis, así como los requisitos científicos de originalidad y rigor metodológicos para ser defendida ante un tribunal. Esta Comisión ha tenido también en cuenta la evaluación positiva anual del doctorando, habiendo obtenido las correspondientes competencias establecidas en el Programa.

Para que así conste a los efectos del depósito de la tesis, se firma en Alcalá de Henares a 12 de julio de 2019

Francisco Escobar Martínez

Firmado digitalmente por Francisco Escobar Martínez
Nombre de reconocimiento (DN): cn=Francisco Escobar Martínez, o=Universidad de Alcalá, ou=UD Geografía, email=francisco.escobar@uah.es, c=ES
Fecha: 2019.07.15 09:23:08 +02'00'

Fdo.: _____

Programa de Doctorado en

Tecnologías de la Información Geográfica

An epidemiological and GIS-based analysis of mortality in selected Rare Diseases

Tesis Doctoral con mención internacional presentada por

GERMÁN SÁNCHEZ DÍAZ

Dirigida por:

Dr. Francisco Javier Escobar Martínez

Dra. Verónica Alonso Ferreira

2019

This doctoral thesis was possible thanks to the funding of the Ministerio de Educación Cultura y Deporte under the program *Ayudas para contratos predoctorales de Formación del Profesorado Universitario* (FPU14/03914).

This doctoral thesis was developed within the research project *Mortalidad atribuible a enfermedades raras en España 1981-2010. Sistema de información interactivo de la variabilidad temporal y geográfica* (TPY 1238/15, AESI 2014, ISCIII).

Publications arising from this Doctoral Thesis

This doctoral thesis consists of a compendium of articles, three of which have already been published, while a fourth has been submitted. Either way, the doctoral student was the first author of the manuscripts.

1. Sánchez-Díaz G, Alonso-Ferreira V, Posada de la Paz M & Escobar F. (submitted on July 2019). Multi-scale aggregation data displaying rare-disease epidemiology. *Statistical Methods in Medical Research*.

2. Sánchez-Díaz G, Arias-Merino G, Villaverde-Hueso A, Morales-Piga A, Abaitua-Borda I, Hens M, Bermejo-Sánchez E, Posada de la Paz M & Alonso-Ferreira V. (2016). Monitoring Huntington's disease mortality across a 30-year period: Geographic and temporal patterns. *Neuroepidemiology, 47*:155-163. doi: 10.1159/000452860.

3. Sánchez-Díaz G, Escobar F, Villaverde-Hueso A, Posada de la Paz M & Alonso-Ferreira V. (2019). Temporal and cartographic analyses of the distribution within Spain of mortality due to granulomatosis with polyangiitis (1984–2016*). International Journal of Environmental Research and Public Health, 16*, 1388. doi: 10.3390/ijerph16081388.

4. Sánchez-Díaz G, Escobar F, Badland H, Arias-Merino G, Posada de la Paz M & Alonso-Ferreira V. (2018). Geographic analysis of motor neuron disease mortality and heavy metals released to rivers in Spain. *International Journal of Environmental Research and Public Health, 15*, 2522. doi: 10.3390/ijerph15112522.

Acknowledgements

En este momento que supone la culminación de esta tesis doctoral quiero dedicar unas líneas a todas aquellas personas que han formado parte de este proceso o que me han apoyado para conseguirlo.

En primer lugar quiero agradecer la labor realizada durante estos años por mis directores de tesis, Vero y Patxi. El que pertenezcáis cada uno a campos de investigación totalmente diferentes, y que cada uno haya contribuido desde su experiencia ha supuesto para mí un aprendizaje profesional enorme. Valoro el esfuerzo que supone conseguir acuerdos y el respeto por las ideas el uno del otro. ¡Así da gusto trabajar en equipo!

De forma individual quiero agradecer a Vero haber apostado por mí, creo que tenía más confianza en lo que podría llegar a hacer en el futuro que yo mismo. Quién me iba a decir que esa beca del CIBER de un año iba a abrirme las puertas para trabajar en el IIER durante todos estos años e incluso acabar terminando una tesis doctoral. Al estar trabajando codo con codo en el Instituto, has sido la que más has padecido mis preguntas absurdas, cabezonerías y despistes. Me quedo con la paciencia que has tenido conmigo durante estos años y todas las cosas que he aprendido de ti sobre epidemiología y enfermedades raras pero sobre todo por lo pragmática que eres y como simplificas las cosas.

A Patxi también tengo que agradecer tanto como a Vero, empezando desde el momento que se ofreció a codirigir esta tesis. Estoy muy contento de haber tenido a un director que realmente cree en lo que es capaz de hacer la geografía en otros campos de investigación. Gracias por la libertad que me has dado para poder organizar la docencia que he impartido en la universidad y todos los consejos que me has dado, fueron muchos nervios al principio pero es algo muy gratificante. Ahora entiendo que lleves tanto tiempo siendo profesor, ¡es que engancha mucho!

Vuelvo de nuevo a dirigirme a ambos directores. El proyecto del atlas de mortalidad de enfermedades raras que está finalizándose y en el cual he tenido el privilegio de participar es un orgullo para mí. Gracias por tener en cuenta mi opinión y criterio pese a mi inexperiencia en este campo. Aún con todas las dificultades surgidas, ha sido un trabajo ingente para las pocas personas que formamos el equipo pero merecerá la pena por lo todo lo que hemos aprendido durante su elaboración así como por lo que aprenderán otras personas cuando puedan contemplar el trabajo realizado.

También quiero agradecer a todos los que me han acompañado durante mi periodo de elaboración de tesis en el IIER, con los que he convivido día a día, tanto por los que están

hoy como por los que estuvieron. Me acogisteis maravillosamente en el centro desde el primer día que llegué y habéis sido muy importantes en lo académico y/o personal: Ana Villaverde, Paloma, Marian, Manuela, Ignacio, Antonio Morales, Manuel Hens, Manuel Posada, Nieves, Ana Carbonero, Enrique, y seguro que me dejo alguno. Especialmente quiero agradecer a Greta por toda la ayuda que he recibido y lo que he aprendido de ella de epidemiología. Hemos hecho buen tándem de epidemiología/mapas. Espero que podamos volver a trabajar juntos.

Many thanks also to Hannah Badland and to the Healthy Liveable Cities Group team. Thanks for hosting me in the wonderful city of Melbourne during 3 months at the RMIT University. You were nice to me in both personal and professional terms, it was a joy to enjoy the Australian hospitality.

En todo este camino a la tesis doctoral ha sido un pilar muy importante los hogares de Madrid donde he vivido estos cinco años, sobre todo por la compañía para los momentos tanto buenos como reguleros. Desde Malasaña, La Latina y ahora Retiro. Que buenos momentos hemos compartido en la terraza (la envidia de todo el mundo): Alicia, Alejandro, Estefanía, Mari Pili, Elena, Clémence, Pedro… Aparte de mis compañeros/amigos de piso también a toda la gente que vive en Madrid, o ha vivido, y que me han hecho pasar tan buenos momentos en esta ciudad: una paella, una tarta desecha, una fiesta andaluza o simplemente compañía cuando ha hecho falta.

También dedico este trabajo a mis amigos del pueblo, los *kalumnianos*. Aunque los caminos de la vida hacen que cada vez sea más difícil juntarnos (menos mal que algunos estáis en Madrid), las cervezas en el Aloha, las vueltas con el coche por la sierra o las montañas que quedan por subir no van a terminar nunca.

También tengo que hacer mención a la enorme colaboración altruista realizada por amigos que me han echado un cable con el idioma y que seguro que al final han aprendido unas cuantas cosas sobre enfermedades raras. ¡Muchas gracias Teresa e Inés!

Por supuesto también le tengo que agradecer toda mi familia del pueblo, de Elche de la Sierra, por el apoyo que me han dado siempre desde que comencé esta aventura académica. Muchas gracias por dejarme seguir siempre mis vocaciones, está claro que la tontería de los mapas desde tan pequeño iba a acabar dando sus frutos. ¡Le dedico está tesis a mis padres, a mis hermanos y a mis sobrinos!

Muchas gracias a todos.

Table of contents

List of Abbreviations

AC – Autonomous Community

ADR – Annual Death Registry

ALS – Amyotrophic lateral sclerosis

ANCAs - Anti-neutrophil cytoplasmic antibodies

APC – Annual Percentage Change

BioNER – Biobanco Nacional de Enfermedades Raras

CI – Confidence Interval

CIBER – Centro de Investigación Biomédica en Red

CIBERER – Centro de Investigación Biomédica en Red de Enfermedades Raras

E-PRTR – European Pollutant Release and Transfer Register

EJP-RD – European Joint Programme on Rare Diseases

EU – European Union

GIS – Geographic Information Systems

GIT – Geographic Information Technologies

GPA – Granulomatosis with polyangiitis

HD – Huntington's disease

ICD – International Classification of Diseases

IGU - International Geographical Union

IIER – Instituto de Investigación de Enfermedades Raras

IRDiRC – International Rare Diseases Research Consortium

IRR - Incidence Report Rates

ISCIII – Instituto de Salud Carlos III

LISA - Local indicators of spatial association

MA – Metropolitan Area

MAFF - Ministry for Agriculture, Fisheries and Food

MAUP – Modifiable areal unit problem

MND – Motor neuron diseases

MRI – Municipal Register of Inhabitants

NGIC - National Geographic Information Center

NMA - National Mapping Agency

NSI – National Statistics Institute

NUTS - Nomenclature of Territorial Units for Statistics

PP – Posterior Probability

PWPS - People Working in the Primary Sector

RD – Rare Disease

SMR – Standardized Mortality Ratio

STD - Standard Deviation

WHO – World Health Organization

Abstract

Introduction

Rare diseases (RD) are defined as those affecting no more than 5 in 10,000 people within the European Union (EU). It is estimated that between 27 and 36 million people might be affected by RD in the EU (which amounts to between 6 and 8% of its total population), with them consequently having become a Public Health issue. The epidemiological research of the over 6,000 accounted for RD proves harder than that of more frequent diseases, due to their specific characteristics (e.g. heterogeneity, low number of cases, case dispersion or difficulties in codification, to name a few). When it comes to mortality, the availability of national population-based data and standardized registries are very useful when researching RD, since the reduction of mortality is a public health policy goal. Geographic Information Systems allowing for the integration, analysis and representation of geospatial data have proved to be a valuable tool with which to study the mortality distribution and its spatio-temporal evolution. Knowing the geographical patterns of mortality and their relationships with other spatial variables helps more accurately describe its epidemiology, as well as providing important baseline elements when searching for causal relationships between environmental factors and the diseases object of analysis.

Objective

The general aim of this doctoral thesis is to deepen the knowledge on mortality attributed to RDs in Spain from a geographical and epidemiological perspective. The specific objectives are the following:

Objective 1: To identify difficulties related to the optimal geographic units of representation when working with RDs and to provide sound advice for the election of the most appropriate unit, aiming for a better epidemiological and cartographic result.

Objective 2: To evaluate the Spanish population's time trends in mortality due to RDs over three decades. This target will be achieved by selecting Huntington's disease (HD) and Granulomatosis with polyangiitis (GPA).

Objective 3: To identify geographic patterns in mortality attributed to RDs, in order to trace variations in mortality among spatial units throughout Spain. This objective will be tackled through the analysis of GPA, HD and Motor neuron diseases (MND) as the RDs of reference.

Objective 4: To describe the characteristics of heavy metal emissions to river basins, and to explore the possible correlation that exists between their location and the mortality attributed to MND, in order to help understand the etiology of this group of diseases.

Methodology

This doctoral thesis has the Spanish territory as its study framework. It covers the years 1984 to 2016, at maximum, and uses the annual death registry provided by the National Statistics Institute as the basis for analysis. The decease-related data relevant to each of the three analyzed RDs were selected taking into account the underlying cause of death, whilst data relative to the heavy metal emissions to river basins were collected from the European Pollutant Release and Transfer Register (E-PRTR). Shapefiles containing information on geographic units and hydrography were downloaded from the National Mapping Agency.

As a starting point, the suitability of the available aggregation units (municipalities, districts and provinces) was analyzed, by means of the epidemiological and geospatial comparison of different indicators, such as the Standardized Mortality Ratio (SMR), smoothed SMR and local indicators of spatial association (LISA).

Secondly, regarding epidemiological mortality indicators, time trends in HD and GPA were analyzed by calculating age-adjusted mortality rates, and subsequently smoothed. Time trends were assessed by a *joinpoint* regression. As to the geographical distribution of mortality for the three selected diseases, the SMRs and their 95% confidence intervals were calculated. SMR is defined as the existing ratio between observed and expected deaths within a geographical unit. Smoothed SMR and its posterior probability were also calculated. The smoothing process takes into account the observed and expected values of the neighboring geographical units and helps identify spatial patterns of death risk.

Lastly, the sites that release heavy metal emissions into the rivers were geocoded. In order to study their influence on MND deaths, we took exposed municipalities to be those downstream from the point of emission, within 20 km of it. The mortality of the exposed municipalities was compared to that of the non-exposed ones through a Poisson regression model. The obtained epidemiological results of a spatial nature were represented cartographically.

Results

The modifiable areal unit problem (MAUP) proved relevant when analyzing different aggregation units for the study of RDs, since significant differences were found when comparing epidemiological indicators among the three existing units. The number of cases occurring due to disease was a determining factor when choosing the most appropriate level of aggregation. The district level (*comarca*) has proved the most appropriate to display the variability of mortality in Spain, while still displaying relevant spatial detail.

Regarding the evolution of mortality rates for the analyzed RDs, significant time trends were identified in both HD and GPA within the periods of study. The mortality rate from HD increased annually by 3.44% ($p < 0.001$) throughout, while that for GPA displayed an annual increase of 20.6% between 1984 and 1992, with a 1.91% annual increase thereafter. ($p < 0.05$). The average ages of death from GPA and EH rose by 0.78% and 0.59% ($p < 0.05$), respectively, within the studied timeframe.

As for the spatial analysis of RD-related mortality, the three reference diseases showed geographical variability within the Spanish territory. Regarding HD, higher SMRs and smoothed SMRs were found in some districts to the Southwest of the peninsula. GPA posed an increased death risk for the inhabitants of two specific districts in the North of Spain, and for a further two in the South, while in the case of MND, a comparison between municipalities showed those located in the North to be at a higher risk of death from the disease than those to the South.

The Poisson regression analysis conducted on the variation of mortality attributed to MND in relation to the emission of heavy metals to river basins showed that MND mortality rates were 18.4% higher in the exposed municipalities than it was in non-exposed ones. The figure remained relevant both globally and for the analyses carried out individually for each type of metal.

Conclusions

RDs are of increasing interest due to the high impact they have on affected patients, their families and society in general. Therefore, it is necessary to improve the information available on them through research, carried out over multidisciplinary perspectives. This doctoral thesis has enriched the body of knowledge on spatial variability in mortality from RDs, by adding a geographical approach to its analysis. Some of the difficulties presented by the epidemiological study of these low prevalence diseases have been analyzed. The low number of cases and the need for spatial aggregation have the MAUP as a consequence. Relevant information regarding the temporal and geographic variability in mortality for

selected RDs has been provided, as well as having explored the possible associations with geographic factors. The results of this study offer very valuable information for health planning, since they support and justify the need for preventive and assistive action in those areas where there have been more inequalities in the risk of death due to these diseases. In addition, by establishing a correlation between mortality from RDs and suspect environmental factors, we offer clues for future studies that may delve into the discovery of RD causes.

Resumen

Introducción

Las Enfermedades Raras (ER) son aquellas que no superan los 5 casos por cada 10000 habitantes en la Unión Europea (UE). Se estima que podrían estar afectadas entre 27 y 36 millones de personas por ER en la UE (entre un 6 y un 8% de la población), por lo que se ha convertido en un problema de Salud Pública. Las características propias de este conjunto de más 6000 enfermedades (heterogeneidad, bajo número de casos, dispersión de casos, dificultades de codificación, entre otras) hacen más difícil su investigación epidemiológica. En el caso de la mortalidad, de la que existen datos a nivel nacional de base poblacional y estandarizada en su metodología de registro, su seguimiento es de gran utilidad ya que su reducción es un objetivo perseguido por las políticas de Salud Pública. Los Sistemas de Información Geográfica, gracias a su capacidad para la integración, análisis y representación de datos geoespaciales, se muestran como una valiosa herramienta para estudiar su distribución y evolución espacio-temporal. El conocimiento de patrones geográficos de la mortalidad y su relación con otras variables espaciales contribuyen a describir de forma más precisa su epidemiología. Además, permite aportar importantes elementos de base a la hora de establecer relaciones causales entre factores ambientales y las enfermedades analizadas.

Objetivos

El principal objetivo de esta tesis doctoral es profundizar en el conocimiento de la mortalidad debida a ER en España desde una perspectiva geográfica y epidemiológica. Los objetivos específicos se detallan a continuación:

Objetivo 1: Identificar las problemáticas actuales en la elección de la unidad geográfica para el trabajo con ER y proporcionar recomendaciones para escoger la unidad más apropiada buscando el mejor resultado epidemiológico y cartográfico.

Objetivo 2: Evaluar las tendencias temporales de la mortalidad debida a ER a lo largo de tres décadas en España. Para lograr este objetivo se seleccionarán las siguientes ER: Enfermedad de Huntington (EH) y Granulomatosis con poliangeitis (GPA).

Objetivo 3: Identificar los patrones geográficos de la mortalidad debida a ER para observar su variación entre las diferentes unidades geográficas en España. Este objetivo será desarrollado con las siguientes enfermedades: EH, GPA y enfermedades de la neuronas motoras (ENM).

Objetivo 4: Describir las características de los metales pesados emitidos a cuencas hidrográficas, así como explorar las posibles asociaciones entre su localización y la mortalidad atribuida a ENM con el fin de ayudar a entender su etiología.

Métodos

Esta tesis doctoral se circunscribe al territorio español y a un periodo máximo comprendido entre los años 1984 y 2016. Se utilizó como base del estudio la estadística de defunciones proporcionada de forma anual por el Instituto Nacional de Estadística. Se seleccionaron aquellos fallecimientos atribuidos a cada una de las tres ER, teniendo en cuenta la causa básica de defunción. Los datos de emisión de metales pesados a las cuencas hidrográficas, desde complejos industriales así como otras instalaciones, fueron recogidos del European Pollutant Release and Transfer Register (E-PRTR). Los ficheros de las unidades geográficas e hidrografía en formato shapefile fueron descargados del Instituto Geográfico Nacional.

En primer lugar se analizó la idoneidad de las unidades de agregación disponibles (municipios, comarcas y provincias) mediante la comparación epidemiológica y cartográfica de diferentes indicadores como la Razón de Mortalidad Estandarizada (RME), RME suavizada, así como indicadores locales de asociación espacial (LISA).

En cuanto a indicadores epidemiológicos de mortalidad, se analizó la evolución temporal de la EH y GPA mediante las tasas de mortalidad ajustada por edad (TAE) así como TAE suavizada. Las tendencias fueron evaluadas mediante una regresión *joinpoint*. Respecto a la distribución geográfica de la mortalidad para las tres enfermedades seleccionadas, se calcularon las RMEs y sus respectivos intervalos de confianza al 95%. La RME se define como el ratio entre los fallecimientos observados y los esperados en una unidad geográfica. También se calcularon las RMEs suavizadas y su probabilidad posterior. El proceso de suavizado tiene en cuenta los valores observados y esperados de las unidades geográficas vecinas y ayuda a encontrar patrones espaciales en el riesgo de fallecimiento.

Por último, se codificaron las instalaciones que habían emitido metales pesados a ríos. Para estudiar su influencia en los fallecimientos atribuidos a ENM, se consideraron municipios expuestos aquellos que estaban incluidos 20 kilómetros aguas abajo desde el punto de emisión, y se comparó la mortalidad con los no expuestos mediante un modelo de regresión de Poisson. Todos los resultados epidemiológicos de carácter espacial fueron representados cartográficamente.

Resultados

El Problema de la Unidad Espacial Modificable (PUEM) se mostró relevante al analizar diferentes unidades de agregación para el estudio de las ER, ya que se han encontrado diferencias significativas al comparar indicadores epidemiológicos entre las tres unidades de análisis. El número de casos por enfermedad es determinante para elegir el nivel de agregación, siendo el nivel comarcal en España el que mejor muestra la variabilidad de la mortalidad manteniendo un buen detalle espacial.

Respecto a la evolución de las TAE de mortalidad para las ER analizadas, se han observado cambios significativos en las dos enfermedades analizadas (EH y GPA). La TAE para la EH se incrementó anualmente un 3,44% ($p<0,001$) durante todo el periodo, mientras que en GPA se detectó un incremento anual del 20,6% desde 1984 hasta 1992 y, posteriormente un descenso del 1,91% ($p<0,05$). Tanto para GPA como EH, la edad media de fallecimiento aumentó un 0,78% y un 0,59% anual ($p<0,05$) durante sus respectivos periodos de estudio.

Respecto a la mortalidad espacial, las tres ER estudiadas mostraron variabilidad en el territorio español. En EH se encontraron RME y RME suavizadas más altas en algunas comarcas del suroeste peninsular. Para GPA se encontró un mayor riesgo de fallecimiento en dos comarcas del norte y dos del sur. En el caso de ENM, se encontró un mayor riesgo de defunción en municipios localizados al norte del país en comparación con municipios situados en la mitad sur.

En el análisis de regresión de Poisson de la variación de mortalidad por ENM y emisión de metales pesados a cuencas hidrográficas, se encontró que la mortalidad por este grupo de enfermedades fue un 18,4% mayor en los municipios expuestos que en los no expuestos, tanto de forma global así como en los análisis individuales por tipo de metal.

Conclusiones

Las ER suscitan cada vez mayor interés debido al alto impacto que tienen sobre los pacientes afectados, familiares y sociedad en general. Por ello, es necesario mejorar la información disponible mediante la investigación a través de perspectivas multidisciplinares. Esta tesis doctoral ha enriquecido el conocimiento de la variabilidad temporal y espacial de la mortalidad atribuida a ER añadiendo un enfoque geográfico. Se han analizado algunas de las dificultades que presenta el estudio epidemiológico de estas enfermedades poco prevalentes debido al bajo número de casos y la necesidad de agregación espacial que provoca como consecuencia el PUEM. También se ha aportado información relevante sobre la variabilidad temporal y geográfica de la mortalidad para las ER seleccionadas, además de explorar posibles asociaciones con factores geográficos. Los

resultados de este estudio ofrecen una información muy valiosa para la planificación sanitaria, ya que apoyan y justifican acciones de prevención y de asistencia en aquellas zonas donde se han detectado más desigualdades en el riesgo de defunción debido a estas enfermedades. Además, al relacionarlo con factores ambientales sospechosos se ofrecen pistas para futuros estudios que puedan ahondar en el descubrimiento de las causas que provocan las ER.

Chapter 1

Introduction

This introduction summarizes the state of the art of the geographic analysis of mortality attributed to Rare Diseases (RD). The evolution of the relationship between geography and health and how they have increasingly converged from the time of the ancient Greeks to the present day will be commented here. At present, Health Geography is a growing research discipline on which Geographic Information Technologies (GIT) plays a key role. Simultaneously, epidemiology has also benefited from this integration and from the introduction of new procedures and techniques to analyze issues concerning geographic space. A brief allusion to the concept of epidemiology is essential in understanding the term of RD, which refers to a joint of diseases counted by the thousands. For the purpose of this thesis, three of them have been selected and are briefly described in this chapter. Finally, this introduction also includes the justification of this doctoral thesis, the hypothesis established and the proposed objectives.

1.1 From Medical Geography to Health Geography

This section comprises a brief summary of the evolution of the interaction of geography and health throughout history which is mapped out in **Figure 1**. The relationship between the disciplines of medicine and geography is very ancient and was born as a way of associating the habitats (geographical component) with the pathogens (medical component). The Greek philosopher Hippocrates wrote one of the most celebrated medical works called *Air, Water and Places* in the 4th century BC [Hippocrates *et al* 1752]. This medical piece of writing introduced the concept of environmental theory based on physiology, in which the physical space was fundamental in the internal balance of temperaments, leading to diseases and death [Cruz-Coke 1999]. This thought being adopted by the Greek philosopher, *geographic place* is a key element in understanding health and disease. In so doing, the geographic characteristics such as the topography, climate, and atmosphere in which a person lives should be analyzed in order to prevent certain diseases [Jori 2013]. The environmental tradition remained during centuries as a basis for what is known as Medical Geography. This subdiscipline can be described as the study of the physical and social environment and its relationship with pathological problems being hygiene, sociological and geographical empirical research and urban space-related problems its main topics [Lyseen *et al* 2014]. However, the definitive development of Medical Geography took place during the Enlightenment thanks to its institutionalization and to the disciplines of geography and medicine merging together into a subdiscipline [Urteaga 1980]. During the Enlightenment, the studies known as medical topographies and associated to hygiene theories, flourished in the European continent. Medical topographies were studies taking place in specific geographical places from a hygienic-sanitary

perspective which comprised physical descriptions (situation, climate, soil, and hydrography), biological environment, historical background, information about inhabitants, dominant pathologies and disease distribution [Solís 2001]. Aligned with the progressive implementation of the scientific reasoning in the 19th century, new research lines and the enlargement of scales of analysis were deployed. This led to the publications of numerous works on mapping of diseases establishing a causality link between geographic and pathologic factors.

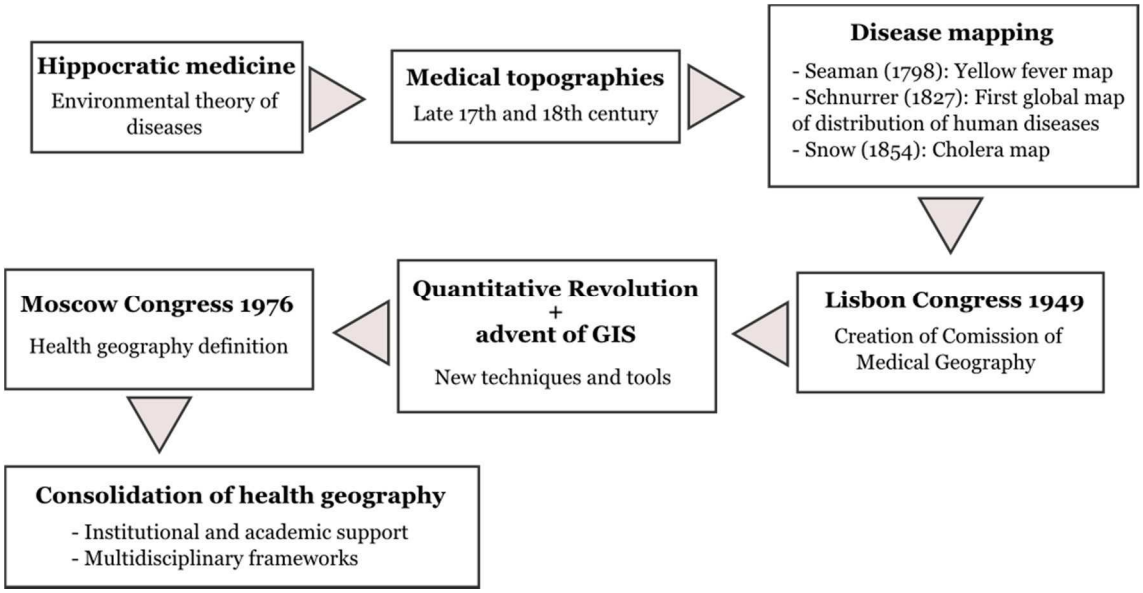


Figure 1. Main milestones in the development of Medical Geography.

The first epidemiologically relevant map is attributed to Seaman, which showcased a map of the incidence of yellow fever in New York in 1798. However, he cannot be considered the father of spatial epidemiology because he was a strong supporter of the miasmatic theory [Garfield 2013]. This obsolete medical theory attributed the epidemiological diseases to *miasma*, or in other words, bad air emanating from rotting organic matter [Urteaga 1980]. At that time, there was an important development of maps of diseases. For example, Schnurrer published the first global map of the distribution of human diseases in 1827, considered a milestone for geography in the health field [Brömer 2000]. Shortly after, Medical Geography started acquiring a scientific basis. The treatises of Snow, Lombard, Boudin or Hirsh became highly relevant before the end of the 19th century.

John Snow's work deserves a special mention as he is considered one of the founders of modern epidemiology. Snow studied several epidemic cholera outbreaks that were

devastating London in the 1850s. He associated the location of water wells with the deaths due to cholera. The value of this study is rooted in the location map showing water wells and death events. When people stopped drinking from suspect water wells, deaths fell down dramatically. What is more remarkable is that he introduced the concept of place in epidemiological studies [Snow 1855, Bonita *et al* 2006].

Decades after, the discovering of microbial agents of the diseases resulted in a loss of interest in studies relating environmental influences and human health. Pioneer geographers in Human Geography such as Ratzel or Vidal de la Blache did not show almost any interests for Medical Geography. The geographer Bhrunes demanded a greater presence of health topics in geographical studies, noticing the scant interest in the first decades of the 20th century [Deffontaines 1926]. In 1949 the International Geographical Union (IGU) created the Commission on Medical Geography in the Lisbon Congress in order to give a greater institutional coverage to this subdiscipline and to promote dialogue and collaboration among researchers. In the same decade, the World Health Organization (WHO) (1946) defined the concept of health in its constitutional document as *"a state of complete physical, mental and social well-being, not merely the absence disease or infirmity"*. This definition conveys a holistic understanding of the concept as well as a need to tackle it from a multidisciplinary perspective. In this way, the importance of geography in explaining and understanding the balance between health and disease is evident, as geography studies the relationship between humans and their environments.

Between the 1950s and the 1960s a quantitative revolution took place, considering geography as a spatial science and seeking to develop systematic methodologies. This revolution was possible when new statistical and mathematical techniques, theorems and proofs were introduced in geographical research [Burton 1963]. In relation to health, the Swedish geographer Torsten Hägerstrand tried to explain the spreading of diseases through models based on laws [Carter 2016]. From then on, the evolution of Medical Geography might be divided into three approaches according to the problems that the world faced at that moment, namely the ecological, the spatial and the social approach. Spatial studies together with time-trend studies acquired relevance in order to analyze location patterns of indicators such as mortality or morbidity. These studies were aimed at contributing to the spatial understanding of epidemics and infectious diseases, particularly in underdeveloped countries [Jori 2013].

From the 1970s, studies in developed countries focused not only on non-infectious diseases as cancer, but also on infectious diseases such as influenza or tuberculosis. At the Moscow Congress, the IGU Commission changed the name of Medical Geography to Health Geography in 1976. Verhasselt (1993) summarized the concept as follows: Health

Geography encompasses the disease ecology (study of the spatial morbidity and mortality patterns) and the care delivery (grouping the study of distribution and planning of health-care services). In line with the definition above, Health Geography not only focuses on quantitative methods but also pays attention to diseases related to the anthropogenic influence over the environment, inequalities at a global level or risk habits. Under this more integrative definition, Health Geography acquires a social perspective as well. At first glance, it may seem that the objectives of Health Geography overlap with those pursued by Public Health, however there are some differences concerning the concept of *place*. Anthamatten and Hazen (2001) pointed out that *place* responds not only to the location where an event occurs but also the reason that propels it. Thus, a quantitative and qualitative integrated approach is necessary when conducting research. Health Geography, as a subdiscipline of Human Geography, acquired remarkable prestige in recent decades as shown by the numerous research lines and existing publications [Anthamatten and Hazen 2011, Gatrell and Elliot 2014, Emch *et al* 2017, Brown *et al* 2018].

In the 21st century, advances on GITs facilitated the takeoff of geovisualization linked to the internet, benefiting classic cartography and mapmaking processes [Cauvin *et al* 2010]. The use of RDs mortality indicators in disease mapping is one of the basis of this doctoral thesis, together with the panoply of techniques provided by the GITs.

Nowadays, Health Geography is a consolidated field of study and its results are taken into account by politicians to implement health policies. Carter (2016) highlighted two main reasons to explain the rise of Health Geography in the last decades. On the one hand, health systems and academic institutions promote interdisciplinary frameworks, especially those related to Public Health (preventive medicine, health promotion or surveillance among others), which are therefore associated with geography. On the other hand, there has been a sharp increase of economic and human resources invested on Public Health, not only in developed countries but also in underdeveloped ones. Concerning the later, it is worth noting the support provided by private foundations and international organizations, to which incentive interdisciplinary frameworks, and geography can offer tools to analyze health problems.

The historical development of the relationship between geography and health in Spain has been similar to that seen in the rest of the European countries, although it occurred at a slower pace. Problems about hygiene rose considerable interest. Medical topographies had a great takeoff after the 18th century as it happened in other countries. This type of studies, thoroughly examined by Solís (2001) and Urteaga (1980), emphasized geographical elements rather than disease outbreaks so as to promote health preventive measures and remedies to improve the health of the individuals. More than 250 documented medical

topographies focusing on Spanish cities and regions were published between the 19th century and the beginning of the 20th century. Undoubtedly, scientific institutions, medical academies and the development of hygiene theories played a key role in the emerging medical topography reports. Medical Geography paid attention to problems of public hygiene as well as to the necessity to link the environment and social contexts to pathological processes as a field of research. Within the Spanish context, Iglesias Díaz (1886) provided another comprehensive definition of Medical Geography as the *"study of the telluric, atmospheric or other modifiers that influence human life, births and deaths, health, diseases and reason, nature and treatments"*.

The hygienist movement dominated Spain until the 1950s, when the link between geography and health went into decline. It was not until 1992 that Olivera (1992) systematized the contents of Health Geography in a comprehensive work. From the 1990s onwards, and thanks to the advent of Geographical Information Systems (GIS) and their application in health studies, the situation has changed. A set of publications and reports has emerged, focusing mainly on environmental health, health services and disease mapping [López-Abente *et al* 2001, Benach *et al* 2001, Aránguez-Ruiz *et al* 2005, Aránguez-Ruiz *et al* 2012]. For the purpose of this doctoral thesis, disease mapping work will be described as the estimation and summary of different health indicators such as description, hypothesis generation, location of natural resources, assessment in health inequalities and observation of risk variations, to get a global focus in the epidemiological analysis [Olivera 1992]. Disease mapping is currently present in numerous epidemiological projects thanks to the consolidation of GIS and the availability of mathematical models in specialized software programs [Ocaña-Riola 2010].

This summary has brought forth the evolution of geography at a European and Spanish level and the progressive integration of geography in the health field. From the second half of the 20th century, GITs and multidisciplinary work have proved to be fundamental in explaining the current situation. Geographers and health practitioners have improved their knowledge on the spatial distribution diseases, which has an impact in prevention tasks and in the generation of plans to benefits the health of populations.

1.2 Adoption of Geographic Information Technologies in health studies

Until the 1950s, health studies focused on the use of classic maps of health reports in paper format [Boelaert *et al* 1998]. The geographer Wright, member of the American Geographical Society, was able to predict the future necessities of cartography in terms of the advent of computers. In his research, it was highlighted that computers would create new opportunities associated to data management, new technical analysis and improvement in visualization. The development of operational GISs or GITs had a paramount role in the evolution of the relationship between geography and other scientific fields of research [Bosque-Sendra 1997]. In the 1960s, Roger Tomlinson developed the CGIS (Canadian Geographical Information Systems), the first GIS. It was aimed at using the computers to obtain information related to maps and to inputting geographic information. CGIS focused on the handling and the data analysis of the Canadian forest inventory in a digital format to manage rural areas. In the same way, it was meant to be a tool to integrate information from other data sources as the Census [Tomlinson 1973]. Meanwhile, new tools and techniques for mapping and map analysis of problems related to resources assessment, land-use evaluation and planning were developed in other countries [Olaya 2016]. Some examples of institutions implementing these tools are the Harvard Laboratory (United States) and the Experimental Cartography Unit (United Kingdom), which also developed GIS software during the 1960s [Liebenberg and Demhardt 2012]. Successively, the inclusion as a work tool in local and national agencies, the commercial development, and finally the growing and massive interest by institutions and researchers at a global level made the consolidation of GITs in a wide sort of disciplines and applications possible [Santos-Preciado 2004].

The integration of GITs in Public Health studies has occurred following the same line, resulting on an improvement in the quality and quantity of epidemiological studies in lines of research as health planning, access to health care, logistics, spatial differences or health in general [Shaw and McGuire 2017]. This is because GIS has been proven as an essential tool in the health sector to improve the health of populations due its capabilities to carry out operational needs rapidly and to help in decision making [Escobar 2001].

The advantage of adopting GITs in health had enabled a better integration of health data and cartography, the study of spatial and temporal patterns, the study of spatial risk of diseases and the association with environmental factors at different scales [Fletcher-Lartey and Caprarelli 2016]. Shaw and McGuire (2017) defined the concept of Health GISs as "*the integrated systems containing tools for managing, inquiring, analyzing and presenting spatially-referenced health data*". In addition, the field of research of GITs in epidemiology

and Public Health is wide. Dummer (2008) divided it in seven areas as follows: (1) services, infrastructure and land-use planning, (2) surveillance, modelling and mapping, (3) etiology of diseases, (4) assessment of environmental risk factors in health, (5) use of health services, (6) health inequalities, and (7) therapeutic and healthy landscapes. The increment of GITs in health research over the past decades and up to this day is clear, as the great adoption by the community of researchers and the growing number of available academic publications has demonstrated [Lyseen *et al* 2014, Shaw and McGuire 2017, Nykifourk and Flaman 2011]. *Geographic place* has become a basis in Health GISs studies, being proven as a bridge to obtain evidences and as a point of support for policy and decision makers as well as practitioners in the development of actions in Public Health. The success of GITs in health can be seen in numerous journals, papers and in the proliferation of kind of atlases displaying geographic information related to health [WHO 2008, Eurostat 2009, Timmis *et al* 2017].

The benefits of GITs in health science have also had tangible achievements for the scientific community in Spain. At the level of journal publication, there has been a growing interest as demonstrated by some examples such as *Serie Geográfica* (edited by the University of Alcalá). In its special issue dedicated to the application of GITs in health, topics ranged from health atlases, implementation of GIS in environmental health studies, the use of remote sensing to mapping diseases or the use of GIS to monitor disease outbreaks [Glover *et al* 2005, Laffly and Handschummacher 2005, Ramírez 2005, Escobar *et al* 2005]. In addition, it is also remarkable the work done about mortality studies and about cancer disease both in the creation of atlases and in many publications accounting with geographical analysis of associations of environmental factors and diseases [Benach *et al* 2007, López-Abente *et al* 2007, Benach *et al* 2013].

The merging together of GITs and health research is reflected by the existing broad catalogue of disease mapping publications, which also contains interactive websites enabling to display geographic information quickly and easily. Some examples of interactive web with maps displaying geographic data are listed below:

- International Cancer Atlas website, developed by the American Cancer Society, WHO and the Union for International Cancer Control: http://canceratlas.cancer.org/data/#?view=map
- National estimates of cancer incidence and mortality in 2018 provided by the European Cancer Information System, European Commission: https://ecis.jrc.ec.europa.eu/#
- Surveillance atlas of infectious diseases provided by the European Centre for Disease Prevention and Control: https://atlas.ecdc.europa.eu/public/index.aspx

- Interactive Epidemiological Information System provided by the National Center for Epidemiology, Instituto de Salud Carlos III (ISCIII), Spain: http://ariadna.cne.isciii.es/

1.3 Epidemiology and Rare Diseases

The raising awareness about RDs by the scientific community, society and policy makers has generated the need for describing these low prevalence diseases from a medical and social point of view. Likewise, the particular features and the huge number of these rare conditions have resulted in the creation of diverse organizations and institutions engaged exclusively in health research, studying their epidemiology or the hunt of treatments of RDs. The following sections will describe these diseases, they will briefly summarize the main institutions and research lines that deal with them and they will present the RDs selected for this doctoral thesis.

1.3.1 Definition of Rare Diseases

The European Union (EU) defines RDs as those affecting not more than 5 people per 10,000 inhabitants within the UE. This definition also includes a strategic and common action plan that should be carried out for diseases under this threshold [Decision No 1295/1999/EC, de Vrueh *et al* 2013]. It may be added that RDs not only refer to the rarity, but also to the debilitating and the life-threatening nature of these diseases. The choice of this number of cases as maximum refers to pharmaceutical investment. When below 5 patients out of 10,000 people, the pharmaceutical industry may not find profitable to develop new drugs. This threshold has not a worldwide scope, since Japan and Australia set the limit in 4 cases per 10,000 inhabitants while United States considers a disease as rare when the total number of affected patients is lower than 200,000. If the EU would have chosen an absolute number to define the RDs, an update should be carried out due to population variations with the addition or dropouts of countries [Posada *et al* 2002].

Low prevalence diseases is the technical term to describe RDs, however it was not attractive for the general public, hence the adjective *rare* was adopted. The Federación Española de Enfermedades Raras (FEDER) assumed the term RD as well as the patients, mass media and public at large during the 1990s. Today, RD is a transversal definition which includes a wide number of diseases affecting people and their families, with a low risk of stigmatization associated. However, the term of *infrequent diseases* has also been taken up in recent times [Posada *et al* 2016].

The precise number of RDs is unknown and it is based on estimations. More than 8,000 RDs have been declared and, between 25 and 30 million people (6-8%) of the European population might be affected by a RD throughout their lives. There are estimates that already extend the range to between 27 and 36 million [Groft and Posada 2017]. The pathologies are diverse, as well as their clinical characteristics, etiology, physiopathology, disease evolution, effects on the patient, treatments and interferences with other diseases [Le Cam 2014]. Despite the diversity of RDs, some common features can be outlined:

(1) RDs are serious or critical diseases, chronic, degenerative with life-threatening in many occasions.
(2) Most RDs have an early onset in the first years of life.
(3) They are debilitating diseases making it difficult for patients to lead a normal life.
(4) In general, RDs are incurable and the procedures only involve treatment of symptoms and palliative care.
(5) Should there be treatments for RDs, patients and families find it difficult to access them.
(6) There is a heterogeneity and geographic dispersion of patients which makes it harder to complete robust studies in small populations.

Hence, the particular characteristics of RDs make the integration with other disciplines more complicated. In the case of the adoption of TIGs in RDs studies, this integration has been much weaker than in more common diseases or more prevalent pathologies. However, in the last decades several steps have been taken, achieving growing interest. The National Atlas of Rare Diseases 1999-2003 is particularly remarkable in Spain, as it was developed at the provincial level. Other epidemiological reports including disease mapping focused in Autonomous Communities (AC) as Comunidad Valenciana (by municipalities) or Region of Murcia (by health areas) also deserve attention [Arizo-Luque *et al* 2005, Botella-Rocamora *et al* 2006, Titos-Gil *et al* 2011, Mira-Escolano *et al* 2018].

In any case, RDs have become a differentiated group and they suppose a challenge for other disciplines. Due to this, a growing interest by the scientific community and Public Health politicians has arisen.

1.3.2 Rare Diseases research

RDs pose particular challenges to patients who are affected, to the clinicians who diagnose and take care for them, and to the researchers who are studying their conditions. To this day, there is a significant level of funding in RDs research in developed countries as it is demonstrated in numerous consortiums at European and at international level, multidisciplinary projects and platforms that ensure the collaboration among patients, institutions and researchers.

For many decades, research on RDs has suffered from the atomization of scientific groups, companies, funders and the lack of treatments, in part arisen from the complexity of addressing this group of diseases. All this led to the creation of the International Rare Diseases Research Consortium (IRDiRC), in order to promote global partnerships and to advance in RDs knowledge by involving national and international governmental institutions, non-profit funding bodies companies, patient advocacy organizations and scientific researchers at worldwide level [Lochmüller 2017]. The 2020 IRDiRC objective of achieving 200 new therapies and means to diagnose RDs has been already reached. New RD research goals for the next decade have been defined (2017-2027), aimed at improving the medical attention and therapies for patients and their families.

Another remarkable initiative is the European Joint Programme on Rare Diseases (EJP-RD, http://www.ejprarediseases.org/). EJP-RD is an inclusive initiative launched in 2019, participating more than 130 institutions of 35 countries. Its goal is to promote a connecting research, health care and innovation in the RDs field and to develop efficient models of financial support. The EJP-RD actions can be summarized in four main pillars: research funding, coordinated access to data and services, strengthening of capacities and accelerate translation of research projects and clinical trials.

Regarding national actions, Spain was one of the first countries launching initiatives of RDs investigation at a national level. Among other initiatives, the ISCIII created the Institute for Rare Diseases Research (IIER, *Instituto de Investigación de Enfermedades Raras*) as the center to promote research, training, innovation and health planning in RDs in 2003. Additionally, the CIBER (*Centro de Investigación Biomédica en Red*, https://www.ciberisciii.es/) with the specific area of CIBERER, devoted to RDs, undertakes transversal research programs focused on those rare conditions. Finally, both the IIER and the ISCIII created the Spanish Patient Rare Diseases Registry (*Registro de Pacientes de Enfermedades Raras,* https://registroraras.isciii.es/Comun/Inicio.aspx) and the BioNER (*Biobanco Nacional de Enfermedades Raras*, http://bioner.isciii.es). BioNER provides

access to harmonized data and biological samples, representing a key support for Spanish RDs research.

1.3.3 Rare Diseases epidemiology

Epidemiology began to be used at the end of the 19th century as a powerful tool to understand the distribution of infectious diseases. Epidemiology is defined as the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the prevention and control of health problems [Last 2001]. At the same time, epidemiology is currently an essential part of Public Health, whose goal is to develop collective actions leading to improve the health of populations. An epidemiological study tries to answer to the characteristics and determinants of a disease affecting the population. Descriptive epidemiology is oriented to respond to the *what* (health issue of concern), *who* (person), *when* (time) and *where* (place) about a disease. Those questions being answered, analytic epidemiology tries to ascertain the disease etiology (cause) [Bonita *et al* 2006]. Epidemiological studies have as their main task Public Health surveillance, research, analytical studies, evaluation, linkages and support policies development [Dicker *et al* 2006]. National studies of epidemiological surveillance have a strong presence with the use of morbidity and mortality data, especially in the case of infectious diseases. This surveillance is aimed at detecting risk situations in the health of populations, monitoring epidemiological characteristics, and carrying out preventive and controlling interventions.

Due to their particular characteristics, the study of RDs is an additional challenge in the field of Public Health, as it seeks to improve the health of populations. The classification and codification of diseases suppose an important challenge for researchers when facing RDs, and therefore this is a fundamental first step to carry out epidemiological studies. Classifications of diseases are geared towards common diseases, and if a disease is not included in some classifications (as it occurs with many RDs), it will not have visibility in health information systems in terms of counting, temporal trends, and geographic distribution or to compute cost burden [Posada *et al* 2010].

In general terms, RDs can be classified following different criteria. For example, they could be divided into congenital and non-congenital according to the onset of the disease. Also, they can be classified depending on the cause: genetic, environmental, multifactorial or unknown cause. As RDs pathologies are various, the criteria are compatible in many cases, and their implementation depends on the foreseen objectives [Posada *et al* 2016].

The codification of diseases is widely regulated by the International Classification of Diseases (ICD), which is carried out by the WHO, being the bedrock for clinical coding and mortality data. It is structured in chapters according to the type of disease or system affected [WHO 2011]. ICD 9th revision (ICD-9) was used to codify the cause of death from 1981 to 1998 in Spain. ICD 10th revision (ICD-10) is used to codify the cause of death since 1999 to this day. Not all countries have adopted the same classification over the same period of time, finding it hard to establish a comparison among countries. ICD-10 only includes 300 RDs, each of them having their own code, whereas varying situations are given to the rest. Some difficulties that may affect the coding of RDs are (1) the attribution of the same code to more than one RD, (2) the lack of code of a RD, (3) the use of different codes for the same RD according to its interpretation, or (4) the possibility that the disease affects multiple body systems (problems to allocate it in only one chapter) [Posada *et al* 2016]. There has been an improvement in the codification of RDs in ICD-10 in respect to the last revision whose list only included 100 RDs with their own code. The new classification of diseases (ICD-11) is expected to gather codes of more than 5,000 RDs [Aymé *et al* 2015]. Apart from the ICD, there are other specific classification systems for RDs such as Orphanet, which is based on multiple-entry codes called Orpha [Orphanet Report Series 2019].

The correct identification and codification of diseases is essential to carry out epidemiological studies about RDs morbidity and mortality. To reduce RDs mortality—which is high in both pediatric and elderly groups due to their chronicity—, it is key to identify and monitor health indicators both in time and space. However, the problems in codification for the majority of RDs make work with mortality or other statistic indicators harder since most of the studies are oriented to more common diseases. Nonetheless, information provided by mortality registries shares a number of characteristics that prove its usefulness for RDs research: (1) they are uniform and codified following the ICD classification (with continuity of data despite the revisions), (2) they are universal and population-based data (this allows comparison in space and time), and (3) they enable the analysis of diseases using the underlying cause of death [Arias-Merino *et al* 2017].

1.3.4 Description of Rare Diseases analyzed in this doctoral thesis

Currently, there are between 6,000 and 8,000 RDs identified, characterized by being a heterogeneous group of pathologies with different clinical characteristics and etiology [Posada *et al* 2016]. This means that global studies are challenging and sometimes it is preferable to focus on individual or specific RD groups. RDs selected in our study had to fulfill a number of pre-established requirements:

- Being codified in classification systems and maintaining a consistency with a single code equivalence between the ninth and tenth edition of ICD, thus allowing a long-term period analysis.

- Accounting for enough number of deaths throughout the Spanish territory in order to obtain powerful statistical results after epidemiological calculation.

- They had to be RDs in which geographic analysis at a great level of detail would be of great interest for researchers, health professionals and affected patients.

- Being suspected to have an environmental factor influencing the disease etiology.

A brief clinical and epidemiological description of the three diseases we have selected for our study is presented below.

Huntington's disease

Huntington's disease (HD) is a rare neurological disorder with a genetic basis and autosomal hereditary dominant whose etiology is still unknown [Pupo *et al* 2013, Bates *et al* 2015]. The most frequent symptoms affect the motor system, cognitive deficit characterized, psychiatric and behavioral disturbances, appearing any of these symptoms in indifferent order [Walker 2007]. The manifestations of HD generally occur in the fourth decade of the patient's life [Orth and Schwenke 2011].

The progression of HD leads to growing dependence in patient's daily lives with death occurring, even in some cases due to pathologies associated to this RD (such as pneumonia, heart disease or suicide) [Farrer 1986, Dorey *et al* 2012]. To date, it was not found an effective treatment and drug therapy is limited to symptoms treatment, such as chorea, depression or anxiety. The majority of HD patients affected are close relatives of people affected, so there are few new isolated cases. In this way, the propagation to successive generations can be prevented by the correct genetic counselling [Ramos-Arroyo *et al* 2005].

In recent years, the identification of affected patients has improved as well as the patient registries. However, population-based studies are still scarce. According to the prevalence and incidence, the HD distribution shows higher rates in North-America, Oceania and Western Europe, while lower rates in Asia [Pringsheim *et al* 2012]. Mortality population-based studies have been carried out in a few countries at a national level [Ekestern and Lebhart 2005, Sipilä *et al* 2015]. In Spain, the analyses of HD mortality are limited to temporal studies but geographical studies had not yet been carried out [Ramalle-Gomara *et al* 2007].

Granulomatosis with polyangiitis

Formerly known as Wegener's Granulomatosis, Granulomatosis with polyangiitis (GPA) is a systemic disease characterized by an inflammation of the blood vessels leading to tissue death. The apparition of necrotizing granulomas is the most frequent clinic characteristic [Martínez-Morillo *et al* 2012]. GPA can be diagnosed using a test in which positive levels of anti-neutrophill cytoplasmic antibodies (ANCA) are yielded. Left untreated, this disease has a life-threatening development in three or four years, mainly due to renal affectation. The treatments with cyclophosphamide, corticoids and rituximab have a high cure rate in affected patients (approximately the 95%) [Phillip *et al* 2008, Tavakolpour and Alesaeidi 2019].

The etiology of this disease is still unclear, although there are suspicions about the combination of a genetic predisposition and environmental triggers. There are a wide variety of studies relating GPA rates with geographic or environmental factors, such as the seasonality in the onset of GPA, urban versus rural differences as well as migration-related studies [Mahr *et al* 2006].

Epidemiological data about GPA show relevant geographic differences in the incidence and prevalence rates. Such rates are usually higher in Northern European countries than in Southern ones [Mahr *et al* 2006]. In Spain, until the publication of one article included in this doctoral thesis, mortality studies at a national level have not been yet carried out, either with temporal assessment or geographic analysis.

Motor neuron diseases

Motor neuron diseases (MND) are a group of neurodegenerative disorders, unknown etiology, characterized by a progressive loss of motor neurons in the motor cortex, and damage in the brainstem as well as in the spinal cord [Kiernan 2018]. MND, as other

related diseases as Alzheimer or Parkinson, are determined by the degeneration of nerve cells. The amyotrophic lateral sclerosis (ALS) is the most known and frequent pathology of MND, which affects more than 85% of patients [Ingre *et al* 2015]. It is a debilitating disease, with a progressive course and drastic outcome in short-term. There is no cure for MND and the treatments are based solely on palliative measures, patient support and rehabilitation aimed at improving the patient's quality live [Miller and Appel 2017, Arias-Merino 2017].

Risk factors are not clear, but these diseases have been associated to ageing, masculine gender, family history, or the mutation of a certain gene related with this RDs group (C9ORF72 and SOD1 genes) in several specific geographic areas [Ingre *et al* 2015, Arias-Merino *et al* 2017].

Epidemiological indicators show a higher prevalence and incidence in North European countries than in Southern ones [Chiò *et al* 2013]. However, there is still a lack of studies at a national level since they used to be focused on regions, cities or provinces [Alonso *et al* 2011,]. MND mortality and incidence indicators are similar, which is advantageous in order to study it. The reason of their being similar is these diseases' quick progression from the moment of diagnosis to the death of the patient.

1.4 Justification

One of the goals of Public Health has always been the understanding the inequalities in mortality and its causes, both in spatial and temporal terms. For this purpose, national population-based death registries have been proved as a very useful inventory for mortality studies thanks to their standardized methodology and to their covering long periods of time. At the same time, RDs are a complex and understudied phenomenon in Spain despite the fact that they account for more than 6,000 different diseases and it is estimated that they may affect more than 3 million people in Spain. Overall, RDs usually present high chronicity and mortality in affected patients, which generates high costs of treatment and care for the public health system, as well as for patients and families. Hence, there is a pressing need to improve knowledge on RDs at every possible level.

At the same time, it is known that geography studies the relationship between individuals and the environment in which they live. As a result, its essential role in understanding the possible environmental factors that may be influencing mortality due to these rare conditions is clear. However, from an epidemiological point of view, RDs have not yet been analyzed through an extensive use of GIS. Thanks to their diverse possibilities to integrate, analyze and represent geospatial data, GIS can allow a better understanding of the

variability of mortality attributed to RDs. Although in many cases the etiology of RDs is still unknown, revealing spatial patterns may help to elucidate environmental factors that explain the distribution of mortality in the geographic space and ascertain the causes of disease. This thesis dissertation focuses on the analysis of three low prevalence diseases (HD, GPA and MND) which were selected according to the criteria described in the previous section.

Finally, this doctoral thesis justifies the identification of spatio-temporal inequalities in mortality attributed to RDs throughout the Spanish territory and its possible causes. In so doing, it will help organizations, professional and health politicians carry out actions aimed at reducing mortality.

1.5 Hypothesis

As socio-economic conditions are similar throughout Spain and the National Health System works equally in the whole territory, the spatial differences found in RDs mortality rates might be related to either individual factors or environmental conditions. Moreover, some factors such as the increment of life expectancy, better treatments and health care, and improved diagnosis suggest that RDs mortality rates should have experienced a decline during the past decades.

1.6 Main aim and specific objectives

The overarching aim of this doctoral thesis is to enhance knowledge on mortality attributed to RDs using Geographic Information Systems and epidemiological analysis tools with the purpose of checking the given hypothesis and making a contribution to the RDs field. This main aim can be disaggregated in the following four specific objectives:

Objective 1: To identify the current problematic in the choice of geographic units of representation when working with RDs and to provide recommendations for the election of the most appropriate unit so as to obtain a better epidemiological and cartographic result. To achieve this purpose, Huntington's disease will be selected as an example of RD. This objective will be developed in Chapter 3.

Objective 2: To evaluate the time trends on mortality attributed to RDs in Spain over three decades. This objective will be achieved by selecting Granulomatosis with polyangiitis and Huntington's diseases, and it will be developed in Chapters 4 and 5 respectively.

Objective 3: To identify geographic patterns in mortality attributed to RDs in order to observe variations in mortality among spatial units through the Spanish territory. This objective will be tackled with the following three RDs: Granulomatosis with polyangiitis, Huntington's disease and Motor neuron diseases. The attainment of this objective can be found in Chapters 4, 5 and 6.

Objective 4: To explore the association between RDs mortality and pollutant emissions using the existing scientific literature as a starting point. To achieve this objective, location, activity, type and quantity of heavy metals emitted to river basins exceeding the annual thresholds set by the E-PRTR will be described. The attainment of this objective will deepen our knowledge on the associations between mortality attributed to motor neuron diseases and heavy metals. This last objective will be developed in Chapter 6.

Chapter 2

Methodological considerations

This doctoral thesis is heavily supported by the use of diverse data sources and by the application of tools and techniques both from the area of GITs and epidemiology. This is geared to obtain results about the health status of populations concerning to mortality attributed to RDs in Spain. For this reason, on the one hand it has been necessary to collect data from a wide variety of national agencies and, on the other hand, to use specific software provided by several enterprises. All of this will be conducted in this Chapter.

2.1 Data sources

As a part of the methodological process, this thesis has invested a great deal of effort in collecting data from various sources such as organizations and national administrations. Despite being secondary, the data of our studies are highly reliable as they belong to official information sources. Furthermore, they are very stable over time and publicly accessible. The data used in this doctoral thesis can be classified depending on spatial or thematic features (which constitutes the geographic information). The management and integration of all of them are possible thank to the GITs and statistical tools. A summary of the main sources and data collected are shown in **Table 1,** and more detailed explanations will be provided in the following section.

Table 1. Main data sources and databases used to carry out this doctoral thesis.

Logo	Name of data source	Data provided
INSTITUTO GEOGRÁFICO NACIONAL cnig	National Mapping Agency	Administrative division units 1:25000
		Spanish hydrography scale 1:500000 (BCN500)
INE Instituto Nacional de Estadística	National Statistics Institute	Annual Death Registry
		Municipal Register of Inhabitants
		Population Census
E-PRTR	European Pollutant Release and Transfer Register	Pollutant emissions of heavy metals into rivers

2.1.1 National Mapping Agency

Geographic data are composed of spatial information (X, Y) as well as thematic information (Z). Using GIS software, thematic information can be integrated in a database to enable subsequent operations such as analysis or visualization. The National Mapping Agency (NMA, *Instituto Geográfico Nacional*) is a Spanish government agency attached to the Ministry of Public Works. It is responsible for the maintenance of the spatial data infrastructure, cartographic production, as well as other works related to astronomic observations, geodetic networks, remote sensing systems, geophysics and seismographic and volcanic activity analyses [Instituto Geográfico Nacional 2019]. In 1989, the NMA created the National Geographic Information Center (NGIC, *Centro Nacional de Información Geográfica*) to provide free of charge access to data according to specific quality standards. For our study purposes, two types of data were downloaded from the NGIC:

a) **Geographic units:** A shapefile is a simple, nontopological format for storing the geometric location and attribute information of geographic features. Several shapefiles in vector format of type polygon have been downloaded. Each shapefile contains basic information such as the name of each polygon contained into it (which corresponds to a municipality, province or other geographic unit), and associated numeric identifiers which can be used to integrate thematic variables from other databases using a GIS tool (as for example the health indicators). For our thesis purposes, we downloaded the shapefile at a scale 1:25000 of Spanish municipalities, containing 8,125 polygons in 2016. District (326 polygons) and *province* shapefiles (52 polygons) were generated by aggregation of municipal polygons. To achieve this, we consulted the information provided by both the NMA and the Spanish Ministry of Agriculture, Fisheries and Food (MAFF, *Ministerio de Agricultura Pesca y Alimentación*).

   For the map production, a world's countries shapefile (at a scale 1:1000000) was also downloaded from the Eurostat website. The objective was to include the neighboring and close countries to the Spanish territory in the cartographic results. The Geodetic Reference System chosen was ETRS89, which is based upon the UTM projection (zone 30 North).

b) Spanish hydrography: This shapefile corresponds to the National Cartographic Database (*Base de Datos Cartográfica Nacional*), scale 1:500000. It is a file in vector format of type polyline which contains information about the name of the rivers, their spatial location, the hierarchy level and the river basins they belong to.

2.1.2 National Statistics Institute

In 1946, The National Statistics Institute (NSI, *Instituto Nacional de Estadística*) was created a as government agency in charge of the official statistics in Spain. This institute collects and manages data on demographic and economic census, national accounts, social statistics as well as on the electoral registry. Specifically, the Demographics and Population subsection contains key information databases that were collected for our doctoral thesis such as the Annual Death Registry, the Municipal Register of Inhabitants and the information about municipal economic activities. All available data can be found and downloaded at the website www.ine.es.

a) Annual Death Registry: Mortality is one of the most important indicators to measure the heath-status of populations. In fact, monitoring mortality for epidemiological analyses, surveillance and research is a pivotal task. The interpretation of mortality data can be used for setting up preventive plans and Public Health actions for the benefit of the populations. In this sense, national mortality population-based registries provide a uniform database highly useful and reliable for temporal and spatial analyses [Alonso *et al* 2018]. In Spain, the Annual Death Registry (ADR, *Estadística de Defunciones según la Causa de Muerte*) collects all the deceases occurred throughout the national territory according to the cause of death. The majority of data records from ADR are available on the NSI website, although only data at a national, AC or provincial level are provided through open access. Certain data concerning to municipal level can be transferred to third parties (as research centers) by setting collaborative agreements. These data make part of the Vital Statistics (*Estadísticas del Movimiento Natural de la Población*), one of the works being undertaken by the NSI. The ADR database is based on the information provided by two sources: The Statistical Death Bulletin and the Statistical Birth Bulletin. Likewise, the NSI has established agreements with the ACs, as they conduct the process of codification, recording and revision of death registry occurred in each AC since 1983. Later, a national database is published by the NSI.

The underlying cause of death is one of the variables provided by the ADR. It was defined by the WHO as "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury" [WHO 1967]. The death certificate must match with the physician report which corresponds to the initial or fundamental disease code. Sometimes, the underlying cause of death is not reported as the initial or fundamental one, but another one more rigorous or detailed may be added later. In some cases, even the underlying cause may be the combination of both described previously. The underlying cause of death is a result from this process and it will be used as the official one for INS statistics using ICD classifications. The classification of causes of death dates back to the end of 19th century in Spain. However, the modern deaths statistics, according to the cause of death, were introduced as of 1951. In this year, in addition to using the ICD, international regulations were implemented to select the cause of death, keeping these regulations until our days [Instituto Nacional de Estadística 2019].

According to the different versions of ICD through history in Spain since 1981, the underlying cause of death has been identified and collected for the ADR database using the following two revisions: Ninth revision until 1998 and tenth edition from 1999 onwards. For this doctoral thesis, RDs deaths were selected using the underlying cause of death taking into account the following set periods: ICD 9th revision Clinical Modification (ICD-9-CM) was used from year 1984 to 1998, while ICD 10th revision from 1999 until 2016 as can be seen in **Table 2**.

In addition to the underlying cause of death, the ADR also provides sociodemographic information of the deceased person such as sex, date of birth, date of death and municipality of residence, among others [Instituto Nacional de Estadística 2019].

Table 2. Classification codes and study period of diseases analyzed in this doctoral thesis:

Disease	ICD-9-CM	ICD-10	Study period	Variables
Granulomatosis with Polyangiitis	446.4	M31.3	1984-2016	Case, sex, age at death, death year, and municipality of residence
Huntington's disease	333.4	G10	1984-2013	
Motor neuron diseases	Not used	G12.2	2007-2016	

b) **Municipal Register of Inhabitants:** The Municipal Register of Inhabitants (MRI, *Padrón Municipal de Habitantes*) is an updated administrative database in which citizens are registered according to their usual municipality of residence. MRI data are published every year after being checked by the NSI. It contains details on sex, age, place of birth or nationality.

   For the purposes of this doctoral thesis, the municipal population by sex-and-age distribution, corresponding to years 2001, 2006, 2011 and 2016 were downloaded. The average population of those years was taken as representative for the corresponding period for each RD, as well as for the calculation purposes of the epidemiological indicators.

c) **Population Census:** Population Census is defined as the exhaustive account of population in one country aimed at knowing social and demographic characteristics. In Spain, Population Census (together with Housing and Building censuses) is condcuted by the NSI every ten years, with the latest data available corresponding to 2011 [Subdirección General de Estadísticas de la Población 2011].

   For the purposes of the research carried out in Chapter 6, municipal data about the branch of economic activity in the active population of each municipality were collected from 2011. This basic classification allows to know the number of People Working in the Primary Sector (PWPS), which includes agriculture, forestry and fishing in each municipality.

2.1.3 European Pollutant Release and Transfer Register

The European Pollutant Release and Transfer Register (E-PRTR) (https://prtr.eea.europa.eu) is a registry providing access to environmental data from industrial facilities, as well as other installations within the EU. This registry is hosted and run by the European Environment Agency, but every country is responsible for compiling its available national information. Each country, including Spain, has its own individual website (https://prtr-es.es).

E-PRTR records have the mission of "enhancing public access to environmental information, prevent and reduce environmental pollution, and the public participation in decision-making about environmental issues". Data inventory is annually reported, together with information on the quantity and location of pollutants releases emitted to

water, soil and air that had exceed certain thresholds [Regulation (EC) No 166/2006]. The pollutants are diverse: heavy metals, pesticides or inorganic substances. This information is contained in a list of 91 elements and more than 30,000 industrial complexes as well as other facilities that emitted pollutants.

For the study purposes performed in Chapter 6, data concerning pollutant emissions of heavy metals emitted into river basins were collected and classified: by industrial complex location and by issued quantity. The heavy metals chosen were the following: arsenic, cadmium, chromium, cooper, lead, mercury and zinc. Data were collected from 2007 through 2015 (the last year available when the research was carried out).

2.2 Methods and procedures

In order to achieve the expected results of our thesis, diverse methodological procedures and tasks were undertaken by using both GIS and statistical/epidemiological tools. A brief summary of the tasks carried out are presented below. To do this, we have divided them into spatial techniques and epidemiological indicators.

2.2.1 GIS techniques

As part of this thesis, software programs for spatial data management are essential in order to input data, as well as to process and visualize the results. In this section, all the tasks that can be carried out with a geographic software are not included, but only the main of them used for achieving our objectives.

a) **Geocoding:** Sometimes, spatial data have a format not valid to be loaded in a GIS software. The purpose of the codification is to transform the description of location data (in our case collected in UTM coordinates) into a point-type shapefile for visualizing and managing in a GIS software. This operation was carried out with the emission point database provided by the E-PRTR to allow subsequent tasks such as analysis and visualization (see Chapter 6).

b) **Features edition:** GIS software enables the edition of geographic information both in vector or raster format. This tool was used to split the rivers affected by heavy metal emissions, by indicating a concrete distance from the emission points (more information can be found in the methods section of Chapter 6).

c) Overlay analysis: This is a basic operation of spatial analysis that can be carried out with a GIS software. In our case, such analysis has been used in Chapter 6 through the overlying of two shapefiles in vector format: river sections and municipalities (see **Figure 2**). This tool was used to answer the question "which municipalities are upstream or downstream of rivers affected by pollutant emissions?"

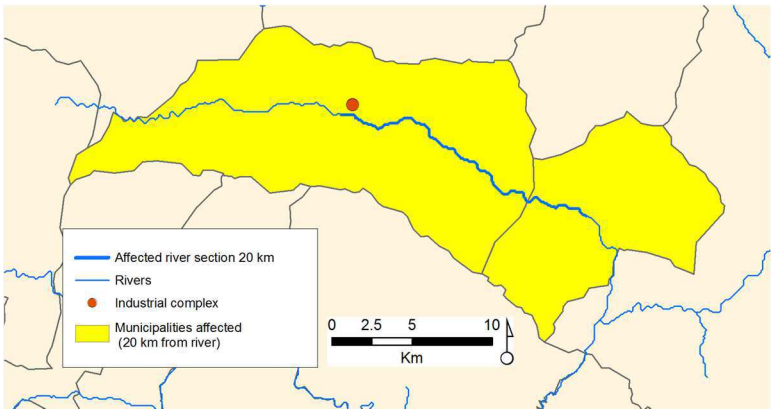


Figure 2. Example of an overlay analysis of which municipalities are within the 20 km section of Ojailén River downstream (Ciudad Real province) from an emission point.

d) Adjacency matrix: Spatial heterogeneity describes the different distribution of an event in a region [Anselin 2010]. The adjacency matrix indicates whether polygons are sharing a side or not. This information can be relevant to further statistical procedures. In order to calculate the Smoothed-SMR indicator in Chapters 3, 4 and 5, we created adjacency matrices for provinces, districts (*comarcas*) and municipalities using the *rook* criteria of neighboring areas. As can be seen in **Figure 3**, a spatial contiguity of first order is shown, which takes into account that *i* neighbors are those sharing a side with *i* [Pfeiffer *et al* 2008].

Figure 3. Example of adjacency of first order (*rook* criteria) in the municipality of Burgos.

e) Cartographic production: Maps have been created in a GIS software in order to transfer the results of statistical and geographic analysis during the execution of this doctoral thesis. Chapters 3, 4, 5 and 6 include our own maps with the purpose of helping to understand the results obtained. They are presented in the most appropriate way to the specialized audience (geographers, epidemiologists, healthcare professionals or scientific community in general).

2.2.2 Epidemiological indicators

Epidemiological indicators are necessary as they provide better knowledge on the health status of the populations. A wide range of indicators used for its quantification are available. In this doctoral thesis, we have worked with mortality indicators based on the population health databases, which enable the comparison among geographic units, as well as the evaluation of time trends with accuracy and reliability [Bonita *et al* 2008]. The understanding of these indicators has much importance since the reduction of mortality is a priority objective of national Public Health policies. In the research field of RDs, this issue is particularly relevant inasmuch as many of rare conditions have a high mortality in early life or in elderly people suffering chronic illnesses [Posada *et al* 2008]. In the following paragraphs, mortality indicators that have been calculated for our study are described. Some of the procedures also consider adjusting or standardization of rates. This is intended to reduce the age and/or sex bias and to facilitate comparison among population groups.

a) Crude death rate: This is one of the most commonly used mortality indicators. It is described as the number of deceases occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year [United Nations 1991]. In the case of RDs, the result is usually multiplied by 100,000. This indicator offers a first approach to the mortality status of the population as it is obtained quickly. However, the crude rate is not the most appropriate indicator due to the internal differences in population characteristics. An example of this would be the greater proportion of elder people in some geographical unit than in others. In this way, crude death rates hide the heterogeneity of the specific mortality per stratum [Schoenbach and Rosamond 1999].

b) Age-adjusted mortality rate: The adjusting for age and sex was introduced in the 19th century as a way of calculating the rate of a disease. It can be carried out by the direct or indirect method [Curtin and Kelin 1995, Bowling 1997].

The age-adjusted mortality rate, which is obtained by the direct method, relies upon reference a single standard population which has a known population structure. In the case of Spain, the European Standard Population is usually the reference for this calculation [Eurostat 2013]. This is an appropriate estimator for comparing groups as well as evaluating temporal trends. Such indicator has the advantage of maintain the consistence in the structure of the study population. For RDs, results are often expressed per 1,000,000 inhabitants as the number of deceases is very low and, it can be performed by years or by established periods to increase consistency. Additionally, age-adjusted mortality rates can be smoothed using a TH4253 procedure for a better observation of time trends, thereby avoiding the peaks caused by the scarce data [Velleman 1980]. *Joinpoint* regression model enables the identification of periods with significantly increased or decreased trends in age-adjusted mortality rates. Using a Monte Carlo Permutation method, such regression model evaluates the annual percentage changes during the selected period and, it also provides statistical significance [Kim *et al* 2000].

c) Standardized Mortality Ratio (SMR): This is a quantity expressed as a ratio which is obtained by the indirect method. It is necessary to know the number of persons at risk by age group and sex as well as the observed deaths in the study population. It is also required to know the age-specific death rates of the general population in the same age groups of the study population. SMR is advantageous due to its low standard error and for the comparisons among geographic areas [Inskip *et*

al 1983]. The SMR indicator is normally represented taking as a reference the number 1 or 100, which represents the rate of the global population. In our study, SMR = 1 is the rate observed for Spain. Thus, a lower value than 1 in a geographic unit means a reduced risk than expected, while a higher value means a greater risk than expected.

A Byar procedure can be carried out in order to obtain the Confidence Interval (CI) for each geographic unit [Regidor et al 1993]. Depending on the SMR value, CIs located outside of 95% are considered as significantly higher or lower in comparison with the Spanish rate.

d) Smoothed Standardized Mortality Ratio (Smoothed-SMR): Obtaining distracting noise or extreme values in rates is very common when working with spatial data with a low number of cases. In the smoothing procedures, such as Bayesian methods, statistical bias is reduced and the precision in death risk is increased [Haining 2003]. These methods take into account the uncertainty of local measurements as well as the spatial dependence between neighboring measurements [Pfeiffer *et al* 2008]. In the task of calculating the smoothed SMR in each geographic unit, three measurements are required: (1) the observed cases, (2) the expected cases and, (3) the geographical contiguities of each geographic entity. The geographical contiguities can be obtained through the adjacency matrix (see section 2.2.1). The death risk calculation uses the empirical Bayes estimation, which combines the local risk (also called likelihood) and the risk in neighboring geographic units [Besag *et al* 1991]. Smoothed-SMR results also are represented on the basis of 1 or 100 (as SMR), being higher than expected values over 1, while lower than expected values under 1.

Posterior Probability (PP) is measured in a scale from 0 to 1. The values under 0.20 show a significant lower than expected death risk, while values above 0.80 indicate a higher than expected mortality risk (i.e. when Smoothed-SMR is higher than 1, the probability of occurrence is more than 80% in comparison with the territory of reference).

2.3 Relevant geographic issues

2.3.1 Spatial database

The study area in this doctoral thesis is the Spanish territory, which comprises more than half a million square kilometers (sqkm), and about 46 million inhabitants according to the statistics provided by NSI in 2018. The three geographic units we worked with are as follows:

a) Provinces: The Spanish territory encompasses 50 provinces and 2 autonomous cities (Ceuta and Melilla). In turn, the provinces are grouped into ACs. According to the 2011 Census, province average population is 900,306 inhabitants (max. 6,421,874 and min. 81,323 inhabitants) and average surface is 9,725 sqkm (max. 21,751 and min. 13 sqkm). Provinces are the geographic units offering the lowest level of detail in our study.

b) Districts: These geographic units are known in the most part of Spain as *comarcas*. They do not have administrative jurisdiction as other Spanish units such as ACs, provinces or municipalities have. The ACs of Aragon and Catalonia are exceptions in Spain as their districts have representative political institutions and certain competences [Parlamento de Catalunya Ley 6/1987, Gobierno de Aragón Decreto Legislativo 1/2006]. There are different district classifications based on not only physical and human affinities, but also in given objectives proposed by policy makers. For instance, there are district divisions based on land use, geology, agriculture, or even "protected designation of origin" for certain products. In this doctoral thesis, we have used the 326-district classification provided by the MAFF in the work called *Caracterización de las Comarcas Agrarias de España* [Fernández 2011]. These districts have an average population of 143,168 inhabitants (max. 4,785,999 and min. 344 inhabitants) and an average surface of 1,542 sqkm (max. 5,396 and min. 13 sqkm).

c) Municipalities: They are the most basic territorial entities with a legal personality as well as financial and administrative autonomy. There are 8,125 Spanish municipalities in 2016 (last year included our study), and there are continually changes through separation and fusion processes [INE 2019]. There are large inequalities in the size and population of municipalities. The average

municipality has 5,768 inhabitants (max. 3,198,645 and min. 1 inhabitants) and a surface of 62 sqkm (max. 1,750 and min. 0.03 sqkm).

2.3.2 Classification of epidemiological indicators

As said previously, the SMR and Smoothed-SMR are epidemiological indicators that are measured on the basis of 1 or 100 as the reference value. This central value represents the ratio for the whole Spanish territory, which in our study is the general reference unit for comparing the different geographical areas. A value lower than 1 represents a mortality or death risk lower than expected and, a value greater than 1 represents a higher than expected mortality or death risk. For this reason, the number of intervals is usually uneven, with a view to having a unique medium interval around the 1 value, and two or three intervals for both lower of higher expected rates. Various examples of this kind of classifications can be found in epidemiological publications [Nagy *et al* 2013, Gómez-Barroso *et al* 2015, Villaverde-Hueso *et al* 2019].

In the case of the PP indicator, the range of possible values moves between 0 and 1. Two intervals were established for values above 0.8 (significantly higher risk), and two other intervals for values below 0.2 (significantly lower risk). An intermediate interval for non-significant values was also established (between 0.2 and 0.8).

2.3.3 Choropleth maps symbolization

Choropleth maps are a form of quantitative cartographic representation of discrete phenomena associated with areal units (or enumeration units), which are symbolized according to their assigned value [O'Sullivan and Unwin 2010]. Choropleth maps are very useful to represent projected attributes on each areal unit according to criteria of quantitative or ordinal measurement levels [Cauvin *et al* 2010]. Although they are considered as the most effective to represent discrete data, the represented data obtained in our study are continuous. These types of maps have been frequently used in spatial epidemiology [O'Sullivan and Unwin 2010], and for our study, we have taken advantage of the circumstance that mortality data are provided in aggregated form at the municipal level.

To symbolize the results, the visual variables allow to describe the differences in the variation and perception of the signs that are used to represent a cartographic phenomenon [Slocum *et al* 2005, Cauvin *et al* 2010]. In our case, two visual variables have been chosen to symbolize the epidemiological results: color (through the use of complementary colors),

and value (through the hues of the colors chosen). A combination of both visual variables in a divergent color palette is the most appropriate option and it is also widely used in epidemiological studies to represent ratios such as SMRs, Smoothed-SMRs or the PP indicators [Williams-Pickle *et al* 1996, López-Abente *et al* 2007, Benach *et al* 2013, Beale 2016]. In this thesis, a double-ended scheme of hues has been followed in which red colors are associated to greater mortality or increased death risk, while shades of green have been used to symbolize a lower mortality or lower than expected mortality risk. Therefore, the shades of color increase towards the end of intervals with this divergent scheme, remaining the yellow color in the medium interval (which represents mortality about the same as expected) [Williams-Pickle *et al* 1996].

It must be taken into account that larger choropleths cause more attraction than smaller ones when they are colored in a map [Cromley and McLafferty 2011]. This issue is particularly visible in terms of size of the choropleths at the municipal level in Spain.

2.3.4 Data aggregation: the ecological fallacy and the MAUP

The privacy and confidentiality issues make that a point distribution map is not always the most desired option for health policy makers. In these cases, the aggregation unit into polygons is the most commonly chosen option by Public Health agencies in order to protect privacy but also to incorporate demographic and socioeconomic data, enabling analyses with multiple variables [Lai *et al* 2009]. The fact of working with spatial data at polygon level results in later consequences such as the ecological fallacy. To better understand this concept, knowledge on ecological studies purposes is needed. An ecological study refers to variables analysis in group in which it is sought the association between an exposure (independent variable) and a result at a group level [Piantadosi *et al* 1988]. Their advantages are that they allow the comparison among population groups, and there is usually a major accessibility to data at group level. The ecological fallacy occurs when individual data are grouped and the analysis variable is assumed to be equally distributed throughout the whole aggregation unit (choropleth). That is, there is an incorrect inference of the statistical relationship observed when changing the scale [Eagleson *et al* 2010, O'Sullivan and Unwin 2010]. Thus, this problem is defined as the error made in accepting associations among events when in reality they do not exist, supposing that the results of an ecological study are the same that those obtained in a study at an individual level [Morgenstern and Thomas 1993, Steenland and Deddens 1997].

The nature of administrative areas is a significant challenge in mapping works, and therefore it affects all countries with various available aggregation levels [Manley 2014]. The modifiable areal unit problem (MAUP) described by Openshaw, a form of ecological fallacy, is a classic problem when working with areal units [Openshaw 1984, Eagleson and Escobar 2003]. Such issue occurs in all spatial studies with aggregated data. MAUP leads to distort the hotspots due to the artificial characteristics of areal units. Thus there will be variations in results because of data aggregation [Lai *et al* 2009]. The number, size and configuration of areal units will be fundamental determinants in the appearance of a choropleth map and the message to be transmitted [Cromley 2011]. The MAUP is inherent to any field of knowledge carrying out spatial analyses with areal units. For example, one of the most known examples of MAUP effect is the technique named gerrymandering, which consists in the manipulation of district boundaries for electoral purposes. It is assumed that this issue is also reflected in epidemiological studies since the analysis units have been usually defined arbitrarily and not for the purpose of a concrete research work. Some authors have proposed alternative maps to alleviate this effect, such as for example the *dasymetric* maps [Wright 1936, Eicher and Brewer 2001]. In these maps, the statistical areas are divided into areas of relative homogeneity based on complementary information. This means that an attempt is made to get a closer delimitation of the limits for the event studied and to obtain a more realistic description. MAUP has been widely discussed, since it is present in many areas of research when working with spatial data [Arbia 1989, Wong 2009, Flowerdew 2011, Manley 2014].

There are two relevant aspects of the MAUP that must be considered when working with aggregated data into choropleths. These are the aggregation or zonation effect and the scale effect:

a) **Zonation effect:** It consists in variations of results within the geographic units depending on their configuration in a given scale and the same number of spatial units. Additionally, it is assumed that an area in a given scale can be divided up in an infinite number of ways (see **Figure 4**). In the most of cases, the aggregation effect reveals the arbitrary nature of many areal units [Manley 2014]. The zonal effect was studied by Openshaw and Taylor (1979, 1981) assessing its impact through correlation analyses in multiple configurations of areas in Iowa (United States).

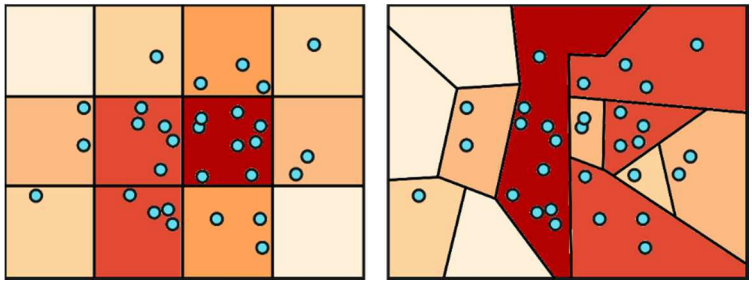


Figure 4. Zonation effect with a given scale and the number of units remained.

b) Scale effect: The hierarchies in areal units give rise to a scale effect when choosing among distinct geographic units for the same spatial data (see **Figure 5**). This means that given the same dataset, different aggregation scales will lead to different analytical results [Parenteau and Sawada 2011]. The more extensive the unit areas the more they hide wide variations within its territorial boundaries. In Spain, the scale effect can be seen when comparing the population density through different levels of aggregation. For example, some provinces have a global high population density but they might contain municipalities with a very low density. This is also observable in epidemiological studies. Small areas depict better the spatial distribution of a health event, while large areal units might hide local differences, reducing in this way the spatial variability [Cromley 2011]. For example, the MAUP has been explored at different scales of aggregation in the city of Madrid using a deprivation index [Cebrecos *et al* 2018]. Additionally, the scale effect might improve or worsen a process of spatial modelling such as removing noise or smoothing data [Manley 2014]. This is a question raised in this doctoral thesis as we face the task of choosing among three Spanish geographic units (provinces, districts and municipalities) for achieving our study purposes.

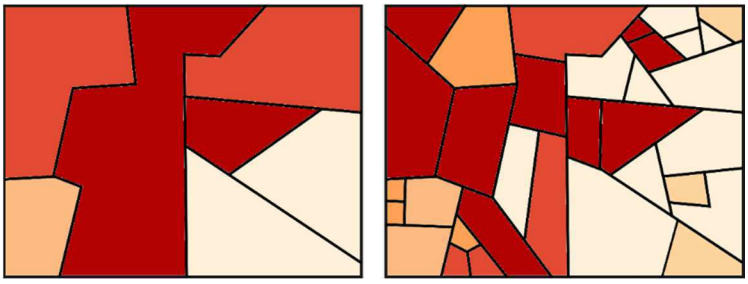


Figure 5. Scale effect given the same dataset.

2.4 Software

Being a doctoral thesis which incorporates different research fields such as geography or epidemiology, a wide variety of tools and software programs have been used. The tasks carried out encompass database management, formatting changes, epidemiological calculations, data query as well as cartographic production. Below, the software programs and their main characteristics are listed in order to achieve the tasks proposed in the achievement of our study.

a) **Stata (version 14):** It is a statistical software program created by StataCorp. Its capabilities include data management, statistical analyses, chart creations, simulations and regression models. This software program was used as a basic tool for the following works: selection and management of RDs database and, statistical calculation of SMRs. These tasks have been carried out in Chapters 3, 4, 5 and 6. More information about this software program can be found at the website https://www.stata.com/.

b) **ArcGIS Desktop (version 10.3):** This is a GIS software used to create, share and analyze spatial data. It has been developed by Environmental Systems Research Institute (ESRI). For the purpose of this doctoral thesis, we have used the desktop application named ArcMap. The main tasks carried out with ArcMap were: data geocoding, integration of geographic information with health variable data (obtained after statistic calculations), feature edition of shapefiles, overlay analysis, as well as the production of maps. Such GIS software has been used in Chapters 3, 4, 5 and 6 and more information can be provided at the website http://www.esri.com/software/arcgis.

c) **R (version 3.3):** This is a language free software program for statistical computing and graphics. It is supported by the R Foundation for Statistical Computing. R functions can be extended via packages such as R-INLA that was used to create the adjacency matrix and to perform the Bayesian computing in order to obtain the Smoothed-SMRs and PPs. Besides, the R software program was used with the purpose of calculating age-adjusted mortality rates in temporal analyses. This programming software was used in Chapters 3, 4 and 5. More information can be found at the websites https://www.r-project.org/ and http://www.r-inla.org/.

d) **Joinpoint (version 4.2.0.2):** This is a kind of free statistical software program to analyze changing points in temporal trends and the level of significance. It was

developed by the National Cancer Institute, United States. This program was used in Chapters 4 and 5 in the task of detecting changing points in both age-adjusted mortality rates and in the average age at death. Likewise, it was used to calculate the annual percent changes (APC) in indicators during our study period. More information can be found at the website https://surveillance.cancer.gov/joinpoint/.

e) **IBM SPSS Statistics (version 22):** This is a software package developed by IBM Corporation for statistical analysis in social sciences. This program was used for smoothing temporal age-adjusted mortality rates, a task carried out in Chapters 4 and 5. Additional information about this software can be consulted at https://www.ibm.com/products/spss-statistics.

f) **GeoDa (version 1.8.12):** This is a free open source software application developed by Luc Anselin to conduct spatial data analysis, spatial autocorrelation and spatial modelling. It was used to perform autocorrelation and cluster analyses in Chapter 3. This program can be downloaded from the website http://geodacenter.github.io/.

Chapter 3

Multi-scale aggregation data displaying rare-disease epidemiology

Multi-scale aggregation data displaying rare-disease epidemiology

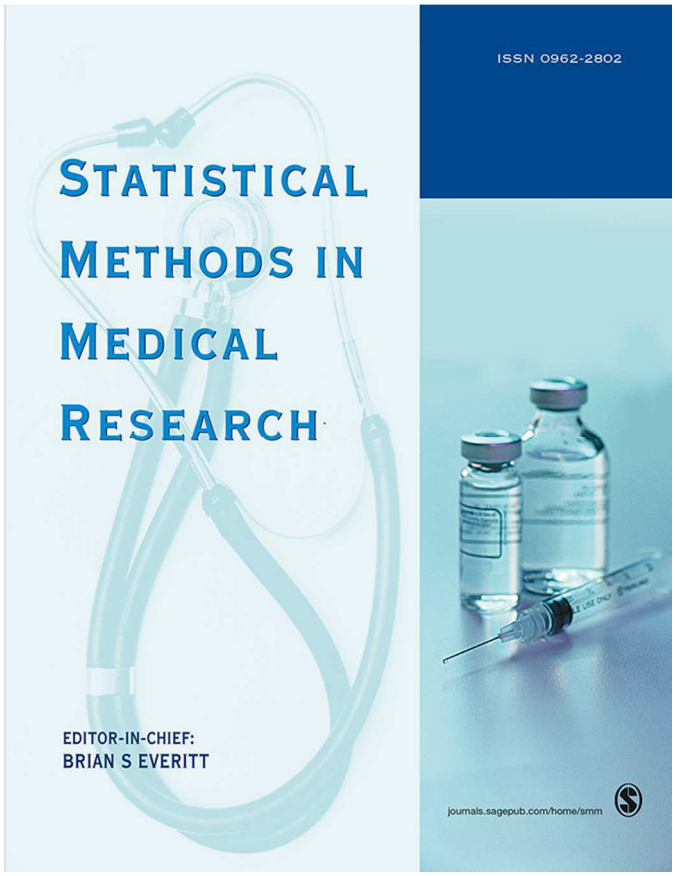
Article submitted on 12 July 2019 to: Statistical Methods in Medical Research

Authors: Germán Sánchez-Díaz [a-c], Verónica Alonso-Ferreira [a,b], Manuel Posada de la Paz [a,b] and Francisco Escobar [c].

[a] Institute of Rare Diseases Research (IIER), Instituto de Salud Carlos III (ISCIII) Madrid, Spain

[b] Centre for Biomedical Network Research on Rare Diseases (CIBERER), Madrid, Spain

[c] Department of Geology, Geography and Environmental Sciences, University of Alcalá, Spain



Multi-scale aggregation data displaying rare-disease epidemiology

Abstract

This study analyzes the effect of modifiable areal unit problem (MAUP) comparing mortality data of a specific rare disease (Huntington) in three different Spanish levels of spatial aggregation. The objective is to compare mortality indicators and cartographic visualizations in order to soundly advise on the most optimum aggregation level according to its population, covered area and number of cases. We designed an adjacency ratio to observe the effect of neighborhood relationships among three geographic units: province, district and municipality. For each level of aggregation, we performed epidemiological indicators of mortality as well as local indicators of spatial association. Maps were plotted with user-defined intervals to compare visual and statistical differences. MAUP-related effects are particularly noticeable in these relatively infrequent events as rare diseases. We found that district displayed the highest indicator stability in the adjacency ratio and showing optimum characteristics for spatial resolution and amount of information revealed through plotting. This help in the election of the working scale can be used with other diseases or levels of aggregation as a first step of more advanced epidemiological analyses.

Key words: aggregated data; epidemiology; geographic unit; scale effect; rare disease; cartography.

3.1 Introduction

Since the advent of Geographic Information Systems (GIS) in the late 1960s, they have been increasingly used by the health sector.[1-2] While numerous examples of their application range from service planning to epidemiological analysis of disease outbreaks, emergency-service delivery and health promotion campaigns[3], analysis of rare diseases (RD) has not yet fully benefited from this integration with the same degree of intensity.[4]

RDs are defined as those affecting not more than 5 people in 10,000 in the European Union (UE).[5] Some 6,000-8,000 RDs have been identified, they may affect 6% to 8% of the EU population, and thus they have been recognised as posing a major global public health problem.[6,7] GIS integration in the field of RDs is not at an advanced stage, with the bulk of research focusing on the most common diseases.[8,9] This lag is essentially caused by the analytical difficulties deriving from its low prevalence, leading to an excess of zeros in results and cartographical representations, especially when it comes to smoothing data with exploratory data-analysis tools.[10] Moreover, disease-classification systems are mainly geared to high-prevalence conditions, and this is an important concern for epidemiological RD studies and subsequent map plotting or analysis tasks since both analysis and cartographical representations can therefore be said to be constrained by data availability and quality.[10-12]

The spatial distribution of diseases is one of the most important tasks in spatial epidemiology in order to allocate resources in areas that present a higher risk.[13] Mortality is one of the most widely used epidemiological indicators for the study of diseases. It helps ascertain the course and severity of a given disease, enhance its visibility, and highlight the scale of impact that its inequalities have on public health.[10] Recent decades have witnessed great progress in spatio-temporal studies aimed at observing time trends in the spatial behaviour of mortality and its relation with other variables with applications in fields as diverse as the environment, economy or health.[14,15] For example, in the case of epidemiology, the best known and most numerous spatio-temporal studies are those on the monitoring and control of infectious-disease outbreaks.[14,16-18]

On the other hand, spatial hierarchies in which data are aggregated do not necessarily reflect their nature, this is an alteration associated with the so-called Modifiable Areal Unit Problem (MAUP). According to the MAUP, when a continuous geographical event is represented in artificial units such as census tracts or municipalities, the effect of unit size, configuration of forms, and especially scale may lead to diverse results and, by extension, differing interpretations.[19] The MAUP has been extensively reviewed in the scientific literature since the mid 1980's. However, in the case of RD, there is a remarkable gap in the

bibliography. A new insight in the MAUP in relation to RD is much needed as the aggregation of RDs data does not respond to mainstream analytical approaches. In RDs aggregation is a must for two reasons not found in other geographical phenomena: the need to preserve confidentiality[20], and the infrequency of RDs cases, meaning that the more detailed the scale, the scarcer the data.[10] This begs the question of which aggregation level would be the most appropriate for the cartographic representation of a RD. Local government (municipality) and postal-code units display a high degree of heterogeneity in size and population, and their adoption would thus result in an excess of geographical units with zero cases. But also a smaller number of entities having larger populations, such as states or regions, would undermine the work of geographical analysis and visualisation.[21,22] It has been proposed the application of a set of criteria such as population, shape unit, event distribution or biological relevance in order to allow the choice of the most appropriate geographical unit in an epidemiological study.[23]

Accordingly, this study aimed to analyse the effect of different levels of spatial aggregation on the analysis of RD mortality data, using Huntington's disease as example. The final objective is to show the process in the choosing of the most appropriate spatial units proposed in terms of number of cases, population and area by comparing statistical results and cartographic visualizations.

3.2 Methods

This study was conducted in Spain, covering an area of approximately 500,000 sq.km. and a population of about 47 million inhabitants in 2018, according to the National Statistics Institute (*Instituto Nacional de Estadística*/INS). The Spanish territory is divided first in autonomous communities, and then into 50 provinces and two autonomous cities, as delimited by a statutory administrative division dating back to 1833 (under the original Act as amended).[24,25] Each province contains a variable number of municipalities (ranging from 34 to 371), with more than 8,000 entities in all. These constitute the most basic level of government having independent legal personality. Districts (*comarcas*) are entities which are larger than municipalities but smaller than provinces, and group neighbouring municipalities according to historical and geographic criteria. For study purposes, we used the 326-district classification provided by the Ministry of Agriculture, Fisheries, Food and Environment (*Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente/MAPAMA*).[26] **Figure 1** depicts the above-mentioned administrative subdivisions, showing the top-down scale adjacency model (from provinces to municipalities) used in the spatial smoothing process.

Mortality data according to cause of death for the last 15 years (1999-2013) were drawn from the death statistics published on the INS website. By way of example of a low-prevalence disease, Huntington's disease was selected as the underlying cause of death, using the International Classification of Diseases, 10th Revision (ICD-10) code G10. This yielded a total of 1,090 HD-related deaths for the study period.

Population data for the period 1999-2013 were obtained from the Annual Municipal Register published on the *INS* website, and then aggregated at a district and provincial level. Spatial information (shapefiles in vector format) for each level of aggregation were supplied by the MAPAMA and the Spanish National Geographical Institute (*Instituto Geográfico Nacional*).

Figure 1. Spatial administrative boundaries included in this study in Spain (with number of entities).

Municipalities and districts were grouped by province. Adjacency ratios for districts and municipalities were estimated (i.e., in the case of district: the total number of entities per district in a province divided by average number of adjacent entities). Both ratios were normalised on a scale ranging from -2 to +2 to allow for comparison between the spatial distributions in districts and municipalities.

The standardised mortality ratio (SMR) is a commonly used indicator in spatial-epidemiology studies because it enables comparison between observed and expected deaths in each spatial unit according to an age-related reference rate (in our case the Spanish rate).[27] Indirect standardization by five-year age groups was used for expected deaths calculation. The SMR is the ratio between the observed number of deaths and the number of deaths would be expected. In our case, we calculated the respective SMRs for provinces, districts and municipalities for the study period. The results obtained were divided into user-defined intervals, following a classification based on a value of around 1.00 (0.80-1.20) representing the normal risk. Intervals above 1.20 show a higher than expected risk, while those below 0.80 show one lower than expected. This kind of user-defined interval chosen, widely used in epidemiology, was supported by visual legibility concerns as well as the need to split the variable in two upper and lower intervals around the medium interval (0.80- 1.20).[28] It must be noted that SMR results have an additional interval for entities with value 0.00. SMR 95% confidence intervals (CIs) were calculated, and mortality distribution by aggregation level was assessed using the Chi-square test.[29]

The smoothing process method for SMR is used in order to reduce the distortion produced by the excess of zeros and the extreme values due to variation in population of the geographic units. Smoothed SMRs were estimated using the number of observed and expected deaths in each spatial unit as well as the equivalent values in neighbouring areas (adjacency criteria), thus enabling spatial risk patterns to be detected or discarded. To this end, an adjacent matrix had to be created for each administrative level using the *rook's* case contiguity which is based on neighbourhoods sharing physical borders.[30] After the creation of the matrices, we used the conditional autoregressive model proposed by Besag, York and Mollié to carry out the estimation of the risk taking into account two randomised effects: spatial and heterogeneous.[31] To enable visual and statistical comparison with cartographic results, the smoothed SMR results were classified in the same way as the SMR results.

We examined spatial autocorrelation and clustering to assess the results at each level of aggregation and to compare which would provide the richest information for the purpose of further in-depth study or displaying data in a cartographical software programme. In order to avoid biased results, global and local tests were only carried out with SMR, not with smoothed data. Moran's Index was calculated to assess autocorrelation in SMR values for provinces, districts and municipalities (Moran, 1950), and a local indicators of spatial association (LISA) analysis was performed to detect significant SMR clusters.[32,33] This local analysis made it possible to ascertain the existence of geographic units with neighbours sharing similar values, i.e., high values or dissimilarities.

Statistical estimation for SMR purposes was performed using the Stata computer software program. SMR smoothing process was carried out in R software with the use of R-INLA package (*Integrated nested Laplace approximation*). This smoothing process is faster in computer time than Monte Carlo (McMC) methods in model fitting and inference which makes it clearly advantageous.[34] GeoDA software was used for autocorrelation and clustering analysis. Finally, mapping visualisations were performed using the ArcGIS software.

3.3 Results

As shown in **Table 1**, at each level of aggregation entities had different sizes and population, especially in provinces and municipalities. The differences in sizes, which amounted to as much as 10-fold among provinces, 20-fold among districts and 5,000-fold among municipalities, were important and meant greater uniformity across provinces and greater heterogeneity across municipalities. Similarly, population differences were also in evidence, ranging from 70-fold among provinces to 30,000-fold among districts and 1 million-fold among municipalities. To sum up, size and population displayed great diversity, not only among but also within aggregation levels.

Table 1. Size and population statistics for the three aggregation levels used in this study.

		Province	**District**	**Municipality**
	n	52	326	8,123
Area	Average (sq.km.)	9,725	1,542	62
	Standard deviation	5,095.82	896.03	92.05
	Min (sq.km.)	13	13	0.03
	Max (sq.km.)	21,751	5,396	1,750
Population*	Average (inhabitants)	853,588.10	136,155.20	5,467.67
	Standard deviation	1,074,190.70	342,422.30	15,520.64
	Min (inhabitants)	71,379	1625	5
	Max (inhabitants)	5,956,777	4,540,006	3,116,900

* Average population for the 1999-2013 period.

Comparison of the normalised adjacency ratios showed that districts were more stable than municipalities (**Figure 2**). Whereas districts displayed more balanced ratios, municipalities had few cases in the intermediate range (14 provinces in the district analysis versus 5 in the municipal analysis, with an index ranging from -0.20 to 0.20). The greater the inter-provincial differences in the index, the more unstable the smoothed SMR results.

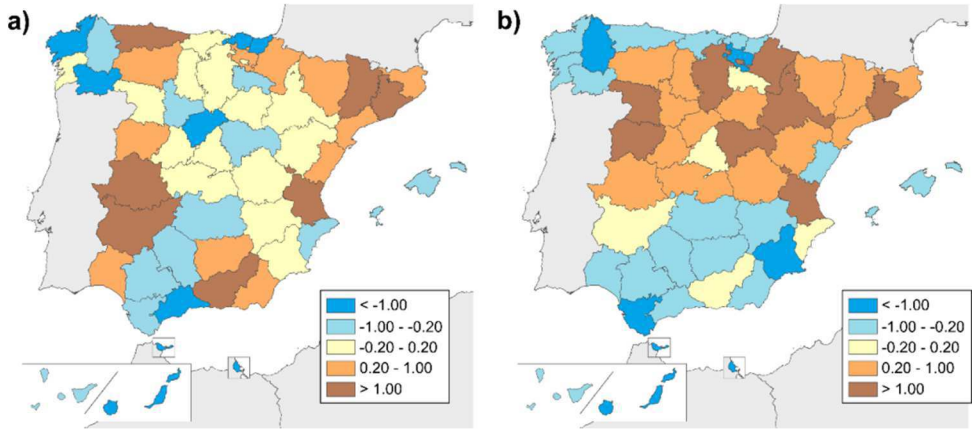


Figure 2. Normalised adjacency ratios in provinces. Aggregation by: **a)** district; and **b)** municipality.

As can be observed from **Table 2**, the distribution of SMR values was significantly different ($p<0.001$) depending on the level of aggregation used. Note should be taken of the percentage of geographic units with no cases, since this points to a lack of information (96.44% at the municipal level). The SMR value will be zero if there are no observed deaths, a common occurrence in the majority of municipalities. Among the selected intervals, it was "province" that registered the most uniform distribution of SMRs, while at the district level, 42.94% corresponded to the "No cases" category. The number of units and percentages of values not coming within the 95% CI accounted for only 10% and 5% of the provincial and district totals respectively ($p<0.001$), while in municipalities there were hardly any values lying outside the CI.

Table 2. Number of provinces, districts and municipalities (percentage in brackets) by interval of Standardized Mortality Ratio and Smoothed Standardized Mortality Ratios after Bayesian model process for Huntington's disease (1999-2013 period).

	Province n (%)		**District** n (%)		**Municipality** n (%)	
Interval	**SMR**	**Smoothed SMR**	**SMR**	**Smoothed SMR**	**SMR**	**Smoothed SMR**
No cases (0.00)	4 (7.69)	-	140 (42.94)	-	7,835 (96.44)	-
< 0.60	7 (13.46)	1 (1.92)	31 (9.51)	7 (2.15)	11 (0.14)	0 (0.00)
0.60 - 0.80	10 (19.23)	12 (23.08)	16 (4.91)	35 (10.74)	7 (0.09)	0 (0.00)
0.80 - 1.20	15 (28.85)	30 (57.69)	39 (11.96)	202 (61.95)	17 (0.21)	8,123 (100)
1.20 - 1.40	2 (3.85)	4 (7.69)	11 (3.37)	41 (12.58)	8 (0.10)	0 (0.00)
> 1.40	14 (26.92)	5 (9.62)	89 (27.31)	41 (12.58)	245 (3.02)	0 (0.00)
Total	52 (100)		326 (100)		8,123 (100)	

Smoothed SMRs after Bayesian modelling process are depicted in **Figure 3** by province, district and municipality. In a first visual analysis, the SMR smoothing procedure blurs the mortality differences previously detected at a municipal level. Depending on unit size, the percentage of units with a significantly higher or lower risk death in terms of smoothed SMRs varied ($p<0.001$) from 26% in provinces and 19% in districts down to 0% in municipalities.

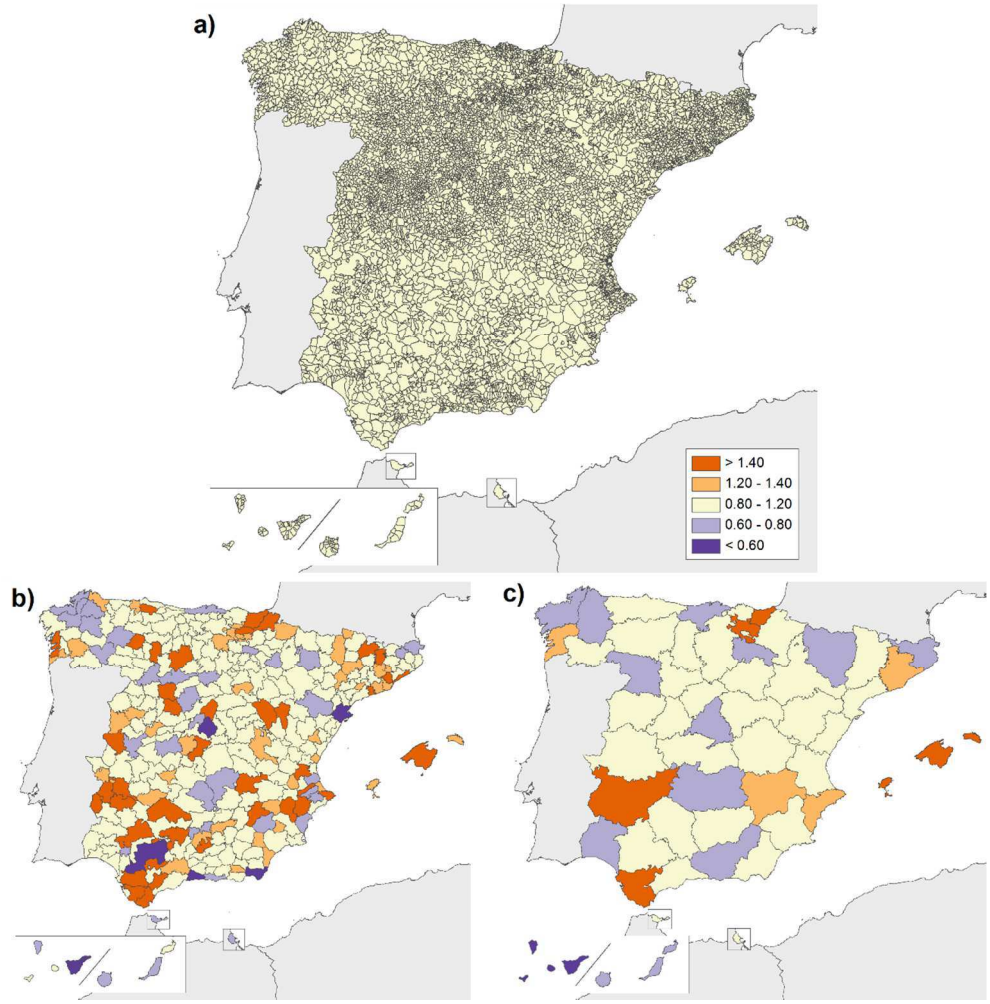


Figure 3. Smoothed Standardized Mortality Ratios for Huntington's disease 1999-2013. Levels: **a)** municipal; **b)** district; **c)** provincial.

Moran's Index showed no overall positive or negative autocorrelation at any of the three levels of aggregation. Indeed, the negative z-score for all levels implies that there is a disperse pattern in SMR results in Spain. None of the results was statistically significant ($p>0.05$).

In **Figure 4**, a choropleth map depicts the units with a significant local result Moran (or LISA) classified by type of spatial statistical correlation, using the SMR results. The values "High-High" have been observed in the West of Spain while they were detected in the North and in the South at a district and provincial level. "Low-Low" cluster values are distributed throughout the country at the municipal level, in the North and Southwest at a district level, while it was no detected at a provincial level. "High-Low" is result of an entity with outlier value surrounded by low values (it was found in a disperse way at municipal level, in the

Northern half of the country in districts and not found at a provincial level). Lastly, "Low-High" entities (significant low outliers of SMR surrounded by high values) are random dispersed at a municipal level, more predominantly in the Northern half at a district level, and in the Southwest at a provincial level. The percentages of statistically significant values were as follows: 9.61% at a provincial level (5 to 52 entities); 13.50% at a district level (44 to 326 entities); and 8.10% at a municipal level (658 to 8,123 entities).

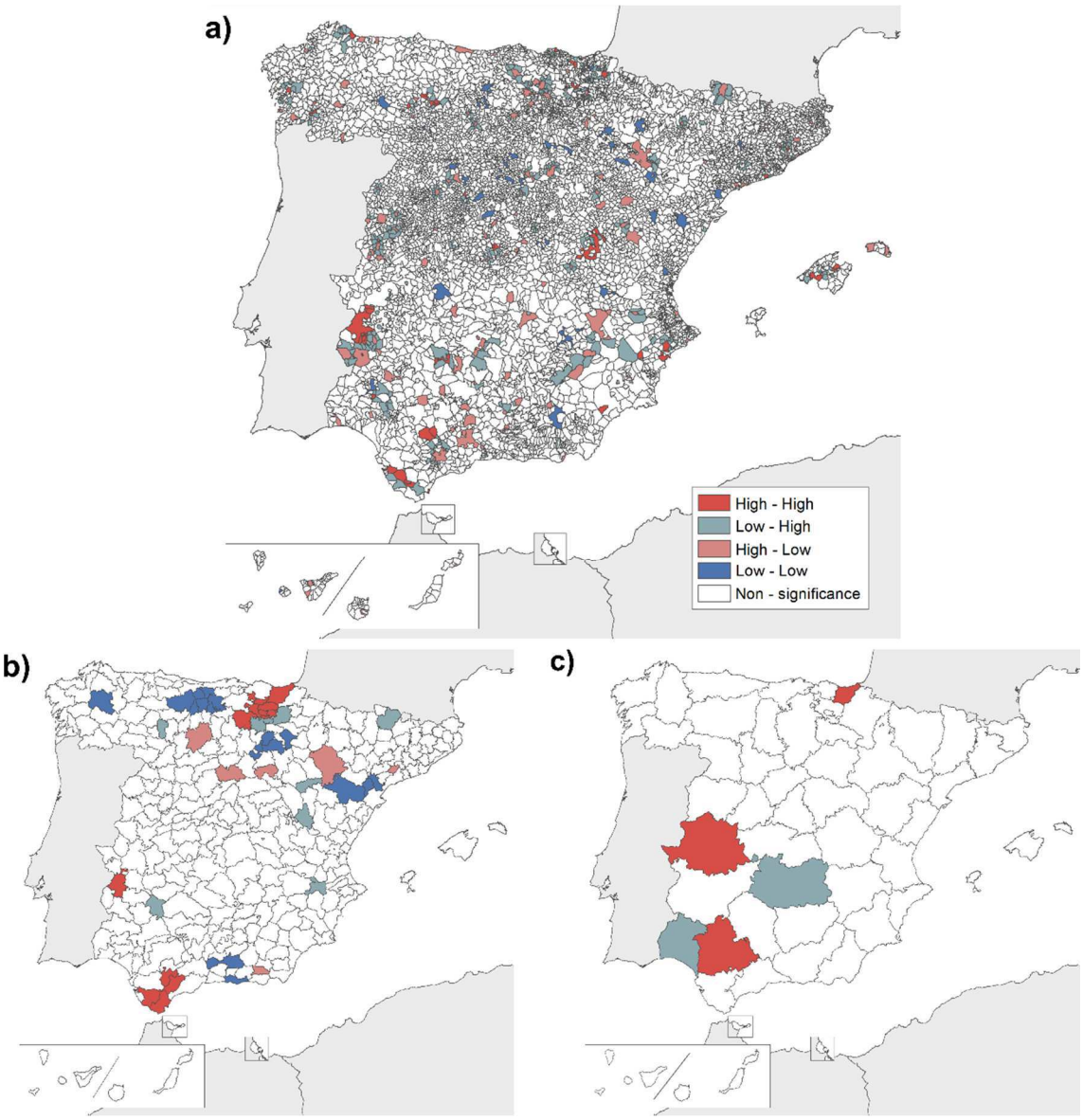


Figure 4. LISA cluster map of Huntington's disease Standardized Mortality Ratios: 1999-2013. Levels: **a)** municipal; **b)** district; **c)** provincial.

3.4 Discussion

This study performed a statistical and a descriptive analysis of HD mortality at a municipal, district and provincial level in Spain. The results are shown in map and table form, thereby facilitating the goal of observing the differences among and within three data-aggregation levels. In addition to comparing the estimated rates, the significance of each result obtained is of key importance when it comes to supporting decision-making and reducing uncertainty surrounding unstable data. In this respect, cartography is an integral part of spatial and temporal studies and any kind of study complementing numerical results.[35]

As will be seen from the tables and figures, death data obtained from the INS at municipal level were used as a basis and then added at the two other levels of aggregation (district and province) to enable the comparison of results. In so doing, spatial patterns change as a consequence of MAUP[36], something that must be faced when aggregating data in administrative units of different size and distribution. This said, rather than proposing solutions for MAUP, the aim of this study was to explore and identify this issue in order to help other researchers when they are working with their own datasets. Also, it is sought that researchers can determine which is the more stable geographic unit in that context ensuring the most optimum cartographic work for data displaying. In regards to this, the total number of geographic units a key element for numerous types of statistical tests in a geospatial context.[37,38] To have geographic death data at a municipal level and not individual responds to the pressing need to preserve the confidentiality of the deceased's personal data, which are especially sensitive when dealing with sparsely populated municipalities.[37] In such cases, anonymity is quite difficult to maintain, thus rendering it necessary to adopt any of the measures outlined above.

Rather than calculating crude death rates, we chose instead to estimate SMRs because mortality is strongly related to age group and entities with ageing population can distort data when comparing with other entities no sharing the same characteristics. Even so, SMRs may constitute a highly unstable data source when working with very low numbers of cases.[39] The enormous number of geographic units with unobserved cases and the heterogeneity of population data in municipalities make for many zeros, extreme positive values and, as a consequence, a lack of robustness.[10]

The adjacency ratio is a purpose-designed index created to understand the relationship among a number of entities and their contiguous areas. Obtaining extreme results, lying far from the central value, highlighted the irregularity in the subdivisions that make up Spain, with particularly striking inequalities in municipalities. This, combined with the low number of cases in each entity, results in a blurring in the smoothing procedure when

applied to municipal datasets, making the search for spatial patterns much more difficult. When a rate is smoothed, the outcome is influenced by the value of neighbouring units. Taking into account the fact that each municipality is contiguous to a further 5.94 municipalities (average number of neighbours) and that 96% of these have no observed cases, the SMRs of units with observed cases were smoothed by removing the extreme values. While the number of contiguous units is slightly lower in districts and provinces (5.50 and 4.46 respectively), the difference is that the percentage of units with no observed cases is significantly lower. In our case, we have found that a municipality is not the most suitable level of aggregation for the study of spatial patterns in rare diseases when it comes to countries such as Spain or others with similarly small administrative bodies. The loss of information is less severe when taking a district or province as the frame of reference, since the smoothing process may lead to a better result for a higher level of aggregation than that of a municipality, as supported by other studies.[40] At a provincial level, the low spatial resolution is a disadvantage observed, since aggregated data are divided into a low number of entities (52) when compared to the other two levels available. It should be borne in mind that the aggregation levels were not designed for RDs research, and this is why distortions appear when it comes to interpreting the results obtained.[41]

The LISA analysis showed that the percentage significant values in the geographic units was lower in municipalities than in provinces. It can thus be concluded that results present lower variability in areas with no registered cases at the municipal level. It is also be noted that if it is observed a disease cluster at a high level of detail as the municipal, the aggregation in less detailed scales could lead to loss this information.

The low number of cases in municipalities (or census sections when available) used to be a general limitation in epidemiological studies, leading to aggregate data at a higher level (district or province). In some cases, using a very detailed scale may not ensure the reveal of the desired pattern through geospatial analysis.[42] Mortality of diseases such as cancer, which bring together a high number of cases (although there are infrequent types), allows for large-scale statistical analysis at a census level.[43] Even it is possible to work at a census level. Some of the criteria proposed by Arsenault[23] and his colleagues regarding the selection of geographical units, have been considered in this study: Communicability of results, intra-unit homogeneity, variation in population size and variation in area size. Districts may be the most adequate geographical unit within which to implement public health objectives, due to their consisting of a manageable area. The competent authorities in Spain are the seventeen Regional Health Services. Each Service could implement policies within its own districts, taking into account a greater awareness of health issues at an intermediate level. The application of the second criterion also reveals the district as the

preferable geographical unit, showing better intra-unit homogeneity than that within a province or a region. Lastly, the district also meets the two last criteria, regarding variation in both population and area sizes. Big variations may lead to an increased bias in the identification of spatial patterns, due to the fact that epidemiological studies are generally based on certain distributional assumptions.[44]

Obviously, these Spanish geographic units do not correspond to other countries with equal exactness, although this study could be replicated in other study areas with their respective aggregations level from least to greatest detail. For example in the United States it could be replicated from the municipality, to county or state[45] or, in the case of Portugal, from *freguesias* to municipalities and districts.[46] In countries no possessing an administrative unit equivalent to the Spanish district, an aggregation ad hoc of municipal units would be equally appropriate and would count on the same characteristics outlined by Arsenault *et al* (2013).

3.5 Conclusion

The usefulness of this study lies in the possibility of replicating the procedure to another disease or disciplines, provided that a low number of cases of a variable as the basis of the analysis. The results presented could be a support to choosing the optimal way of mapping and analysing an event, thus successfully completing the first step in a bivariate or multivariate analysis. For this purpose, attention must be paid, not only to the characteristics of the available data (quantity and spatial distribution), but also to the heterogeneity of the entities into which the study area is divided. In our case, "Districts", with a medium size per entity of 1,542 sqkm and a medium population of 136,000 inhabitants, was found to be the best option to analyse and plot the mortality indicators of a specific RD, having taken into account the balance between its size, its total inhabitants and the number of events registered in each entity.

3.6 References

1. Olaya V. Sistemas de Información Geográfica. CreateSpace Independent Publishing Platform; 2016.

2. Shaw NT, McGuire SK. Understanding the use of geographical information systems (GISs) in health informatics research: a review. BMJ Health Care. 2017; 24(2):228-233.

3. Lai P, So F, Chan K. Spatial Epidemiological Approaches in Disease Mapping and Analysis. Boca Raton: CRC Press; 2009.

4. Gómez-Rubio V, López-Quílez A. Statistical methods for the geographical analysis of rare diseases. In: Posada M, Groft SC, editors. Rare Diseases Epidemiology. Dordrecht: Springer; 2010. p. 151-172.

5. Decision No 1295/1999/EC of the European Parliament and of the Council of 29 April 1999 adopting a programme of Community action on rare diseases within the framework for action in the field of public health (1999–2003). (1999) Commission of the European Communities, Brussels. Retrieved from: http://ec.europa.eu/health/archive/ph_overview/previous_programme/rare_diseases/raredis_wpgm99_en.pdf

6. Posada M, Alonso V, Bermejo E. Enfermedades raras. Madrid: Catarata; 2016.

7. Groft SC, Posada M. Rare Diseases: joining mainstream research and treatment based on reliable epidemiological data. In: Rare Diseases Epidemiology, Update and overview. Springer International Publishing AG; 2017. p. 3-21.

8. Alonso V, Villaverde-Hueso A, Hens M, Morales-Piga A, Abaitua I, Posada de la Paz, M. Increase in motor neuron disease mortality in Spain: temporal and geographical analysis (1990-2005). Amyotroph Lateral Scler. 2011; 12:192-198.

9. Sánchez-Díaz G, Arias-Merino G, Villaverde-Hueso A, *et al.* Monitoring Huntington's disease mortality across a 30-year period: geographic and temporal patterns. Neuroepidemiology. 2016; 47(3-4):155-163.

10. Posada M, Villaverde-Hueso A, Alonso V, *et al.* Rare diseases epidemiology research. Adv Exp Med Biol. 2010; 686:17-39.

11. Rytkönen M.J. Not all maps are equal: GIS and spatial analysis in epidemiology. Int J Circumpolar Health. 2004; 63(1): 9-24.

12. Grady SC, Enander H. Geographic analysis of low birthweight and infant mortality in Michigan using automated zoning methodology. Int J Health Geogr. 2009; 8:10.

13. Aregay M, Lawson AB, Faes C, Kirby RS. Bayesian multi-scale modeling for aggregated disease mapping data. Stat Methods Med Res. 2017; 26(6): 2726-2742.

14. Lawson AB, Banerjee S, Haining RP, Ugarte MD. Handbook of Spatial Epidemiology. Boca Raton: CRC Press; 2016.

15. Meyer S, Held L, Höhle M. Spatio-Temporal analysis of epidemic phenomena using the R Package surveillance. J Stat Softw. 2016; 77(11):1-55.

16. Waller L, Gotway C. Applied Spatial Statistics for Public Health Data. Hoboken: John Wiley & Sons; 2004.

17. Meliker J, Sloan C. Spatio-temporal epidemiology: Principles and opportunities. Spat Spatiotemporal Epidemiol. 2001; 2(1):1-9.

18. Redondo-Bravo L, Suárez-Rodríguez B, Fernández-Martínez B, *et al*. Epidemia por virus Zika. Respuesta desde la salud púbica en España. Rev Esp Salud Publica. 2018; 92:e1-e16.

19. Openshaw S. The modifiable areal unit problem. Norwich: Geo Abstracts University of East Anglia; 1984.

20. Zandbergen PA. Ensuring confidentiality of geocoded health data: Assessing Geographic masking strategies for individual-level data. Adv Med. 2014; 2014:567049.

21. Moore D, Carpenter T. Spatial analytical methods and Geographic Information Systems: Use in health research and epidemiology. Epidemiol Rev. 1999; 21(2):143-161.

22. Arab A. Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. Int J Environ Res Public Health. 2015; 12(9):10536-10548.

23. Arsenault J, Michel P, Berke O, Ravel A, Gosselin P. How to choose geographical units in ecological studies: proposal and application to campylobacteriosis. Spat Spatiotemporal Epidemiol. 2013; 7:11-24.

24. Moreno L. Ethnoterritorial Concurrence and Imperfect Federalism in Spain. In: Bertus de Villiers, editor. Evaluation Federal Systems. Cape Town: Juta & Co/Martinus Nijhoff Publishers; 1994. p. 162-193.

25. Gómez-Díaz J. División territorial de España. Provincias y partidos judiciales. 175 años. Toletum. 2008; 55:151-175.

26. Fernández J, Checa M, Esteban B, *et al*. Caracterización de las comarcas agrarias de España. Madrid: Centro de Publicaciones del Ministerio de Medio Ambiente y Medio Rural y Marino; 2001.

27. Royo MA, Moreno JD. Método epidemiológico. Madrid: ENS - Instituto de Salud Carlos III; 2009.

28. Cauvin C, Escobar F, Serradj A. Thematic Cartography and Transformations. 1st ed. London: Wiley; 2010.

29. Greenwood PE, Nikulin M.S. A guide to chi-squared testing. New York: Wiley; 1996.

30. Smirnov O, Anselin L. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. Comput Stat Data An; 2001; 35(3):301-319.

31. Besag J, York J, Mollìè A. A Bayesian image restoration with two applications in spatial statistics. Ann Inst Math. 1991; 43:1-59.

32. Moran P. Notes on Continuous Stochastic Phenomena. Biometrika. 1950; 37(1):17-23.

33. Anselin L, Syabri I, Kho, Y. GeoDa: An introduction to spatial data analysis. Geogr Anal. 2006; 38:5-22.

34. Rue H, Riebler A, Sørbye S, *et al.* Bayesian Computing with INLA: A Review. Annu Rev Stat Appl. 2016; 4(1): 1-28.

35. Dykes J, Brunsdon C. Geographically weighted visualization: Interactive graphics for scale-varying exploratory analysis. IEEE Trans Vis Comput Graph. 2007; 13(6):1161-1168.

36. Kanaroglou P, Delmelle E, Páez A. Spatial analysis in health geography 1st ed. London: Routledge; 2015.

37. Goodwin S, Dykes J, Slingsby A, Turkay C. Visualizing multiple variables across scale and geography. IEEE Trans Vis Comput Graph. 2015; 22(1):599-608.

38. Lawson AB. Statistical Methods in Spatial Epidemiology. Chichester: John Wiley & Sons; 2006.

39. Ugarte MD, Ibáñez B, Militino AF. Modelling risks in disease mapping. Stat Methods Med Res. 2006;15:21-35.

40. Cockings S, Martin D. Zone design for environmental and health studies using pre-aggregated data. Soc Sci Med. 2005; 60:2729-2742.

41. Elliot P, Wartenberg D. Spatial epidemiology: Current approaches and future challenges. Environ Health Perspect. 2004; 112(9):998-1006.

42. Beale L, Abellan JJ, Hodgson S, Jarup L. Methodologic issues and approaches to spatial epidemiology. Environ Health Perspect. 2008; 116(8): 1105-1110.

43. Puigpinós-Riera R, Marí-Dell'Olmo M, Gotsens M, *et al.* Cancer mortality inequalities in urban areas: a bayesian small area analysis in Spanish cities. Int J Health Geogr. 2010; 10:6.

44. Berke O. Exploratory disease mapping: kriging the spatial risk function from regional count data. Int J Health Geogr. 2004; 3:18.

45. Pavković A, Radan P. Creating New States: Theory and Practice of Secession. Burlington: Ashgate; 2007.

46. Nunez-Silva C. Political and administrative decentralization in Portugal: Four decades of democratic local government. In: Nunes-Silva C, Buček J, editors. Local Government and Urban Governance in Europe. The Urban Book Series. Cham: Springer; 2017. p. 9-32.

Chapter 4

Monitoring Huntington's disease mortality across a 30-year period: Geographic and temporal patterns

Monitoring Huntington's disease mortality across a 30-year period: Geographic and temporal patterns

Published in: Neuroepidemiology

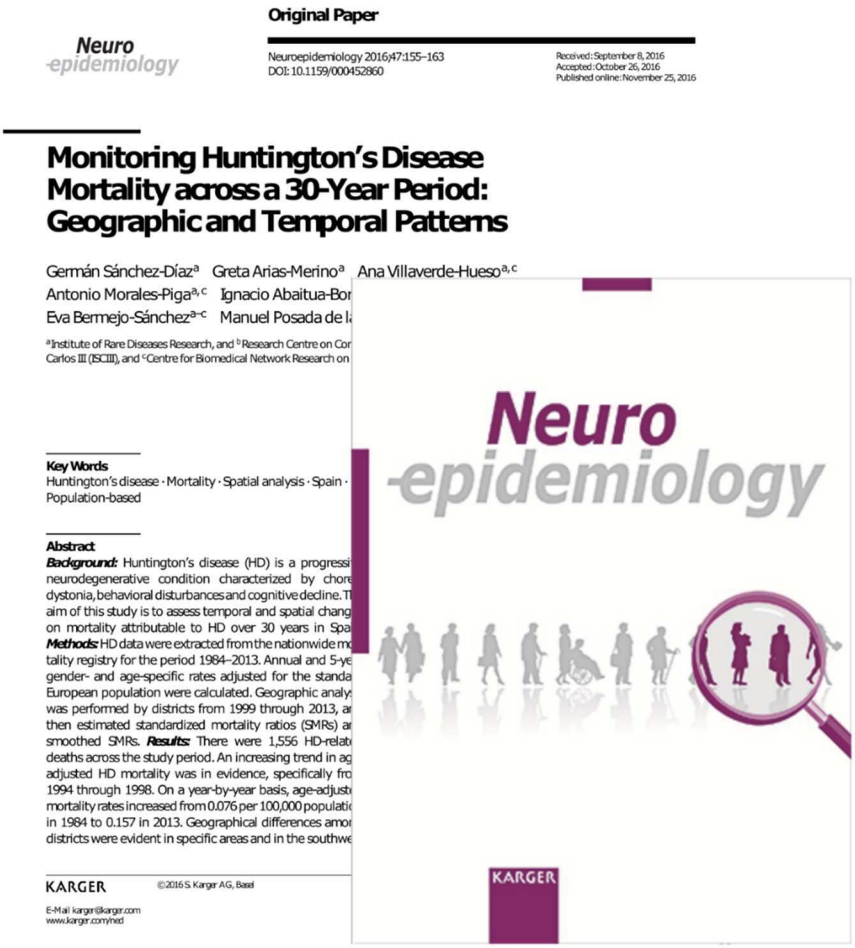
Neuroepidemiology 2016, 47(3-4): 155-163; doi: 10.1159/000452860

Authors: Germán Sánchez-Díaz [a], Greta Arias-Merino [a], Ana Villaverde-Hueso [a,c], Antonio Morales-Piga [a,c], Ignacio Abaitua-Borda [a,c], Manuel Hens [a,c], Eva Bermejo-Sánchez [a-c], Manuel Posada de la Paz [a,c] and Verónica Alonso-Ferreira [a,c].

[a] Institute of Rare Diseases Research (IIER), Instituto de Salud Carlos III (ISCIII) Madrid, Spain

[b] Research Centre on Congenital Abnormalities (CIAC), Instituto de Salud Carlos III (ISCIII) Madrid, Spain

[c] Centre for Biomedical Network Research on Rare Diseases (CIBERER), Madrid, Spain



Original Paper

Neuro
-epidemiology

Neuroepidemiology 2016;47:155–163
DOI: 10.1159/000452860

Received: September 8, 2016
Accepted: October 26, 2016
Published online: November 25, 2016

Monitoring Huntington's Disease Mortality across a 30-Year Period: Geographic and Temporal Patterns

Germán Sánchez-Díaz[a] Greta Arias-Merino[a] Ana Villaverde-Hueso[a,c]
Antonio Morales-Piga[a,c] Ignacio Abaitua-Bor
Eva Bermejo-Sánchez[a–c] Manuel Posada de l

[a]Institute of Rare Diseases Research, and [b]Research Centre on Cor
Carlos III (ISCIII), and [c]Centre for Biomedical Network Research on

Key Words
Huntington's disease · Mortality · Spatial analysis · Spain ·
Population-based

Abstract
Background: Huntington's disease (HD) is a progressi
neurodegenerative condition characterized by chore
dystonia, behavioral disturbances and cognitive decline. T
aim of this study is to assess temporal and spatial chang
on mortality attributable to HD over 30 years in Spa
Methods: HD data were extracted from the nationwide m
tality registry for the period 1984–2013. Annual and 5-ye
gender- and age-specific rates adjusted for the standa
European population were calculated. Geographic analy
was performed by districts from 1999 through 2013, a
then estimated standardized mortality ratios (SMRs) a
smoothed SMRs. ***Results:*** There were 1,556 HD-relat
deaths across the study period. An increasing trend in ag
adjusted HD mortality was in evidence, specifically fro
1994 through 1998. On a year-by-year basis, age-adjust
mortality rates increased from 0.076 per 100,000 populati
in 1984 to 0.157 in 2013. Geographical differences amo
districts were evident in specific areas and in the southwe

KARGER ©2016 S. Karger AG, Basel

E-Mail karger@karger.com
www.karger.com/ned

Neuro
-epidemiology

KARGER

Monitoring Huntington's disease mortality across a 30-year period: Geographic and temporal patterns

Abstract

Background: Huntington's disease (HD) is a progressive neurodegenerative condition characterized by chorea, dystonia, behavioral disturbances and cognitive decline. The aim of this study is to assess temporal and spatial changes on mortality attributable to HD over 30 years in Spain.

Methods: HD data were extracted from the nationwide mortality registry for the period 1984–2013. Annual and 5-year gender- and age-specific rates adjusted for the standard European population were calculated. Geographic analysis was performed by districts from 1999 through 2013, and then estimated standardized mortality ratios (SMRs) and smoothed SMRs.

Results: There were 1,556 HD-related deaths across the study period. An increasing trend in age-adjusted HD mortality was in evidence, specifically from 1994 through 1998. On a year-by-year basis, age-adjusted mortality rates increased from 0.076 per 100,000 population in 1984 to 0.157 in 2013. Geographical differences among districts were evident in specific areas and in the southwest.

Keywords: Huntington's disease · Mortality · Spatial analysis · Spain · Population-based

4.1 Introduction

Huntington's disease (HD) is a chronically debilitating and neurodegenerative disease with an autosomal dominant pattern of inheritance. It typically appears in early adult life, has a progressive course, and reveals a combination of motor, cognitive and behavioral symptoms. There is progressive, selective neural cell loss and atrophy in the caudate and putamen. It is caused by a heterozygous expanded CAG trinucleotide repeats in HTT, the gene that encodes the huntingtin protein [1–3]. Diagnosis and genetic confirmation improved with the mapping of HD gene chromosome 4p in 1983 and the identification of the pathogenic mutation ten years later [4–6]. The current gold standard for the diagnosis is the DNA determination, showing a CAG-repeat of at least 36 on the huntingtin gene on chromosome 4p according to international guidelines [7, 8]. The fact that HD has an estimated worldwide prevalence of around 2.7 cases per 100,000 persons (95% CI 1.55–4.72), means that it meets the criteria for being classified as a rare disease (RD) [9]. While prevalence figures are higher in Caucasian populations (5.7 cases per 100,000), those in Asia and Africa are lower (approximately 0.5 cases per 100,000) [9–11].

Mortality studies provide insight into demographic changes, prompting socio-economic strategies addressed at improving population health and wellness [12, 13]. The robustness of nationwide population-based mortality registries make it possible to carry out analysis over extended periods of time. To enhance mortality studies, geographical information systems (GIS) are a useful tool for understanding spatial and epidemiologic patterns in a fast and effective manner. GIS capture, store, and process large amounts of data, thereby building a decision-making support system useful for disease surveillance and health planning [14]. Geographic analysis of national population-based HD mortality levels may be of great interest when it comes to planning the future of health care systems. The European Parliament provides much encouragement to analyze and enhance the epidemiological information of different RD in member states [15]. Related to this, an increasing number of studies combining geographical analysis with RD epidemiology have been conducted in recent decades. A temporal and spatial analysis of HD mortality was undertaken in Austria, and a regional analysis of prevalence was carried out in Finland [16, 17]. In Spain, our research group conducted a study on HD mortality trends across the period 1981–2004, and additional studies on the prevalence of the disease have also been published, but no spatial analysis has been reported to date [18–22].

Accordingly, the aim of this study was to analyze both the temporal trend and spatial variability in HD mortality in Spain over an extended period of time (1984–2013), in order to better characterize the patterns of mortality attributable to this RD.

4.2 Methods

Population-based data for the period 1984 through 2013 were sourced from the annual death registry kept by the National Statistics Institute (NSI) (*Instituto Nacional de Estadística*). Cases of HD were identified using the International Classification of Diseases, 9th Revision (ICD-9) code 333.4 for the period 1984–1998, and the 10th Revision (ICD-10) code G10 for the period from 1999 onwards. We calculated annual and 5-year period age- and gender-adjusted mortality rates using the standard European population as reference. Both annual age-adjusted rates and average age of death were assessed by Joinpoint analysis. Adjusted rates from 1984 to 2013 were smoothed, using the TH4253 non-parametric procedure to delete random movements in time-series data and improve sharpness [23].

For geographic analysis purposes, we obtained populations, categorized by gender and age at the municipal level, from the NSI. The study area comprised Spanish territory defined by a total of 326 districts. These districts are known in Spain as 'comarcas', divisions of the Spanish territory concerning adjacent municipalities with similar geographical and historical features [24]. Districts have been obtained by aggregating municipalities according to the information provided by the Agricultural Census furnished by the NSI. Standardized mortality ratios (SMRs) were calculated by district for the period 1999–2013 using indirect standardization, and then smoothed (smoothed-SMRs) for the purpose of reducing instability in numerical values in districts with a lower number of deaths, taking into account data from adjacent units. The conditional autoregressive model proposed by Besag et al. [25] considers the spatial contiguity and heterogeneity of each unit.

All statistical analyses were performed using the SPSS, Joinpoint and R-INLA computer software programs, with the ArcGIS software program being used for cartographic representations.

4.3 Results

There were 1,556 HD-related deaths in Spain from 1984 through 2013. Male mortality slightly exceeded female mortality, with men accounting for 51.6% of all such deaths (803 vs. 753 among women). Crude HD mortality rates tripled, going from 0.068 per 100,000 persons in 1984 to 0.182 in 2013. As shown in **figure 1**, the adjusted mortality rate in Spain increased significantly over time by 3.44% per annum ($p < 0.001$), with no point changes observed in this trend. On a yearly basis, adjusted rates varied from 0.076 in 1984 to 0.142 in 2013, reaching a peak in 2011 (0.179 per 100,000 persons).

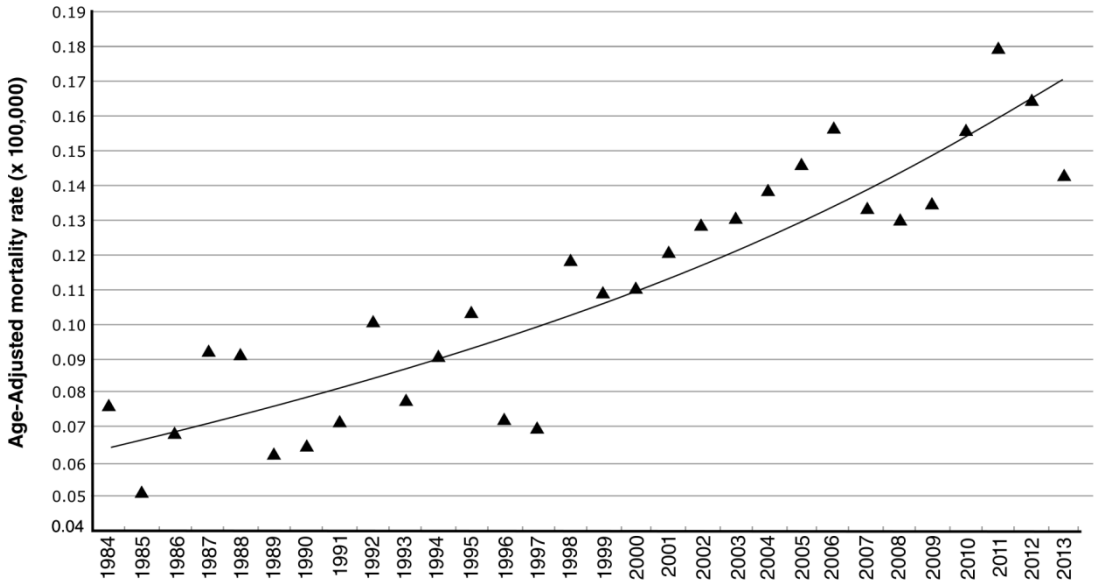


Figure 1. HD mortality trend plotted using a Joinpoint regression model: 1984–2013.

The average age of death increased by 0.59% per year ($p < 0.001$), increasing from 45 years in 1984 to 62 years in 2013. In the disaggregated analyses, the adjusted mortality rate was 25.2%, and was higher among men than among women (annual average). Across the study period, annual increases of 3.82% ($p<0.001$) and 3.31% ($p<0.001$) were observed among men and women, respectively. As can be seen in **figure 2a**, the annual smoothed line reveals a rising trend in mortality rates from the 1990s until 2010, when it stabilized. The apparent decrease among females since 2010 was statistically unremarkable, according to our analyses. In addition, higher rates were observed in the 5-year analysis, especially from 1994 to 1998 onwards. As is clear from **figure 2b**, the last 5-year period witnessed 38.8% of all HD deaths across the entire period, as well as the maximum adjusted rates when considering both genders (0.154 per 100,000) and men only (0.175).

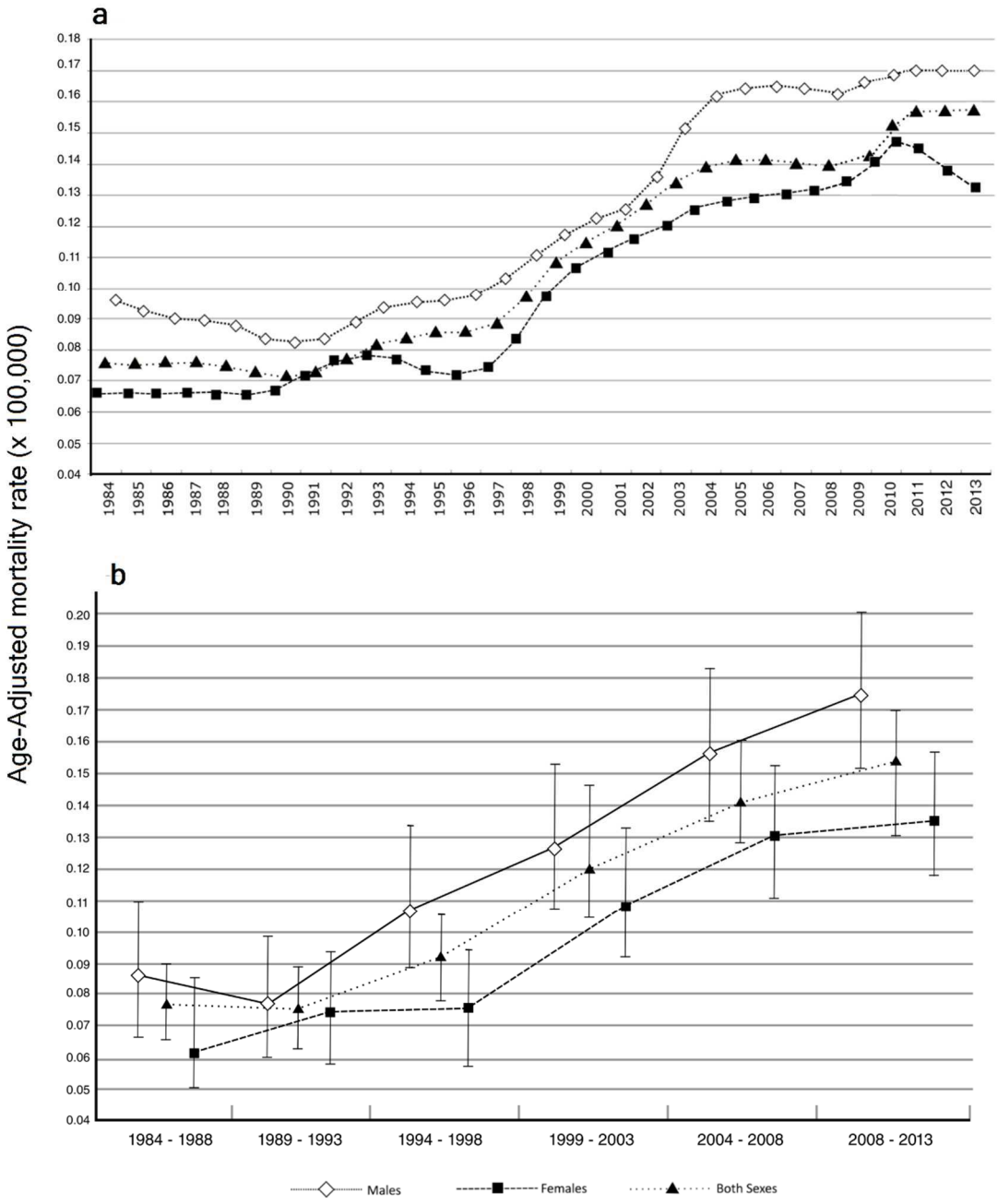


Figure 2. Age-adjusted HD mortality rate in Spain by gender: 1984–2013. **a** Annual smoothed; **b** 5-year period.

Insofar as the geographic distribution was concerned, significant differences were observed among districts. It is seen from **Table 1** that the HD mortality risk was significantly lower in scattered districts throughout the country, for example, usually under 0.60 in densely populated geographic areas (over 150,000 inhabitants). In contrast, a concentration of districts with significantly higher SMRs was observed in an area in southern Spain comprising the provinces of Malaga, Cadiz and Seville, with figures ranging from 0.2 up to 5.26. Results classified by gender displayed similarities to those observed in the aggregate analyses, though there were more districts with higher risk among women than among men. The Madrid metropolitan area was noteworthy, in that it registered the lowest mortality risk, namely, 0.56 (95% CI 0.43–0.72) for both genders, 0.52 (95% CI 0.35–0.76) for men, and 0.60 (95% CI 0.42–0.84) for women.

Table 1. Comparison of SMRs among districts (indirect standardization) across the period 1999–2013, for both genders, and for males and females separately. Only areas with significant results are shown

District	Province	Location	SMR 1999-2013 (95% CI)		
			Both genders	Males	Females
Very low risk					
Campo de Nijar	Almería	SE	0.00 (0.00 -0.73)	-	-
Vélez-Málaga	Málaga	S	0.00 (0.00 - 0.88)	-	-
Norte de Tenerife	Tenerife	SW*	0.11 (0.00 - 0.65)	0.00 (0.00 - 0.82)	-
Sur de Tenerife	Tenerife	SW*	0.19 (0.02 - 0.70)	-	-
La Campiña	Sevilla	SW	0.26 (0.03 - 0.93)	-	-
Septentrional	La Coruña	NW	0.54 (0.26 - 0.99)	-	-
Metro Area Madrid	Madrid	C	0.56 (0.43 - 0.72)	0.52 (0.35 - 0.76)	0.60 (0.42 - 0.84)
Gran Canaria	Las Palmas	SW*	-	0.32 (0.06 - 0.95)	-
High risk					
Mallorca	Baleares	E*	1.59 (1.06 - 2.28)	-	2.17 (1.30 - 3.38)
Vallès Occidental	Barcelona	NE	1.62 (1.09 - 2.31)	-	2.22 (1.36 - 3.43)
Litoral	Pontevedra	NW	1.66 (1.12 - 2.37)	2.02 (1.20 - 3.19)	-
Guipuzkoa	Guipuzkoa	N	1.68 (1.14 - 2.38)	-	1.75 (1.00 - 2.85)
Maresme	Barcelona	NE	1.98 (1.39 - 2.74)	1.92 (1.14 - 3.04)	2.04 (1.21 - 3.22)
Oviedo	Asturias	N	2.01 (1.23 - 3.11)	-	2.15 (1.07 - 3.84)
Campo de Gibraltar	Cádiz	SW	2.03 (1.01 - 3.63)	-	-
La Sagra-Toledo	Toledo	C	2.12 (1.02 - 3.91)	-	-
Almendralejo	Badajoz	SW	2.71 (1.09 - 5.59)	-	-
Badajoz	Badajoz	SW	2.75 (1.32 - 5.07)	-	-
Campiña de Cádiz	Cádiz	SW	2.84 (1.76 - 4.35)	3.16 (1.63 - 5.53)	2.51 (1.14 - 4.76)
Vinalopó	Alicante	E	3.15 (1.92 - 4.87)	2.74 (1.25 - 5.21)	3.57 (1.78 - 6.40)
Sierra Norte	Sevilla	SW	3.51 (1.13 - 8.19)	5.40 (1.45 - 13.82)	-
Serranía de Ronda	Málaga	SW	3.59 (1.16 - 8.39)	5.56 (1.50 - 14.20)	-
La Janda	Cádiz	SW	3.74 (1.61 - 7.36)	-	4.00 (1.07 - 10.26)
Lozoya-Somosierra	Madrid	C	3.81 (1.23 - 8.89)	-	5.04 (1.01 - 14.73)
Nord-Occidental	Navarra	N	4.06 (1.48 - 8.84)	4.88 (1.31 - 12.50)	-
Esla-Campos	León	N	4.65 (1.25 - 11.9)	-	-
Sierra Sur	Sevilla	SW	5.26 (2.27 - 10.38)	5.06 (1.36 - 12.96)	5.47 (1.47 - 14.00)
Cuenca del Jiloca	Teruel	NE	6.73 (1.81 - 17.22)	-	-
Bajo Llobregat	Barcelona	NE	-	-	1.61 (1.21 - 2.10)
Campos	Palencia	C	-	-	3.92 (1.43 - 8.53)
La Costera	Valencia	E	-	-	1.18 - 11.31)

* Island territories (Canary and Balearic Islands). C= Centre; E= East; N=North; NE= Northeast; NW=Northwest; S=South; SE=Southeast; SW=Southwest; W=West.

There was considerable geographical variability in the smoothed-SMRs (**fig. 3a**). As can be seen from the values registered for both genders, the smoothed-SMR yielded by the aggregate analysis stood in sharp relief, with a value of over 1.5 in some areas, such as the Basque Country and other districts located in the southwest of Spain. These results were similar for men (**fig. 3b**) and women alike (**fig. 3c**), with slight differences vis-à-vis the aggregate map. Nevertheless, there was insufficient evidence to indicate any spatial pattern of HD mortality across the country as a whole.

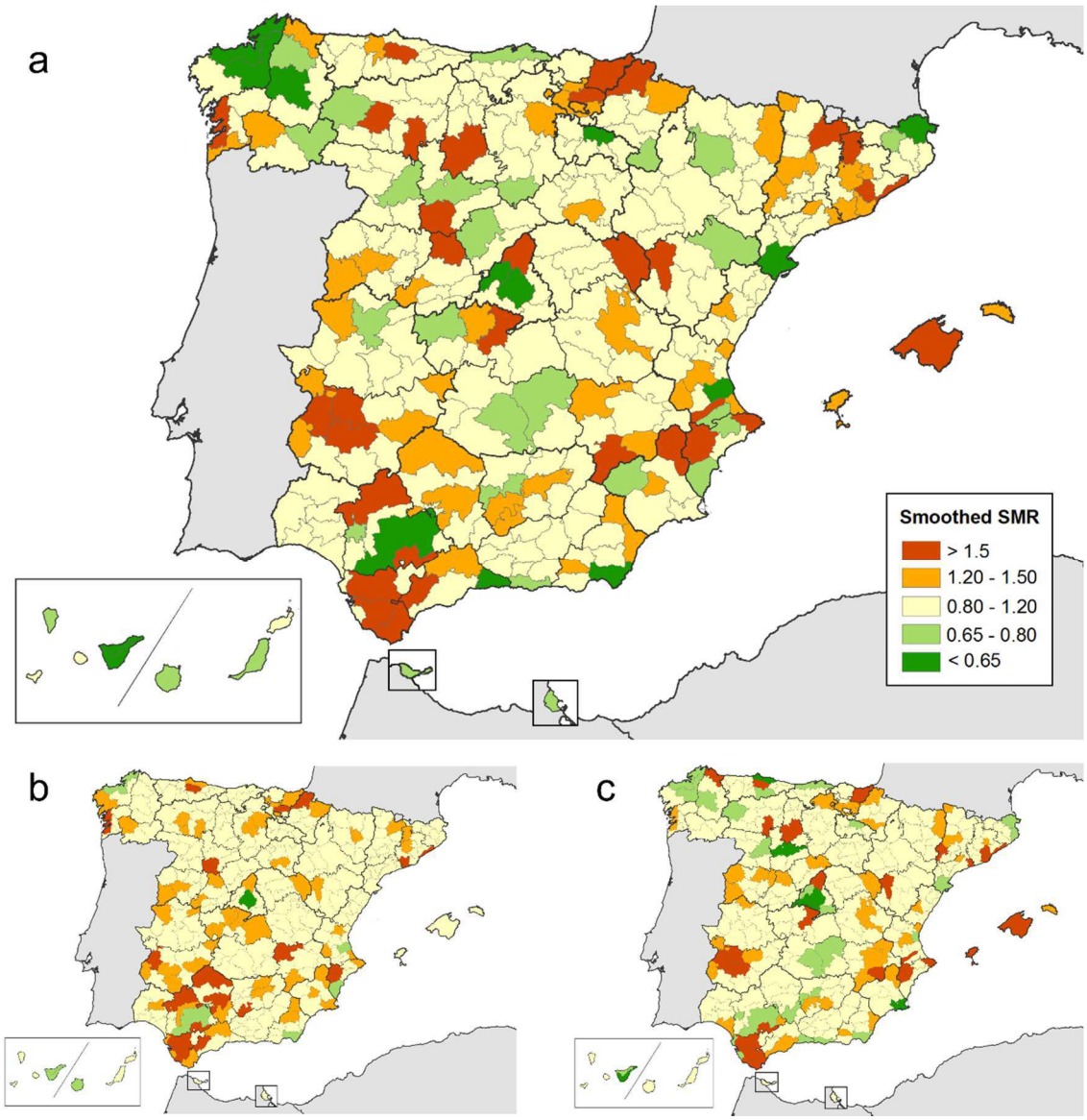


Figure 3. Smoothed-SMRs for HD by gender: 1999–2013. **a** Both genders; **b** males; **c** females

Figure 4a depicts posterior probability (PP) by geographic location. While districts with values of over 0.80 had a significantly higher-than-expected mortality risk at a national level, values under 0.20 indicated a significantly lower risk. As can be seen from **figure 4b**, the southwestern district of Campiña de Cádiz (PP = 0.995) and the Madrid metropolitan area (PP < 0.001) represent the extremes of geographical variability in males, in terms of expected HD mortality risk here in Spain. At a national level, the distribution of values was more to be dispersed in females than in males (**fig. 4c**).

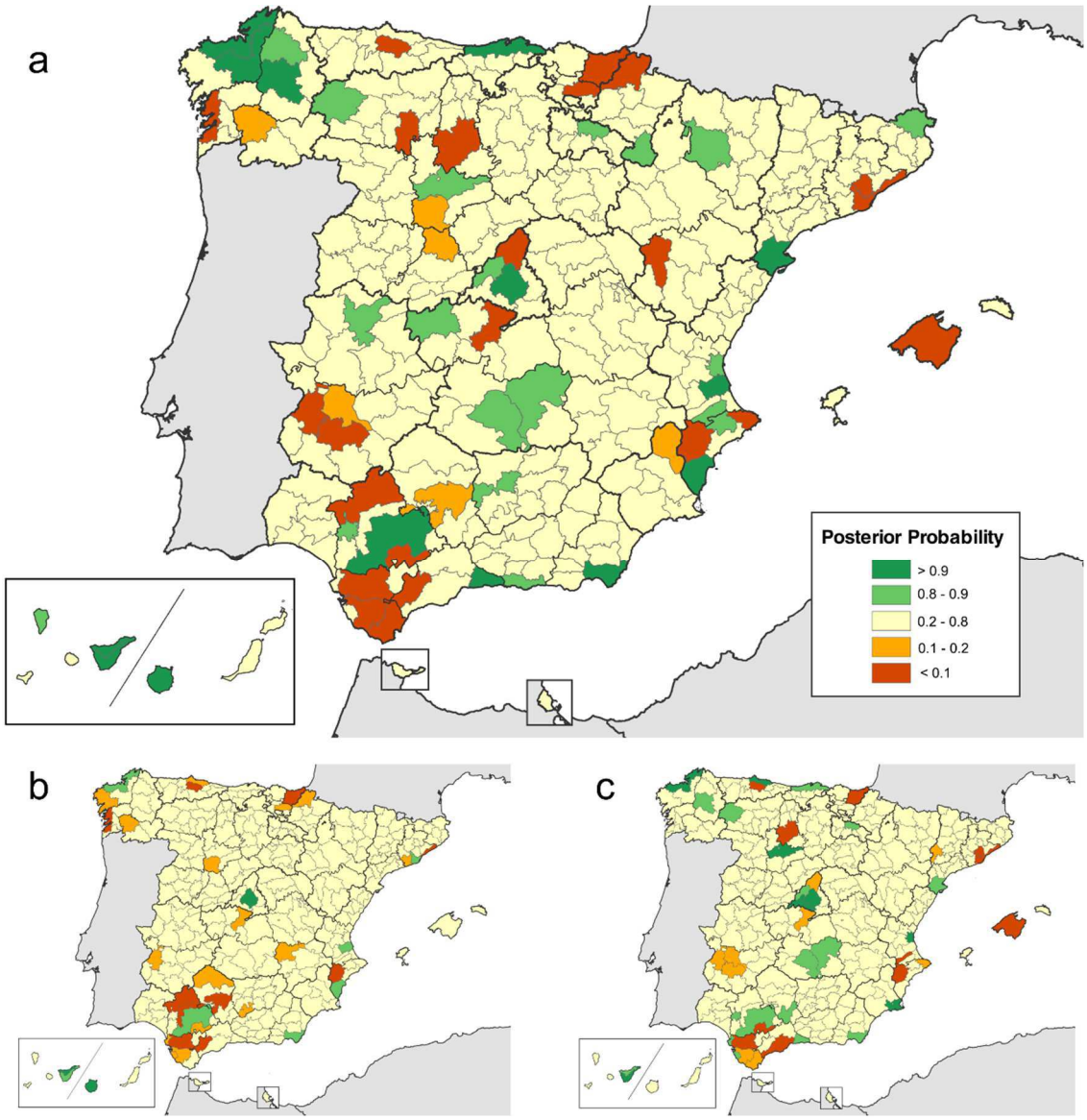


Figure 4. PP of smoothed-SMR due to HD (significantly higher (>0.80) or significantly lower (<0.20)) by district: 1999–2013. **a** Both genders; **b** males; **c** females

4.4 Discussion

This study shows the rise in HD-related mortality in the last 3 decades and its geographical distribution at a district level. These results have been achieved, thanks to one of the largest detailed-scale mapping exercises conducted at a national level in assessing a low-prevalence disease.

Adjusted rates are higher in men than in women. The excess male HD mortality reported by this study was within the expected range, according to data reported by other studies undertaken in various European countries and the United States [14, 26]. The time trend showing an annual 3.44% increase in HD mortality is in line with the annual change described in a previous study undertaken in Spain [18]. The increase observed in the mortality rates over the period might reflect a rise of prevalence, probably due to better health care of severe complications, which seems to be supported by the gradual increase in the average age of death. While the period covered by us coincided in part with that of a previous study, our study not only extended and updated the time trend by 9 years, but also contributed a spatial analysis. Another study which analyzed HD mortality in Spain is not comparable to ours, in that it failed to furnish adjusted values and was conducted at an earlier point in time (1974–1979) [27]. In all these studies, however, the evolution in the number of deaths would appear to be in line with the trend observed by us.

In the absence of other time-series studies on HD mortality in Spain, comparison with the scientific literature on prevalence studies conducted in specific locations is required. Some examples included the city of Cadiz in 1970 (1.4 cases per 100,000 inhabitants), the province of Salamanca in 1984 (8.4 cases per 100,000), the Valencian region in the period 1987–1992 (5.4 cases per 100,000), and more recently, Navarre (5 cases per 100,000) [19–22]. When chronologically arranged, these studies do not indicate a rising trend, and the latest data to be reported are close to the reported Caucasian average [9]. At the European level, a British journal pointed to a wide variation in the prevalence rate, ranging from 5.4 cases per 100,000 inhabitants in 1990 to 12.3 in 2010 [28]. This can be explained – at least partly – by the greater availability of predictive genetic tests, particularly in newly diagnosed patients with no family history [29, 30]. The British results may signal a similar trend in other European countries, if national studies were carried out. A family history with clinical and morphological verification in at least one of the parents or grandparents was obligatory to diagnose HD before 1993 [31]. After that, clinical diagnosis is complemented with DNA determination as the gold standard in Spain. The progressive implementation of genetic confirmation available since 1993 does not seem to modify the continuing rising trend in Spanish mortality rates after that, according to Joinpoint results. Other authors suggest that increased mortality rates could be related to the change in ICD

coding since 1999, but this reason alone cannot account for this pattern because in 1998–1999, mortality rates rose in the same way as they did during the remainder of the study period.

In terms of geographic distribution, different Spanish districts show a higher HD mortality risk, mainly in the southwestern provinces of Seville, Cadiz and Malaga (particularly among males). Despite these results, at the level studied, there is not enough evidence to find any spatial patterns of mortality due to HD in Spain. Areas with a higher incidence of HD, such as Venezuela or Tasmania, have been identified due to a founder effect, but no evidence has yet been detected in Spain, not even in the island territories [32–34].

The introduction of predictive genetic testing in the Spanish National Health System has made for earlier, accurate diagnosis [5]. Notwithstanding this, the scant evidence to support the claim that pharmacologic and nonpharmacologic treatments to control some symptoms can prevent the disease from progressing has generated resentment among persons with a family history of HD to undergo the predictive test [35–37]. In addition, no indications have been found to explain the increase in mortality rates during the study period, though this might be related to a likely increased prevalence in Spain, in line with other Western countries. In the coming years, the influence on diagnosed patients of HD predictive genetic testing in family planning could lead to a reduction in birth rates and, by extension, in mortality rates [38]. On the other hand, the fatal outcome of HD causes a four- to eightfold increase in the risk of suicide, as shown by studies in the United States [39, 40]. The fact is that the risk of suicide among diagnosed HD patients is associated with depressive behaviors. In some cases, it is accompanied by aggressiveness and anxiety as a consequence of a high burden of disease not only in patients but also in relatives and caregivers [41]. Therefore, support services for both patient and family are vital [42–44].

It is important to note that HD mortality data could be underestimated in cases where the underlying cause of death is not duly completed on the death certificate. Since the underlying cause of death was used to identify HD cases in this study, the rates obtained could thus be lower than expected. Lastly, positive predictive value was assessed for HD on the basis of medical records and primary-care databases, but the mortality database could not be validated for HD [12, 22]. Even so, the strength of this study is supported by the nationwide, population-based death registry and a constant (over time) and standardized, geographically homogeneous (in all regions) methodology.

In conclusion, this epidemiologic study has shown an increase in the HD mortality rate and a somewhat heterogeneous regional distribution in Spain. At this point in time, there is no known evidence to link this rising trend and geographical distribution with environmental or other types of factors. Future studies should focus on analyzing the risk in regions with

higher mortality rates in order to improve and optimize health care planning by correctly allocating the necessary resources. Furthermore, the identification of some areas with lower rates could help when it comes to checking whether correct diagnoses are being made in all cases, and implementing any remedial measures, if needed – something that would also have an impact on health care systems.

4.5 References

1. Bates GP, Dorsey R, Gusella James F, et al. Huntington disease. Nature Reviews Disease Primers 2015.

2. Armstrong MJ, Miyasaki JM. Evidence-based guideline: Pharmacologic treatment of chorea in Huntington disease. Neurology 2012;79:597-603.

3. Walker FO. Huntington's disease. Seminars in Neurology 2007;27:143-150.

4. Gusella JF, Wexler NS, Conneally PM, et al: A polymorphic DNA marker genetically linked to Huntington's disease. Nature 1983; 306: 234–238.

5. Falush D, Almqvist EW, Brinkmann RR, et al: Measurement of mutational flow implies both a high new-mutation rate for Huntington disease and substantial underascertainment of late-onset cases. Am J Hum Genet 2001; 68: 373–385.

6. Bates G. The molecular genetics of Huntington disease - a history. Nature Reviews Genetics 2005; 6:766-773

7. Benitez J, Fernández E, García Ruiz P, Robledo M, Ramos C, Yébenes J. Trinucleotide (CAG) repeat expansion in chromosomes of Spanish patients with Huntington's disease. Hum Genet 1994; 94(5): 563-564.

8. Losekoot M, Van Belzen MJ, Seneca S, Bauer P, Stenhouse SA, Barton DE, European Molecular Genetic Quality Network (EMQN). EMQN/CMGS best practice guidelines for the molecular genetic testing of Huntington disease. European Journal of Human Genetics 2013; 5:480-486.

9. Pringsheim T, Wiltshire K, Day L, et al. The Incidence and Prevalence of Huntington's Disease: A Systematic Review and Meta-analysis. Movement Disorders 2012;27:1083-1091.

10. Nakasima K, Watanabe Y, Kusumi, M, et al. Epidemiological and Genetic Studies of Huntington's Disease in the San-in Area of Japan. Neuroepidemiology 1996;15: 126-131.

11. Chen Y, Lai CH. Nationwide Population-Based Epidemiologic Study of Huntington's Disease in Taiwan. Neuroepidemiology 2010;35:250-254.

12. United Nations. Data bases for mortality measurement. United Nations Publication 1984, Sales No. E.83.XIII.

13. Joubert J, Rao C, Bradshaw D, et al. Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa. Global Health Action 2012;5:1-19.

14. Ascuntar-Tello J, Jaimes F. Ronda clínica y epidemiológica: Sistemas de Información Geográfica (SIG) en salud. Iatreia 2016;29:97-103.

15. European Union: European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committeee and the Committee of the Regions on Rare Diseases: Europe's Challenges, 11 November 2008, COM (2008) 679 final, available at: http://ec.europa.eu/health/ph_threats/non_com/docs/rare_com_en.pdf

16. Ekestern E, Lebhart G. Long-term monitoring of the mortality trend of Huntington's disease in Austria. Eur J Epidemiol 2005;20:169-172.

17. Sipilä JO, Hietala M, Siitonen A, et al. Epidemiology of Huntington's disease in Finland. Parkinsonism Relat Disord 2015;21:46-49.

18. Ramalle-Gomara E, Gonzalez MA, Perucha M, et al. Mortalidad por la enfermedad de Huntington en España en el período 1981-2004. Revista de Neurología 2007;45:88-90.

19. Calcedo Ordoñez A. Huntington's disease in the Province of Cadiz. Epidemiological and familial study. Revista Clínica Española 1970;119:333-344.

20. Ruiz JJ, Ortin A, Cacho J, et al. Corea de Huntington: estudio epidemiológico en la provincia de Salamanca. Aarch Neurobiol 1985;48:302-303.

21. Burguera JA, Solís P, Salazar A. Estimate of the prevalence of Huntington disease in the Valencia region using the capture-recapture. Revista de Neurología 1997;25:1845-1847.

22. Vicente E, García-Amigot F, Gastón MI, et al. Prevalence on Huntington Disease in Navarra (Spain). Sensitivity and positive predictive value of different sources of ascertainment. J Neurol Neurosurg Psychiatry 2014;85:suppl 1.

23. Velleman PF. Definition and comparison of robust nonlinear data smoothing algorithms. Journal of the American Statistical 1980;75:609-615.

24. Arias Lopez M. La comarca como entidad natural, histórica y administrativa. Dykinson, S.L. 2013. ISBN: 978-849031758

25. Besag J, York J, Mollìe A. A Bayesian image restoration with two applications in spatial statistics. Ann Inst Math 1991;43:1-59.

26. Lanska DJ, Lavine L, Lanska MJ, et al. Huntington disease mortality in United States. Neurology 1988;38:769-772.

27. Burguera Hernández JA, Láinez Andrés JM, Blasco Olcina R, et al. Mortalidad por enfermedad de Huntington en España, 1975-1979. Revista de Neurología 1990;18:13-16.

28. Evans SJ, Douglas I, Rawlins MD, et al. Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. Journal of Neurology, Neurosurgery, and Psychiatry 2013;84:1156–1160.

29. Sipilä JOT, Päivärinta M. Why we still need more research on the epidemiology of Huntington's disease. Neuroepidemiology 2016;46:154-155.

30. Rawlins MD, Wexler NS, Wexler AR, et al. The prevalence of Huntington's disease. Neuroepidemiology 2016;46:144-153.

31. Roos R. Huntington's disease: a clinical review. Orphanet J Rare Dis 2010; 5-40.

32. Young AB, Shoulson I, Penney JB, et al. Huntington's disease in Venezuela: Neurologic features and functional decline. Neurology 1986;36:244-249.

33. Paradisi I, Hernández A, Arias S. Huntington disease mutation in Venezuela: age of onset, haplotype analyses and geographic aggregation. J Hum Genet 2008;53:127-135.

34. Pridmore SA. The large Huntington's disease family of Tasmania. Med J Aust 1990;153:593-595.

35. Rivera-Navarro J, Cubo E, Mariscal N. Analysis of the Reasons for Non-Uptake of Predictive Testing for Huntington's Disease in Spain: A Qualitative Study. J Genet Couns 2015;24:1011-1021.

36. Ramos-Arroyo M, Moreno S, Valiente A. Incidence and mutation rates of Huntington's disease in Spain: experience of 9 years of direct genetic testing. Journal of Neurology, Neurosurgery, and Psychiatry 2005;76:337–342.

37. Yudkin JS, Ritcher B, Gale E. Uncovering the true prevalence of Huntington's disease. Lancet 2010;376:18-19.

38. Solis-Pérez MP, Burguera JA, Palu F, et al. Resultados de un programa de diagnóstico presintomático en la enfermedad de Huntington: evaluación de un periodo de 6 años. Revista de Neurología 2001;16:348-352.

39. Farrer LA, Opitz JM, Reynolds JF. Suicide and attempted suicide in Huntington disease: Implications for preclinical testing of persons at risk. Am J Med Genet 1986;24:305-311.

40. Shoenfeld M, Myers RH, Cupples A, et al. Increase rate of suicide among patients with Huntington's disease. Journal of Neurology, Neurosurgery, and Psychiatry 1984;47:1283-1287.

41. Dorey J, Trigo P, Garcia de Yébenes, Tourmi M. Burden of Huntington's Disease in Spain. Journal of Neurology, Neurosurgery and Psychiatry 2012; 303524.51

42. Hubers AA, Van Duijn E, Roos RA, et al. Suicidal ideation in a European Huntington's disease population. J Affect Disord 2013;151:248-258.

43. Bindler L, Travers D, Millet B. Suicide in Huntington's disease: a review. Red Med Suisse 2009;5:646-648.

44. Fernández M, Grau C, Trigo P. The impact of Huntington's disease on the family. An Sist Sanit Navar 2012;35:295-307.

Chapter 5

Temporal and cartographic analyses of the distribution within Spain of mortality due to Granulomatosis with polyangiitis (1984–2016)

Temporal and cartographic analyses of the distribution within Spain of mortality due to granulomatosis with polyangiitis (1984–2016)

Published in: International Journal of Environmental Research and Public Health

IJERPH 2019, 16, 1388; doi:10.3390/ijerph16081388

Authors: Germán Sánchez-Díaz [1,2,3], Francisco Escobar[2], Ana Villaverde-Hueso [1,3], Manuel Posada de la Paz [1,3] and Verónica Alonso-Ferreira [1,3].

[1] Institute of Rare Diseases Research (IIER), Instituto de Salud Carlos III, 28029 Madrid, Spain

[2] Department of Geology, Geography and Environmental Sciences, University of Alcalá 28801 Alcalá de Henares, Spain

[3] Centre for Biomedical Network Research on Rare Diseases (CIBERER), 28029 Madrid, Spain

International Journal of Environmental Research and Public Health

Article

Temporal and Cartographic Analyses of the Distribution within Spain of Mortality Due to Granulomatosis with Polyangiitis (1984–2016)

Germán Sánchez-Díaz [1,2,3], Francisco Escobar [2], Ana Villaverde-Hueso [1,3], Manuel Posada de la Paz [1,3] and Verónica Alonso-Ferreira [1,3,*]

[1] Institute of Rare Diseases Research (IIER), Instituto de Salud Carlos III, 28029 Madrid, Spain; g.sanchez@externos.isciii.es (G.S.-D.); anavillaverde@isciii.es (A.V.-H.); mposada@isciii.es (M.P.d.l.P.)
[2] Department of Geology, Geography and Environmental Sciences, University of Alcala, 28801 Alcalá de Henares, Spain; francisco.escobar@uah.es
[3] Centre for Biomedical Network Research on Rare Diseases (CI
* Correspondence: valonso@isciii.es; Tel.: +34-91-822-20-89

Received: 14 March 2019; Accepted: 13 April 2019; Published: 17

Abstract: The aim is to conduct a descriptive, population- and spatial changes in mortality due to granulomatosis wi to 2016. Mortality data were obtained from the Spanish GPA was the underlying cause were selected using the 446 Classification of Diseases, 9th and 10th revision. Annual mortality rates were calculated. Geographic analysis was level. Variations in mortality according to the type of m district and geographic location (degrees of latitude) wer ratios (SMRs) and smoothed-SMRs. Over the whole period Age at death increased at an average annual rate of 0.78% adjusted mortality rates increased by an annual average they fell by 1.91% a year ($p < 0.05$). The agro-urban catego municipalities with a significantly higher GPA mortality revealed four districts with a higher risk of death due to G the South. This population-based study revealed an increa Age-adjusted mortality rates went up sharply until 1992, the end of the study period. Geographic differences in m studies will be necessary to ascertain the reasons for the di

Keywords: mortality; Spain; granulomatosis with poly trends; latitude; mapping; Wegener

1. Introduction

Granulomatosis with polyangiitis (GPA), formerly kn systemic necrotizing vasculitis characterized primarily by and medium vessels [1]. It typically affects the respiratory death if not treated [2]. It is considered a rare disease due t per 10,000 inhabitants) and its etiology remains unknown [

Mortality due to GPA is still high, especially in the treatments being available [6]. Although the death rate amo the beginning of the 21st century, it remains higher than th

Int. J. Environ. Res. Public Health **2019**, *16*, 1388; doi:10.3390/ijerph16081388

Temporal and cartographic analyses of the distribution within Spain of mortality due to granulomatosis with polyangiitis (1984–2016)

Abstract

The aim is to conduct a descriptive, population-based study in order to assess temporal and spatial changes in mortality due to granulomatosis with polyangiitis (GPA) in Spain from 1984 to 2016. Mortality data were obtained from the Spanish Annual Death Registry. Deaths in which GPA was the underlying cause were selected using the 446.4 and M31.3 codes from the International Classification of Diseases, 9th and 10th revision. Annual average age at death and age-adjusted mortality rates were calculated. Geographic analysis was performed at municipality and district level. Variations in mortality according to the type of municipality (urban, agro-urban or rural), district and geographic location (degrees of latitude) were assessed using standardized mortality ratios (SMRs) and smoothed-SMRs. Over the whole period, 620 deaths due to GPA were identified. Age at death increased at an average annual rate of 0.78% over the period 1987–2016 ($p < 0.05$). Age-adjusted mortality rates increased by an annual average of 20.58% from 1984 to 1992, after which they fell by 1.91% a year ($p < 0.05$). The agro-urban category had the highest percentage (4.57%) of municipalities with a significantly higher GPA mortality rate than expected. Geographic analysis revealed four districts with a higher risk of death due to GPA, two in the North of Spain and two in the South. This population-based study revealed an increase in the age at death attributed to GPA. Age-adjusted mortality rates went up sharply until 1992, after which they started to decline until the end of the study period. Geographic differences in mortality risk were identified but further studies will be necessary to ascertain the reasons for the distribution of GPA disease.

Keywords: mortality; Spain; granulomatosis with polyangiitis; geographic patterns; temporal trends; latitude; mapping; Wegener

5.1 Introduction

Granulomatosis with polyangiitis (GPA), formerly known as Wegener's Granulomatosis, is a systemic necrotizing vasculitis characterized primarily by a granulomatous inflammation of small and medium vessels [1]. It typically affects the respiratory tract and kidneys leading to premature death if not treated [2]. It is considered a rare disease due to its low prevalence (less than five cases per 10,000 inhabitants) and its etiology remains unknown [3,4,5].

Mortality due to GPA is still high, especially in the first year after diagnosis, despite better treatments being available [6]. Although the death rate amongst GPA patients has been reduced since the beginning of the 21st century, it remains higher than the mortality in the general population [7]. Age at death is increasing amongst patients suffering from this rare disease due to improvement in treatments based on cyclophosphamide, glucocorticoids and rituximab [8,9].

The incidence of GPA at an international level has been increasing since the 1980s and then stabilized in the early 2000s, suggesting an increasing physician awareness following the introduction of routine anti-neutrophil cytoplasmic antibodies (ANCAs) testing [10]. Furthermore, there are considerable variations between the incidence rates in different countries [2]. In Europe, notwithstanding the fact that there are variations in methodology in the different studies, data suggest that the incidence of this disease is two to four times greater in northern than in southern countries. Previous researchers reported an annual incidence rate of 9.8 (95% CI = 7.4–12.2) per million people in Sweden, 9.3 (8.1–10.6) in Finland, 11.8 (10.7–12.9) in United Kingdom, and 7.7 (4.1–11.4) in Poland, indicating a possible spatial variation in incidence rates [11–14]. In Spain, epidemiological information about GPA is still scarce and there are no studies of mortality or other indicators at a national level. The little research so far conducted on this subject has focused on incidence rates at a provincial or local level. These studies found that Lugo province had an annual rate of 2.95 cases per million inhabitants, the city of Sabadell in Barcelona province had 4.1 cases, while the rate in Málaga province was much lower at 2.1 cases per million [15–17].

Geographic analyses of GPA in which populations are stratified into rural and urban areas have been conducted in several European regions and in parts of Australia [11,17–19]. These investigations sought to assess the differences between the incidence or prevalence of GPA in rural and urban areas. In the case of Lugo province (Spain), for example, there was higher incidence in urban areas. Several studies have also analyzed GPA by comparing incidence rates in Europe according to latitude. These found higher rates in northern than in southern European countries [20]. At a Spanish level, we thought it might be interesting

therefore to observe the variation in mortality rates according to latitude, based on the assumption that northern regions will have higher rates than southern ones.

In view of the paucity of national population-based data, this study also seeks to analyze mortality due to GPA in Spain over a long period. The objectives were to: (1) evaluate the time trends in GPA mortality over a 32-year period; and (2) analyze the geographical distribution of mortality at municipal and district level, in rural vs. urban areas and according to latitude.

5.2 Materials and methods

This research focuses on the trend over time and the geographic distribution of GPA mortality for the period 1984–2016. Mortality data were obtained from the Spanish Annual Death Registry kept by the Spanish National Statistics Institute (NSI) (Instituto Nacional de Estadística). GPA as underlying cause of each death is noted by the codes 446.4 in the International Classification of Diseases (ICD) 9th edition (period 1984 and 1998), and M31.3 in the ICD 10th edition (period 1999–2016). Underlying cause of death is defined by the World Health Organization as "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury" [21]. Population data were also obtained from the NSI and categorized by gender and age at municipal and district or comarca level (in Spanish the word "comarca" refers to an area or territory that contains several towns and villages but is smaller than a region).

The age-adjusted mortality rates were calculated, using the European standard population as reference, for males, females, and both sexes (expressed per 1,000,000 inhabitants). Annual trends were smoothing using the TH4253 non-parametric procedure which uses running medians to summarize overlapping segments in the time-series data leading to delete random movements [22]. Average age at death and mortality trends were evaluated using joinpoint regression models to describe changes in the time-series data where line segments are joined in the joinpoints [23].

In order to assess the variation in mortality in rural and urban areas, the Spanish municipalities were broken down into three categories. Normally, municipalities are classified as rural or urban according to population criteria [17,19]. However, this classification seems imprecise due to the varying characteristics of municipalities in Spain. Northern municipalities are usually smaller in terms of population but have more industry, while Southern municipalities tend to have larger populations and are economically more oriented towards the agricultural sector. We therefore decided to include an intermediate

category for those municipalities with both urban and rural characteristics using not only population data but also the number of People Working in the Primary Sector (PWPS), which includes activities such as agriculture, livestock, mining or fishing. The data about the sectors in which people are employed were provided by the NSI. On this basis, in this study we used the following three categories:

(a) Urban: less than 5% PWPS

(b) Agro-urban: more than 5% PWPS and more than 10,000 inhabitants

(c) Rural: more than 5% PWPS and less than 10,000 inhabitants or labor force of less than 50 people.

To assess the variation in mortality according to latitude, Spain was divided into horizontal strips of one degree of latitude. The most northerly strip was at 44° and the most southerly at 36°. The municipalities in the Canary Islands were grouped together under the same latitude strip (from 27° N to 29° N).

For disease mapping purposes, standardized mortality ratios (SMRs) were calculated by municipality using the expected GPA mortality rate in Spain for the period 1999–2016 as a reference (SMR = 1.00). Municipalities were previously divided into three categories as described above: urban, agro-urban and rural areas. Similarly, SMRs were also calculated for groups of municipalities with the same latitude in an attempt to uncover possible North-South variations in mortality.

The smoothing procedure for geographic units was carried out at comarca (district) level. The smoothed-SMRs for the period 1999–2016 were calculated for district using the cases observed in each district together with the expected cases, and taking the sex- and age-specific death rates for the Spanish population as a reference. We used a conditional autoregressive model based on two effects: unstructured spatial random effect and spatially structured heterogeneity [24]. As a measure of uncertainty, the Posterior Probability (PP) of each district being above average risk was computed. PP shows those districts with significantly higher (PP > 0.80) or lower (PP < 0.20) risks of death due to GPA than that expected for the country as a whole.

The statistical analyses and the smoothing procedure were performed using the Stata and R software respectively. The ArcGIS software program (Esri, Redlands, CA, USA) was used for geographic analysis and cartographic representations.

5.3 Results

Overall, we identified 620 deaths from GPA over the period 1984–2016 in Spain, of which 55.5% were males (**Figure 1**). The average age at death for GPA patients was 65.6 years ± 14.6 Standard Deviation (STD); 65.2 years ± 14.6 STD in males; and 65.8 years ± 14.8 STD in females. GPA age at death increased by 0.78% a year (annual percent change, APC) between 1987 and 2016 ($p < 0.05$). By sexes, age at death increased annually by 0.61% in males and 0.74% in females over the whole study period (1984–2016) ($p < 0.05$).

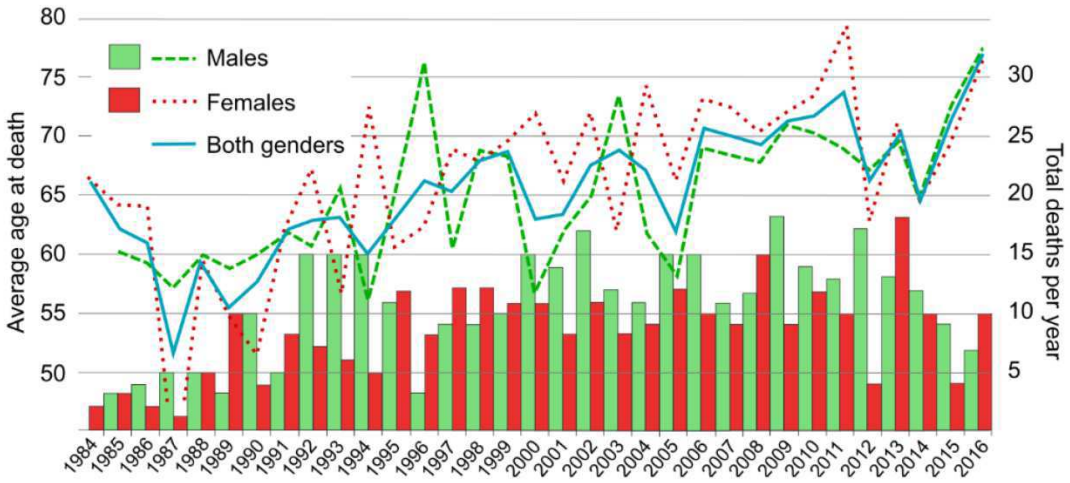


Figure 1. Average age at death (lines) and number of deaths (bars) due to granulomatosis with polyangiitis (GPA) in Spain by gender: 1984–2016.

The overall age-adjusted mortality rate for the period was 0.35 (95% Confidence Interval (CI) = 0.32–0.38) per million inhabitants: 0.44 (95% CI = 0.39–0.49) in males, and 0.28 (95% CI = 0.24–0.32) in females. Joinpoint analysis showed a 20.58% increase in the age-adjusted mortality rate between 1984 and 1992 ($p < 0.001$) and an annual fall of 1.91% from 1992 to 2016 ($p < 0.001$). Rates were higher for males than for females, and there was a significant increase of 21.57% APC ($p < 0.001$) over the period 1984–1992. From then on, the rate declined (–2.13% APC; $p < 0.001$). In females, there was a significant increase in the death rate of 36.1% APC ($p < 0.001$) from 1984 to 1989 after which it remained stable over the rest of the period. These trends for both sexes (males and females) are similar to the smoothed age-adjusted mortality rates and in the aggregation into intervals (**Figure 2**).

According to our classification of Spanish municipalities as urban, agro-urban or rural, 16.2% were considered as urban, 2.7% as agro-urban and 81.1% as rural (**Table 1**). Deaths due to GPA took place in 271 municipalities (3.34% of the total of Spanish municipalities).

According to SMR by municipality, 4.57% of agro-urban municipalities had a significantly higher GPA mortality than expected for Spain as a whole, while this percentage was lower in urban and rural municipalities. Only one urban municipality obtained a death risk lower than expected (Barcelona: SMR = 0.52; 95% CI = 0.24–0.98). However, global SMR in these categories was similar to the expected value (SMR = 1.00): SMRs due to GPA in all agro-urban areas considered as a whole were 1.25 (95% CI = 0.93–1.64), 0.87 (95% CI = 0.68–1.10) in rural and 1.00 (95% CI = 0.89–1.12) in urban.

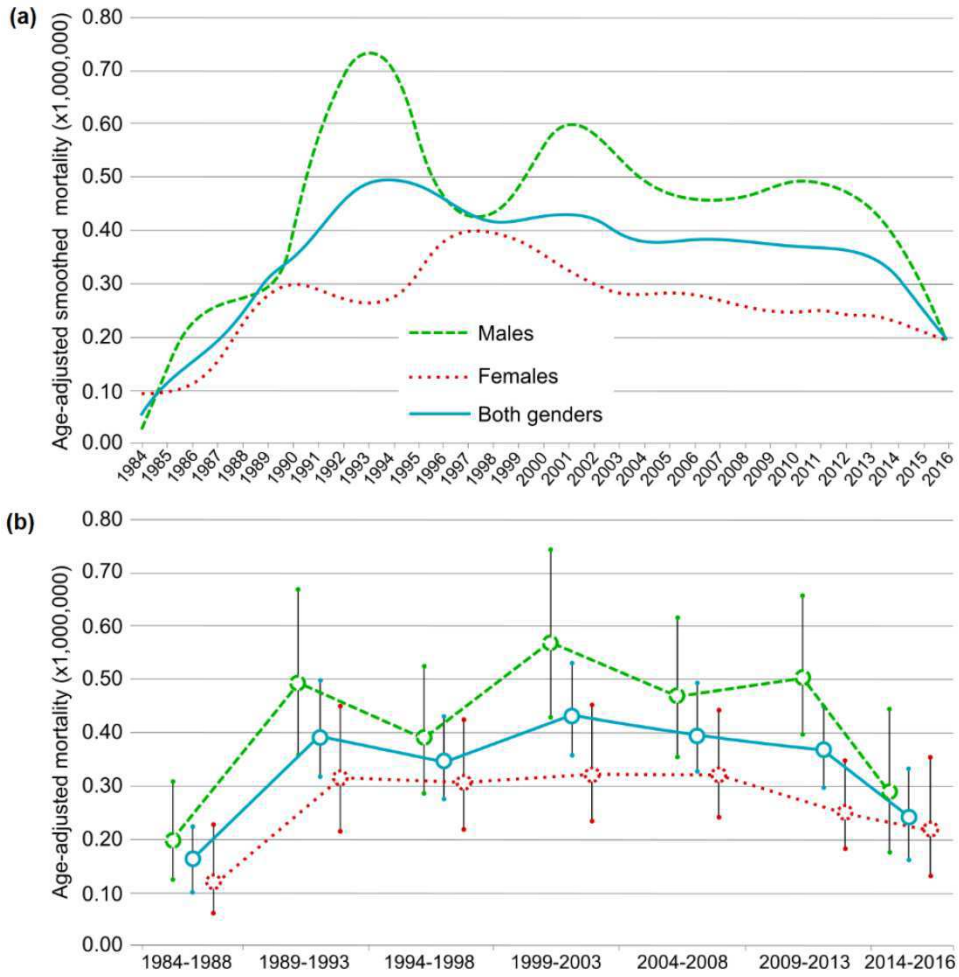


Figure 2. Age-adjusted mortality rates due to GPA in males, females and both genders. (a) Annual smoothed rates; (b) five-year rates (except for 2014 to 2016), bars show 95% confidence intervals.

Table 1. Distribution of municipalities categorized as urban, agro-urban or rural; number of deaths due to GPA and municipalities with higher or lower GPA mortality than expected for Spain in 1999–2016 (only statistically significant differences are shown, $p < 0.05$).

Category	Municipalities (%)	Municipalities Registering GPA Deaths	Deaths Due to GPA	Municipalities (%) with Significant SMR Due to GPA	
				Higher Than Expected	Lower Than Expected
Urban	1316 (16.2%)	162	300	22 (1.67%)	1 (0.08%)
Agro-urban	219 (2.7%)	40	52	10 (4.57%)	0
Rural	6588 (81.1%)	69	71	25 (0.38%)	0

Throughout the Spanish territory, while urban municipalities are frequent along the Mediterranean (Catalonia, Valencia) and Atlantic coasts (northern Spain) and in the center (around Madrid), rural municipalities tend to be concentrated in inland Spain (**Figure 3**). Most agro-urban municipalities are located in the South (Andalusia and Murcia). Agro-urban areas typically have a medium-sized population and a relatively strong primary sector compared with other parts of Spain.

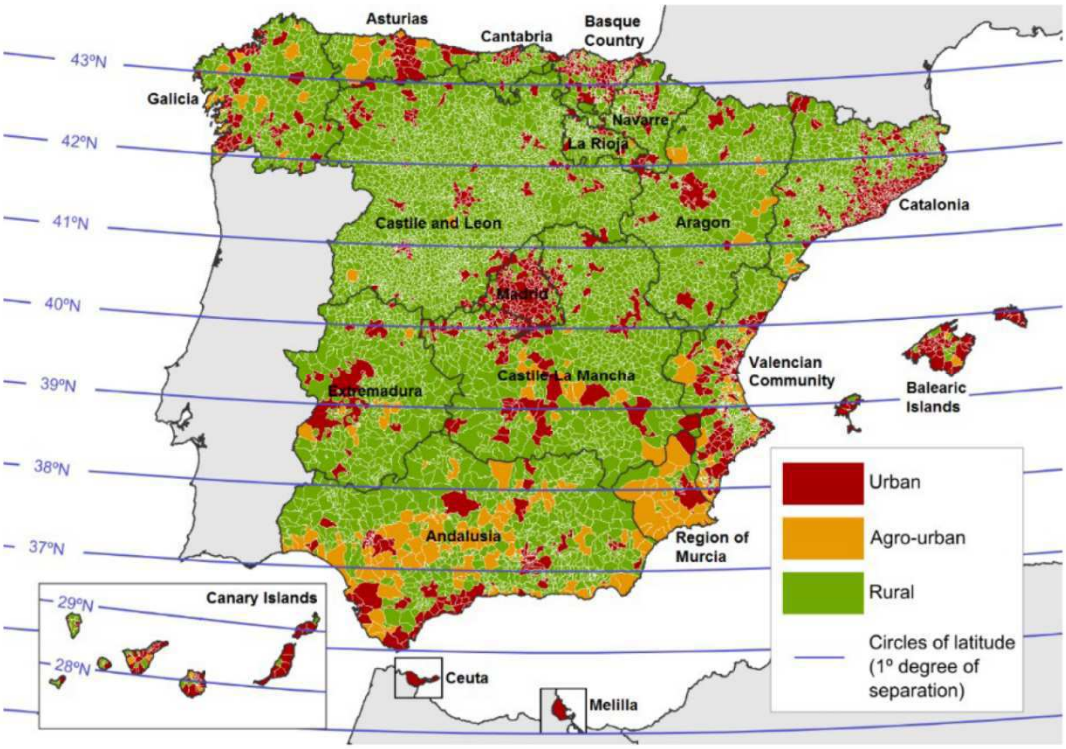


Figure 3. Classification of urban, agro-urban and rural municipalities in Spain. Latitude bands are shown.

According to SMR results for municipalities grouped within the same degree of latitude, an increased death risk due to GPA can be observed not only in the far North of the Iberian Peninsula but also in the South (degree 36° N and statistically significant: SMR = 1.79) for both sexes (**Figure 3**; **Table 2**). The lowest death rate was in the 40° N strip (SMR = 0.75 and statistically significant), which runs across the center of the Iberian Peninsula and includes the capital, Madrid. By sexes, SMR results showed a significantly higher death risk for males in the coastal areas of Andalusia in southern Spain (36° N strip): SMR = 1.91; 95% CI = 1.29–2.72), however it was not significant for women. In the southernmost latitude in Spain, the Canary Islands, there was a higher risk for both sexes together and for women separately, however these results were not significant.

Table 2. Number of municipalities, deaths and standardized mortality ratio (SMR) results for each degree of latitude due to GPA disease over the period 1999–2016 in Spain.

Grade of Latitude	No. Municipalities	GPA Deaths	SMR (95% CI)		
			Both Sexes	Both Sexes	Women
43° N	519	64	1.28 (0.99–1.63)	1.16 (0.79–1.63)	1.44 (0.99–2.04)
42° N	1842	50	1.05 (0.78–1.38)	1.26 (0.87–1.76)	0.77 (0.44–1.25)
41° N	2078	70	0.83 (0.65–1.05)	0.86 (0.62–1.17)	0.79 (0.53–1.13)
40° N	1507	52	0.75 (0.56–0.99)	0.73 (0.49–1.06)	0.79 (0.50–1.17)
39° N	795	40	0.90 (0.64–1.22)	0.91 (0.58–1.37)	0.88 (0.51–1.41)
38° N	579	37	0.97 (0.68–1.34)	1.00 (0.62–1.51)	0.93 (0.53–1.53)
37° N	500	41	0.88 (0.63–1.20)	0.80 (0.50–1.23)	0.99 (0.61–1.53)
36° N	215	49	1.79 (1.33–2.37)	1.91 (1.29–2.72)	1.64 (0.98–2.55)
27°–29° N	88	20	1.28 (0.78–1.98)	0.87 (0.38–1.72)	1.84 (0.95–3.22)

Finally, in smoothed SMRs a similar pattern can be seen as the territory gradation according to latitude: there are several areas in the North and in the South with a higher than expected death rate while in the center there were several districts with lower than expected results (**Figure 4**). The lowest death risks due to GPA were noted in some areas of western Galicia (Northwest) and in metropolitan areas of Madrid and Barcelona. While of the four districts in Spain with a significantly higher risk of death (PP > 0.80), two were in the North (one in Galicia and one in the Basque Country) and two were in the South (Andalusia).

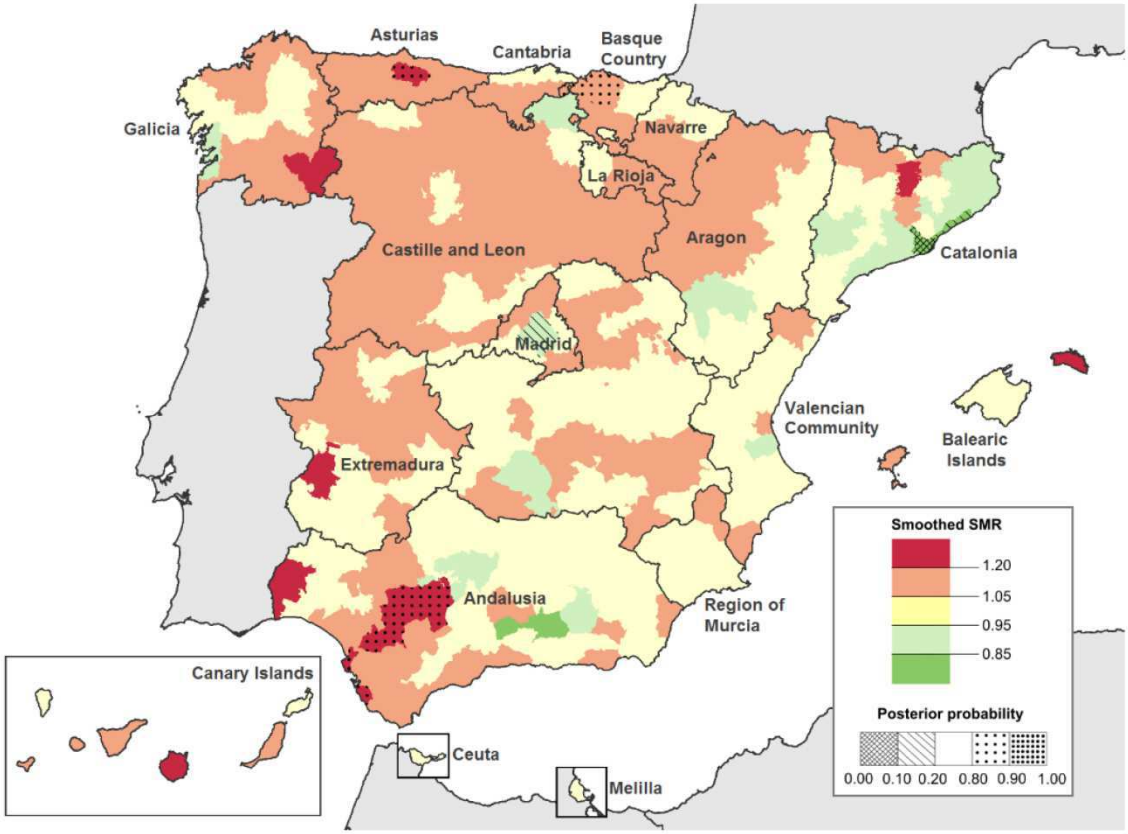


Figure 4. GPA Smoothed-SMR and Posterior probability (PP) by district in Spain (1999–2016, both genders). PP shows those districts with significantly higher (PP > 0.80) and lower (PP < 0.20) risk of death due to GPA than expected for Spain as a whole.

5.4 Discussion

This retrospective epidemiological study on mortality due to GPA is the first of its kind in Spain at a national level and over a long period. In this paper we offer a detailed description of geographic variation in mortality rates from different perspectives so as to understand the spatial distribution of this rare disease in Spain.

The annual increase we noted in the average age at death due to GPA is in step with a similar study in England over the period 1992–2013 [25]. The increasing survival rate has been achieved thanks to quicker diagnosis, which has enabled improved treatments to alleviate damage to affected organs [9]. This has been achieved partly due to the introduction of blood tests for ANCAs, as positive levels of ANCAs proteins are related to certain forms of systemic vasculitis such as GPA disease [26–29]. Better and faster diagnosis result in increased survival of the patients affected. Other authors reported an improvement in the survival of GPA patients in recent decades associated with the use of

less aggressive therapies, so reducing other complications, although this depends on the organ affected [25,30].

We found that age-adjusted mortality rates have decreased significantly since 1992. In a nationwide cohort of patients with end-stage renal disease due to GPA carried out in the United States, mortality rates decreased dramatically between 1995 and 2014, reflecting an improvement in the management of this disease over recent decades [31]. The change in the ICD code in 1999 (from the 9th to 10th edition) does not seem to have affected the trends in GPA age-adjusted mortality rates in Spain. Nevertheless, we do not know the impact of new clinical and molecular classifications for vasculitis on the ICD coding of deaths due to GPA in Spain i.e., "Revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides 2012" [32]. Although our change in mortality time trend was detected in 1992, long before this nomenclature revision for GPA, we cannot rule out some effect in our data. An increase in these rates was observed in the population as a whole and in males until 1994 (the opposite was observed in women), although from then on, the rate started to fall.

Variations in mortality in urban and rural areas could help to clarify whether GPA disease is triggered by environmental factors. In a study comparing vasculitis diseases in Germany, there were higher incidence rates in cities than in rural areas [19]. In a study of the northern province of Lugo in Spain, GPA incidence was also higher in urban municipalities [17]. However, the results of similar studies in other countries have not confirmed these trends [33, 34]. In our study, almost 5% of agro-urban municipalities showed significantly higher GPA mortality than expected for Spain as a whole, and this percentage was much lower in rural and urban municipalities. Agro-urban municipalities are medium-sized municipalities in terms of population (average 21,500 inhabitants) in which a substantial proportion of their citizens still work in the primary sector (more than 5%) in activities such as agriculture and livestock farming. A case-control study on this question in Norfolk (Eastern England) found a significant association between GPA and employment in farming [35]. In our study there is not sufficient ground to establish any positive association between farming in rural areas and GPA given the very low proportion of rural municipalities with higher SMR than expected. Our results are therefore not consistent with the findings obtained by Lane. Further studies must be carried out in order to find out if there is any link between mortality and farming or perhaps with the use of pesticides.

Several studies suggest that the incidence of GPA in Europe might follow a geographic pattern and decreases from North to South [36,37]. This suggests that there is a higher chance of developing GPA in northern countries perhaps due to environmental or genetic factors. Some authors highlight a possible link with the sunlight protection factor (higher in Mediterranean countries) and the benefits of Vitamin D [38]. The role of genetics in the

incidence of GPA was also discussed by various authors, particularly with regard to the more frequent presence of the PTPN22 R620W polymorphism in northern European populations [39]. Another comparative study showed a higher incidence of this disease in Northern Spain than in the south of the country [15]. In our study, mortality rates did not follow any obvious latitude pattern, with higher rates in areas of both Northern and Southern Spain, so coming to the same conclusions as observed in Germany in the period 1998–1999 [36]. In addition, latitude effect could be only noticeable at an international scale, such as European, and not in a single country. It is also possible that a geographic pattern in Spain could have been weakened due to internal migratory flows over the course of history [40].

When we analyzed the results for the different Spanish districts, no clear geographical North-South pattern could be observed. Nor could we find any significant rural/urban differences. This suggests that further, in-depth analysis is required to explain the geographic distribution of GPA in Spain. Family grouping in GPA disease is infrequent except in cases of first-degree relatives, which would suggest that the disease may have environmental causes [41]. Several studies point out that higher incidences of systemic vasculitis might be associated with environmental factors, such as silica exposure [42,43]. However, everything suggests that GPA is a complex, multifactorial disease in which both genetics and the environment play a role [20,35,44].

The present study has several limitations, such as the fact that our population-based death registry does not provide personal information with regard to occupation, disease-related genetic history or personal habits. For instance, a recent Spanish study found that GPA patients were more likely to have a history of smoking than the population as a whole [45]. Findings like these make the relationship between possible triggers for the disease and likely environmental factors increasingly difficult to unravel. In addition, most GPA epidemiological studies focus on prevalence and incidence, which makes it harder to compare mortality rates with other reports not only in Spain but also in other countries. Our research consisted of a nationwide, population-based study over a long period of time of the mortality directly attributed to GPA and followed a homogeneous, standardized methodology. The increase in average age at death identified in our study is an important finding in that it allows national or regional governments to plan the health services required to improve the life of GPA patients more effectively. It is important to emphasize for example that the treatments for GPA disease can be quite costly because patients often suffer relapses [46,47].

5.5 Conclusions

In conclusion, this study has analyzed deaths due to GPA in Spain over a 32-year period. Since 1992, mortality rates have fallen and average age at death has increased due to better treatments and prompt diagnosis. We found that areas categorized as agro-urban had a greater proportion of municipalities with higher than expected mortality in comparison with urban and rural areas. We also analyzed differences in death rates according to latitude, but no spatial pattern was found. More in-depth studies are required to enable us to gain a better understanding of the etiology of GPA disease.

5.6 References

1. Comarmond, C.; Cacoub, P. Granulomatosis with polyangiitis (Wegener): Clinical aspects and treatment. Autoimmun. Rev. 2014, 13, 1121–1125.

2. Lutalo, P.M.; D'Cruz, D.P. Diagnosis and classification of granulomatosis with polyangiitis (aka Wegener's granulomatosis). J. Autoimmun. 2014, 48–49, 94–98.

3. Orphanet Report Series. Prevalence of Rare Diseases: Bibliographic Data. Available online: http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf (accessed on 12 April 2019).

4. Commision of the European Communities Decision No 1295/1999/EC of the European Parliament and of the Council of 29 April 1999 Adopting a Programme of Community Action on Rare Diseases within the Framework for Action in the Field of Public Health (1999–2003). Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31999D1295&from=en. (accessed on 12 April 2019).

5. Relle, M.; Föhr, B.; Fasola, F.; Schwarting, A. Genetics and pathophysiology of granulomatosis with polyangiitis (GPA) and its main autoantigen proteinase 3. Mol. Cell. Probes 2016, 30, 366–373.

6. Pearce, F.; Lanyon, P.C.; Grainge, M.; Watts, R.; Hubbard, R. Mortality in granulomatosis with polyangiitis: A population-based case control study. Rheumatology 2018, 57, key075-204.

7. Tan, J.A.; Choi, H.K.; Xie, H.; Sayre, E.C.; Esdaile, J.M.; Aviña-Zubieta, J.A. All-cause and Cause-specific Mortality in Granulomatosis with Polyangiitis: A Population-based Study. Arthritis Care Res. 2018, 71, 155–163.

8. Tavakolpour, S.; Alesaeidi, S. Rituximab as an effective and probably safe treatment for granulomatosis with polyangiitis (Wegener's Granulomatosis). Int. Inmmunopharmacol. 2019, 70, 67–68.

9. Phillip, R.; Luqmani, R. Mortality in systemic vasculitis: A systematic review. Clin. Exp. Rheumatol. 2008, 26, S94–S104.

10. Watts, R.A.; Mahr, A.; Mohammad, A.J.; Gatenby, P.; Basu, N.; Flores-Suarez, F. Classification, epidemiology and clinical subgrouping of antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis. Nephrol. Dial. Transplant. 2015, 30, i14–i22.

11. Kanecki, K.; Nitsch-Osuch, A.; Gorynski, P.; Tarka, P.; Bogdan, M.; Tyszko, P. Epidemiology of Granulomatosis with Polyangiitis in Poland (2011–2015). Adv. Exp. Med. Biol. 2018, 1116, 131–138.

12. Pearce, F.A.; Grainge, M.J.; Lanyon, P.C.; Watts, R.A.; Hubbard, R.B. The incidence, prevalence and mortality of granulomatosis with polyangiitis in the UK Clinical Practice Research Datalink. Rheumatology 2017, 56, 589–596.

13. Mohammad, A.J.; Jacobsson, L.T.; Westman, K.W.; Sturfelt, G.; Segelmark, M. Incidence and survival rates in Wegener's granulomatosis, microscopic polyangiitis, Churg-Strauss syndrome and polyarteritis nodosa. Rheumatology 2009, 48, 1560–1565.

14. Takala, J.H.; Kautiainen, H.; Malmberg, H.; Leirisalo-Repo, M. Incidence of Wegener's granulomatosis in Finland 1981–2000. Clin. Exp. Rheumatol. 2008, 26, S81–S85.

15. Romero-Gómez, C.; Aguilar-García, J.A.; García-de-Lucas, M.D.; Cotos-Canca, R.; Olalla-Sierra, J.; García-Alegría, J.J.; Hernández—Rodríguez, J. Epidemiological study of primary systemic vasculitides among adults in southern Spain and review of the main epidemiological studies. Clin. Exp. Rheumatol. 2015, 33, 89.

16. Bustamante, E.; Mari, B.; Monteagudo, M.; Casanovas, A.; Jordana, R.; Tolosa, C.; Oristrell, J. Analysis of a series of 55 patients with biopsy proven giant cell temporal arteritis. An. Med. Intern. 2004, 21, 473–476.

17. González-Gay, M.A.; García-Porrua, C.; Guerrero, J.; Rodríguez-Ledo, P.; Llorca, J. The Epidemiology of the Primary Systemic Vasculitides in Northwest Spain: Implications of the Chapel Hill Consensus Conference Definitions. Arthritis Rheum. 2003, 49, 388–393.

18. Ormerod, A.S.; Cook, M.C. Epidemiology of primary systemic vasculitis in the Australian Capital Territory and south-eastern New South Wales. Intern. Med. J. 2008, 38, 816–823.

19. Reinhold-Keller, E.; Zeidler, A.; Gutfleisch, J.; Peter, H.; Raspe, H.; Gross, W. Giant cell arteritis is more prevalent in urban than in rural populations: Results of an epidemiological study of primary systemic vasculitides in Germany. Rheumatology 2000, 39, 1396–1402.

20. Mahr, A.; Neogi, T.; Merkel, P. Epidemiology of Wegener's granulomatosis: Lessons from descriptive studies and analyses of genetic and environmental risk determinants. Clin. Exp. Rheumatol. 2006, 24, S82–S91.

21. World Health Organization. International Statistical Classification of Diseases and Related Health Problems; World Health Organization: Geneva, Switzerland, 2004; Volume 2, pp. 1–125.

22. Velleman, P.F. Definition and comparison of robust nonlinear data smoothing algorithms. J. Am. Stat. Assoc. 1980, 75, 609–615. [Google Scholar] [CrossRef]

23. Yu, B.; Barrett, M.J.; Kim, H.J.; Feuer, E.J. Estimating joinpoints in continuous time scale for multiple change-point models. Comput. Stat. Data Anal. 2007, 51, 2420–2427.

24. Besag, J.E.; York, J.; Mollié, A. A Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Stat. Math. 1991, 43, 1–20.

25. Wallace, Z.S.; Lu, N.; Unizony, S.; Stone, J.H.; Choi, H.K. Improved survival in granulomatosis with polyangiitis: A general population-based study. Semin. Arthritis Rheum. 2016, 45, 483–489.

26. McKinney, E.F.; Willcocksm, L.C.; Broecker, V.; Smith, K. The immunopathology of ANCA-associated vasculitis. Semin. Immunopathol. 2014, 36, 461–478.

27. Tarzi, R.; Pusey, C. Current and future prospects in the managementof granulomatosis with polyangiitis (Wegener's granulomatosis). Ther. Clin. Risk Manag. 2014, 10, 279–293.

28. Holle, J.; Gross, W.; Latza, U.; Nölle, B.; Ambrosch, P.; Heller, P.; Fertmann, R.; Reinhold-Keller, E. Improved Outcome in 445 Patients with Wegener's Granulomatosis in a German Vasculitis Center Over Four Decades. Arthritis Rheum. 2011, 63, 257–266.

29. Eriksson, P.; Jacobsson, L.; Lindell, A.; Nilsson, J.; Skogh, T. Improved outcome in Wegener's granulomatosis and microscopic polyangiitis? A retrospective analysis of 95 cases in two cohorts. J. Intern. Med. 2009, 265, 496–506.

30. Luqmani, R.; Suppiah, R.; Edwards, C.J.; Phillip, R.; Maskell, J.; Culliford, D.; Jayne, D.; Morishita, K.; Arden, K. Mortality in Wegener's granulomatosis: A bimodal pattern. Rheumatology 2011, 50, 697–702.

31. Wallace, Z.; Zhang, Y.; Lu, N.; Stone, J.; Choi, H. Improving mortality in end-stage renal disease due to Granulomatosis with Polyangiitis (Wegener's) from 1995 to 2014: Data from the United States Renal Data System. Arthritis Care Res. 2018, 70, 1495–1500.

32. Jennette, J.C.; Falk, R.J.; Bacon, P.A.; Basu, N.; Cid, M.C.; Ferrario, F.; Flores-Suarez, L.F.; Gross, W.L.; Guillevin, L.; Hagen, E.C.; et al. 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis Rheum. 2013, 65, 1–11.

33. Cotch, M.F.; Hoffman, G.S.; Yerg, D.E.; Kaufman, G.I.; Targonski, P.; Kaslow, R.A. The epidemiology of Wegener's granulomatosis. Estimates of the five-year period prevalence, annual mortality, and geographical disease distribution from population-based data sources. Arthriris Rheum. 1996, 39, 87–92.

34. Carruthers, D.M.; Watts, R.A.; Symmons, D.P.; Scott, D.G. Wegener's granulomatosis increased incidence or increased recognition? Br. J. Rheumatol. 1996, 35, 142–145.

35. Lane, S.; Watts, R.; Bentham, G.; Innes, N.; Scott, D. Are environmental factors important in Primary Systemic Vasculitis? Arthritis Rheum. 2003, 48, 814–823.

36. Reinhold-Keller, E.; Herlyn, K.; Wagner-Bastmeyer, R.; Gutfleisch, J.; Peter, H.; Raspe, H.; Gross, W. No difference in the incidences of vasculitides between north and south Germany: First results of the German vasculitis register. Rheumatology 2001, 41, 540–549.

37. Koldingsnes, W.; Nossent, H. Epidemiology of Wegener's granulomatosis in northern Norway. Arthritis Rheum. 2000, 43, 2481–2487.

38. Gatenby, P.A.; Kucas, R.M.; Engelsen, O.; Ponsobony, A.L.; Clements, M. Antineuthrophil cytoplasmic antibody-associated vasculitis: Could geographic patterns be explained by ambient ultraviolet radiation? Arthritis Rheum. 2009, 61, 1417–1424.

39. Catanoso, M.; Macchioni, P.; Boiardi, L.; Manenti, L.; Tumiati, B.; Cavazza, A.; Luberto, F.; Pipitone, N.; Salvarani, C. Epidemiology of granulomatosis with polyangiitis (Wegener's granulomatosis) in Northern Italy: A 15-year population-based study. Semin. Arthritis Rheum. 2014, 44, 202–207.

40. Bycroft, C.; Fernandez-Rozadilla, C.; Ruiz-Ponte, C.; Quintela, E.; Carracedo, A.; Donnelly, P.; Myers, S. Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula. Nat. Commun. 2019, 10, 551.

41. Trénor-Larraz, P.; Martínez-Cristóbal, A.; Castellano-Cuesta, J.A. Vasculitis sistémicas asociadas a ANCA: Granulomatosis con Poliangeítis (de Wegener) y Poliangeítis Microscópica.

In Enfermedades Reumáticas, 2nd ed.; Ibáñez & Plaza Asociados: Valencia, Spain, 2013; pp. 395–416.

42. Gómez-Puerta, J.A.; Gedmintas, L.; Costenbader, K.H. The association between silica exposure and development of ANCA-associated vasculitis: Systematic review and meta-analysis. Autoimmun. Rev. 2013, 12, 1129–1135.

43. Moulin, P.; Lehucher-Michel, M.P. Wegener's disease and exposure to silica. Study of the physiopathological mechanisms. Presse Med. 2004, 33, 1349–1351.

44. Gibelin, A.; Maldini, C.; Mahr, A. Epidemiology and etiology of Wegener Granulomatisis, Microscopic Polyangiitis, Churg-Strauss Syndrome and Goodpasture Syndrome: Vasculitides with frequent lung involvement. Semin. Respir. Crit. Care 2011, 32, 264–273.

45. Martinez-Morillo, M.; Grados, D.; Naranjo-Hans, D.; Mateo, L.; Holgado, S.; Olivé, A. Granulomatosis con poliageítis (Wegener). Descripción de 15 casos. Reumatol. Clin. 2012, 8, 15–19.

46. Kong, A.; Kim, G.; Michalska, M.; Best, J. Costs of Disease Relapses Among Individuals with Granulomatosis, with Polyangiitis, or Microscopic Polyangiitis in the United States. Rheumatol. Ther. 2018, 5, 159–170.

47. Springer, J.; Nutter, B.; Langford, C.; Hoffman, G.; Villa-Forte, A. Granulomatosis with Polyangiitis (Wegener's): Impact of maintenance therapy duration. Medicine 2014, 93, 82–90.

Chapter 6

Geographic analysis of motor neuron disease mortality and heavy metals released to rivers in Spain

Geographic analysis of motor neuron disease mortality and heavy metals released to rivers in Spain

Published in: International Journal of Environmental Research and Public Health

IJERPH 2018, 15(11), 2522; doi: 10.3390/ijerph15112522

Authors: Germán Sánchez-Díaz [1,2,3], Francisco Escobar [2], Hannah Badland [4], Greta Arias-Merino [1], Manuel Posada de la Paz [1,3] and Verónica Alonso-Ferreira [1,3].

[1] Institute of Rare Diseases Research (IIER), Instituto de Salud Carlos III, 28029 Madrid, Spain

[2] Department of Geology, Geography and Environmental Sciences, University of Alcalá 28801 Alcalá de Henares, Spain

[3] Centre for Biomedical Network Research on Rare Diseases (CIBERER), 28029 Madrid, Spain

[4] Healthy Liveable Cities Group, Centre for Urban Research, RMIT University, Melbourne VIC 3001, Australia

International Journal of Environmental Research and Public Health

MDPI

Article

Geographic Analysis of Motor Neuron Disease Mortality and Heavy Metals Released to Rivers in Spain

Germán Sánchez-Díaz [1,2,3], Francisco Escobar [2], Hannah Badland [4], Greta Arias-Merino [1], Manuel Posada de la Paz [1,3] and Verónica Alonso-Ferreira [1,3,*]

[1] Institute of Rare Diseases Research (IIER), Instituto de Salud Carlos III, Madrid 28029, Spain; g.sanchez@externos.isciii.es (G.S.-D.); gretaariasmerino@gmail.com (G.A.-M.); mposada@isciii.es (M.P.d.l.P.)
[2] Department of Geology, Geography and Environmental Sciences, University of Alcala, Alcalá de Henares 28801, Spain; francisco.escobar@uah.es
[3] Centre for Biomedical Network Research on Rare Diseases (CIBERER), Madrid 28029, Spain
[4] Healthy Liveable Cities Group, Centre for Urban Research, RMIT University, Melbourne VIC 3001, Australia; hannah.badland@rmit.edu.au
* Correspondence: valonso@isciii.es; Tel.: +34-91-822-20-89

Received: 23 October 2018; Accepted: 8 November 2018; Published: 1

Abstract: The etiology of motor neuron disease (MND) is still u
to: (1) analyze MND mortality at a f ne-grained level; and (2) ex
metals released into Spanish river basins. MND deaths were ex
mortality registry (2007–2016). Standardized mortality ratios
a municipal level. Sites that emitted quantities of heavy met
were obtained from the European Pollutant Release and Transf
relative risks for non-exposed and exposed municipalities (co
section) by type of heavy metal were analyzed using a log-lir
higher in central and northern municipalities. SMRs were 1.1
to heavy metals than in non-exposed areas: 0.95 (0.92–0.96)
we found the following increased MND death risks in expos
20.0% for zinc, 16.7% for arsenic, 15.7% for chromium, 15.4%
12.4% for mercury. This study provides associations between
exposed municipalities. Further studies investigating heavy m
in MND understanding.

Keywords: motor neuron disease; population-based mortality; environmental factors

1. Introduction

Motor neuron disease (MND) is a neurodegenerative condi
diminish progressively in the central nervous system [1]. Curre
interventions have been found to stop the progression of sympt
5–10% of MND cases are attributed to familial or genetic origin.
patients, environmental factors may have a key role in the devel

Numerous studies have examined the effects that metals
focusing especially on the role of heavy metals in health [9,1
associated with amyotrophic lateral sclerosis pathogenesis for
heavy metals were found in tissues and f uids from MND patie

Despite the lack of a general agreement in def ning heavy m
those chemical elements having a specif c density of more than

Int. J. Environ. Res. Public Health 2018, 15, 2522; doi:10.3390/ ijerph15112522

International Journal of Environmental Research and Public Health

IMPACT FACTOR 2.145

Understanding Health Promotion Policy Processes: A Study of the Government Adoption of the Achievement Program in Victoria, Australia

Volume 15 · Issue 11 | November 2018

MDPI

mdpi.com/journal/ijerph
ISSN 1660-4601

Geographic analysis of motor neuron disease mortality and heavy metals released to rivers in Spain

Abstract

Background: The etiology of motor neuron disease (MND) is still unknown. The aims of this study were to: (1) analyze MND mortality at a fine-grained level; and (2) explore associations of MND and heavy metals released into Spanish river basins.

Methods: MND deaths were extracted from the Spanish nationwide mortality registry (2007–2016). Standardized mortality ratios (SMRs) for MND were estimated at a municipal level. Sites that emitted quantities of heavy metals above the regulatory thresholds were obtained from the European Pollutant Release and Transfer Register database (2007–2015). The relative risks for non-exposed and exposed municipalities (considering a downstream 20 km river section) by type of heavy metal were analyzed using a log-linear model.

Results: SMRs were significantly higher in central and northern municipalities. SMRs were 1.14 (1.10–1.17) higher in areas exposed to heavy metals than in non-exposed areas: 0.95 (0.92–0.96). Considering the different metals, we found the following increased MND death risks in exposed areas: 20.9% higher risk for lead, 20.0% for zinc, 16.7% for arsenic, 15.7% for chromium, 15.4% for cadmium, 12.7% for copper, and 12.4% for mercury.

Conclusion: This study provides associations between MND death risk and heavy metals in exposed municipalities. Further studies investigating heavy metal exposure are needed to progress in MND understanding.

Keywords: motor neuron disease; population-based mortality; geographical analysis; heavy metals; environmental factors

6.1 Introduction

Motor neuron disease (MND) is a neurodegenerative condition in which motor neuron functions diminish progressively in the central nervous system [1]. Currently, there is no cure for MND, and no interventions have been found to stop the progression of symptoms, leading to mortality [2,3]. About 5–10% of MND cases are attributed to familial or genetic origin. Thus, for the remainder of diagnosed patients, environmental factors may have a key role in the development of the disease [4-8]. Numerous studies have examined the effects that metals have in neurodegenerative diseases, focusing especially on the role of heavy metals in health [9,10]. Heavy metal exposure has been associated with amyotrophic lateral sclerosis pathogenesis for more than 150 years, especially since heavy metals were found in tissues and fluids from MND patients [11]. Despite the lack of a general agreement in defining heavy metals, they are typically referred to as those chemical elements having a specific density of more than 5 g/cm3 [12]. They are involved in a wide range of processes in various receiving environments such as air, soil, or water. Heavy metal emissions have been continually increasing since the middle of the 19th century at the global level [13].

The association between mercury and MND has been studied, being evident its effects on the human nervous system, leading to brain alterations or to shyness, tremors, irritability, and changes in vision or hearing [14]. Lead pollution, described as toxic for more than 100 years, is a major problem in drinking water pipes in some undeveloped countries, but not in Spain [48]. Chronic exposure to lead has been linked to mental retardation, birth defects, and brain damage, among others [16,17]. Lead is one of the most studied heavy metals in terms of exposure and receiving environment, with results confirming its association with MND [18,19]. Cadmium is a water-soluble heavy metal affecting the kidney and bones. Chromium is highly persistent in water sediments and, in elevated concentrations, it is related to ulcers and can induce DNA damage [20]. A case–control study found an association between cadmium exposure and the occurrence of amyotrophic lateral sclerosis (a type of MND accounting for the 85% of MND cases) in Catalonia, Spain [21]. Copper toxicity can lead to hepatic disorders or neurodegenerative changes [22]. Zinc excesses can cause brain damage or affect the respiratory and gastrointestinal tracts apart from prostate [23], with one of the most common exposing sources being drinking water [24].

At the same time, an increase of MND death rates has been reported in the past years [25-28]. In Spain, the patterning of MND has been examined, showing geographic variability at the provincial level, but MND mortality has not been further explored at a finer spatial scale throughout the country, and few studies including exposure types have been carried out [18,21,28]. Examining the associations at a more detailed level (e.g., municipality) using geographically referenced environmental data allows a deeper interrogation of the factors

that potentially contribute to MND. If a relationship is found between increased MND mortality and heavy metal emissions, it will inform future avenues of inquiry to progress our understanding of MND etiology.

The objectives of this study were to: (1) Analyze the mortality distribution of MND at a municipal level in a 10-year period; (2) Characterize heavy metal emissions released to river basins; and (3) Explore associations between heavy metals exposure in water and MND deaths in Spain over a 10-year period.

6.2 Materials and methods

Mortality data were extracted from the annual death registry kept by the National Statistics Institute (NSI) (Instituto Nacional de Estadística), Spain. MND deaths were obtained from 2007 through 2016. These were identified from the International Classification of Diseases, 10th revision, code G12.2, which included all forms of MND. For Standardized Mortality Ratios (SMRs) and confidence intervals (CI) calculations, ages were grouped in five-years intervals up to 84 years old and older.

Using the European Pollutant Release and Transfer Register (E-PRTR) database, we extracted heavy metal quantities released into river basins by emission point in Spain during 2007–2015, and these were assigned to geographic coordinates provided by this registry [29]. E-PRTR is an information platform, a register with a list of facilities that have emitted polluting substances exceeding certain thresholds. This register contains information about the location, emitted substances, quantities per year, activity, and receiving environment. Facilities are required by law to report the release of pollutants, and this law is implemented by all EU member countries, which have the task to impose penalties on sites with emissions beyond the limits in an effective, proportional, and dissuasive way. Of note, information is only reported by an installation to the E-PRTR when the quantities of pollutants released exceed certain thresholds.

Heavy metals included in this study were: arsenic, cadmium, copper, chromium, mercury, lead, and zinc. These heavy metals were selected because there are several studies linking them to MND symptoms and development [11]. We geocoded the sites that released heavy metal emissions into the rivers during the mentioned period. Using geographic criteria, we identified a 20 km river section downstream from the emission point and selected the municipalities that intersected the sections. These were classified as exposed municipalities. The municipalities that did not intersect these river-sections were defined as non-exposed municipalities. We then calculated the times that each emission point had exceeded the thresholds by metal as indicated by the E-PRTR. The annual limits per

emission point and by metal were: mercury 1 kg; arsenic and cadmium 5 kg; lead 20 kg; chromium and copper 50 kg; Zinc 100 kg [29].

We fitted log-linear models on the assumption that the number of MND deaths per stratum followed a Poisson distribution. Observed cases were the dependent variable, and expected cases were included as offset in the models. A term we called “exposure” (distance 20 km or less from the facility) was included as the independent variable. The regression coefficient of this exposure term gave us the logarithm of the ratio between the respective mortality ratios for the exposed and reference zones, which we called incidence report rates (IRR). The advantage of using Poisson regression models is that they provide a magnitude of association (IRR) and a p-value of global significance between the exposure variable and mortality.

Statistical analyses were performed using the Stata computer software program, while the ArcGIS software program (Esri, Redlands, CA, USA) was used for geographic analysis and cartographic representations.

Research Ethics: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Instituto de Salud Carlos III (CEI 50/2013).

6.3 Results

There were 9434 deaths due to MND in the period 2007–2016 in the Spanish territory. As shown in **Figure 1a**, results obtained for municipal SMRs seemed to be random across the period 2007–2016, when it was possible to observe both high and low SMRs throughout the national territory. **Figure 1b** depicts only significant SMRs, highlighting 90 municipalities. The majority of lower-than-expected SMRs values (15 municipalities) were located in the southern half of the Spanish territory, whereas the majority of the higher-than-expected SMRs results (75 municipalities) were located in the northern half of the country.

We identified 129 sites that reported heavy metal quantities emitted to river basins that exceeded the thresholds during the period 2007–2015 (**Figure 2**), noting that a site may release more than one type of heavy metal. Overall, 92 sites were located in the northern part of Spain, mainly in Basque Country and Catalonia. Zinc was the most commonly reported exceeding heavy metal emission, whilst cadmium was the least (99 and 23 sites, respectively). The main activities identified at the sites were: wastewater treatment sludge, thermal power stations, steel industry, paper industry, and chemical industry. Overall, 458

municipalities were affected (5.64% of the total Spanish municipalities) as they intersected 20 km of river downstream from the sites.

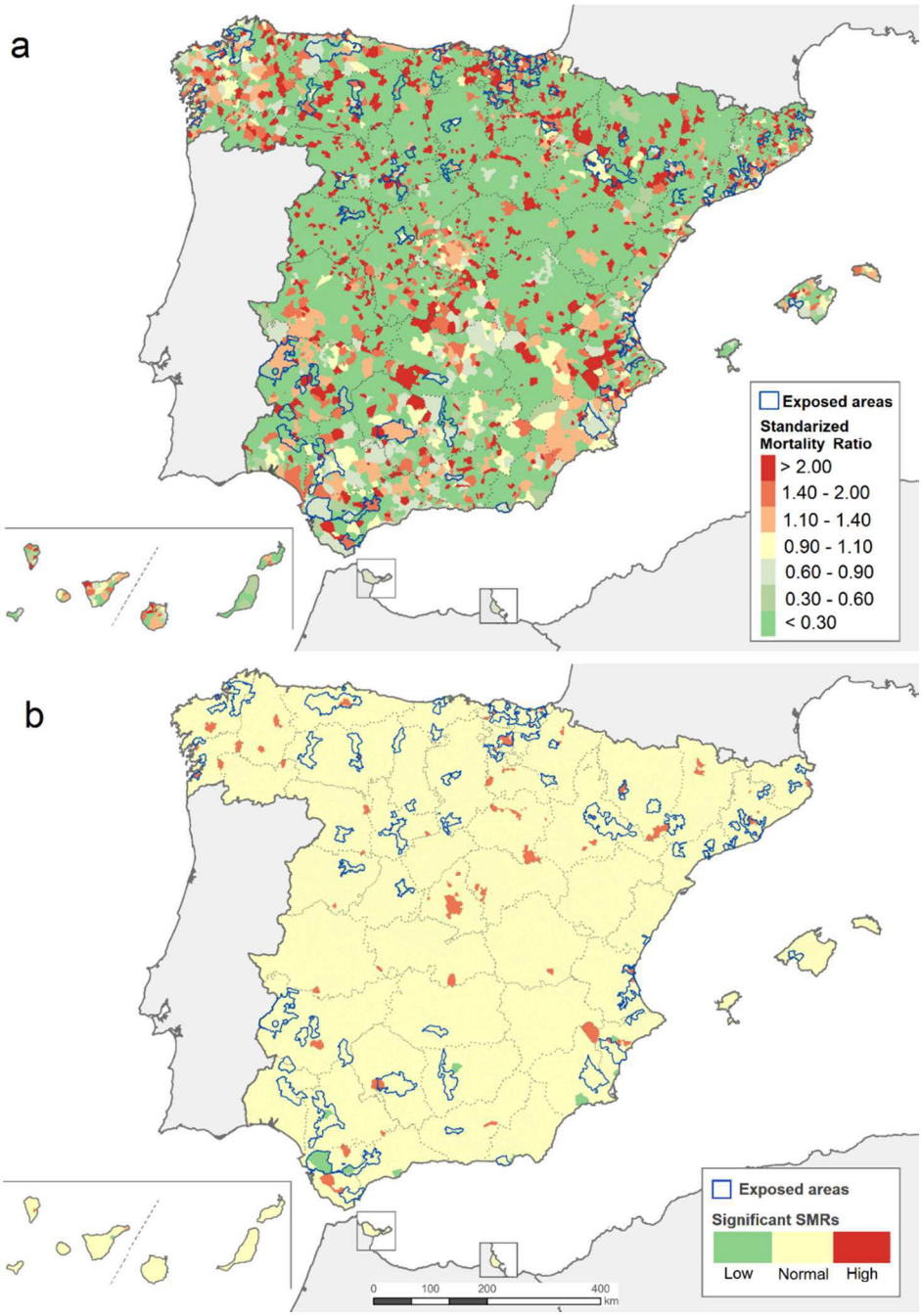


Figure 1. (a) Distribution of standardized mortality ratios (SMRs) due to motor neuron disease (MND) from 2007 to 2016. **(b)** Distribution of lower- and higher-than-expected SMRs.

According to the location of the emission point for each heavy metal, the number of exposed municipalities varied: zinc n = 382, copper n = 235, arsenic n = 198, mercury n = 169, lead n = 163, chromium n = 156, and cadmium n=93.

The results of the Poisson Regressive Model used to compare mortality due to MND in exposed versus non-exposed municipalities is shown in **Table 1**. Significant SMR differences between the exposed and non-exposed municipalities existed. However, SMR attributed to MND was significantly higher in exposed municipalities for all metals and for individual metals. The IRR were higher in exposed than in non-exposed municipalities ($p < 0.001$), with an 18.4% increased risk of MND for people living in municipalities exposed to heavy metals. MND risk varied by metal exposure—lead posed a 20.9% risk (highest), whereas copper and mercury increased the risk by 12.4% (lowest).

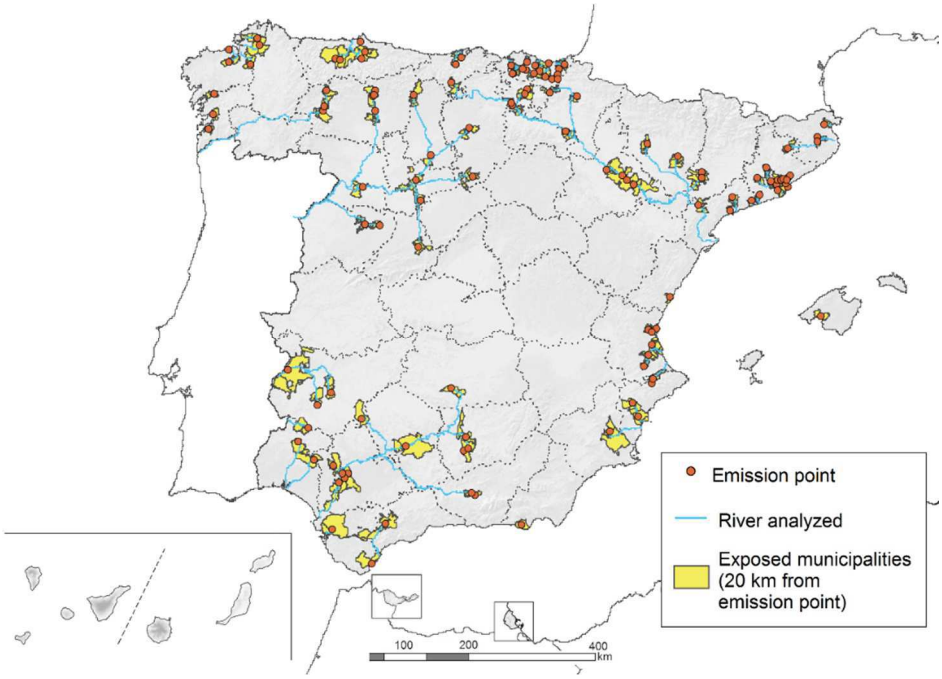


Figure 2. Location of sites (emission point) releasing heavy metals to river basins in the period 2007–2015 in Spain.

Table 1. Comparison of motor neuron disease mortality in exposed and non-exposed municipalities based on individual heavy metals: number of deaths (%), SMR, and incidence report rate (IRR). MND: motor neuron disease. SMR: Standardized mortality ratio.

	Non-Exposed Municipalities		Exposed Municipalities		Comparison between Non-Exposed vs. Exposed Municipalities			
	Deaths from MND (%)	SMR	Deaths from MND (%)	SMR	IRR	95% CI	p-Value	Increased Risk
All heavy metals	6158 (65.27)	0.95 (0.92–0.96)	3276 (34.73)	1.14 (1.10–1.17)	1.184	1.135–1.236	<0.001	18.4%
Arsenic	7348 (77.89)	0.97 (0.95–0.99)	2086 (22.11)	1.13 (1.08–1.18)	1.167	1.112–1225	<0.001	16.7%
Cadmium	8235 (87.29)	0.98 (0.96–1.00)	1199 (12.71)	1.14 (1.07–1.20)	1.154	1.086–1.227	<0.001	15.4%
Chromium	7366 (78.08)	0.97 (0.95–0.99)	2068 (21.92)	1.12 (1.08–1.17)	1.157	1.102–1.215	<0.001	15.7%
Copper	7231 (76.65)	0.97 (0.95–1.00)	2203 (23.35)	1.10 (1.05–1.14)	1.127	1.075–1.182	<0.001	12.7%
Lead	7533 (79.34)	0.97 (0.94–0.99)	1901 (20.66)	1.17 (1.12–1.22)	1.209	1.150–1.271	<0.001	20.9%
Mercury	7397 (78.41)	0.98 (0.95–1.00)	2037 (21.59)	1.10 (1.05–1.15)	1.124	1.071–1.181	<0.001	12.4%
Zinc	6309 (66.88)	0.94(0.92–0.97)	3125 (33.12)	1.13 (1.09–1.17)	1.200	1.150–1.253	<0.001	20.0%

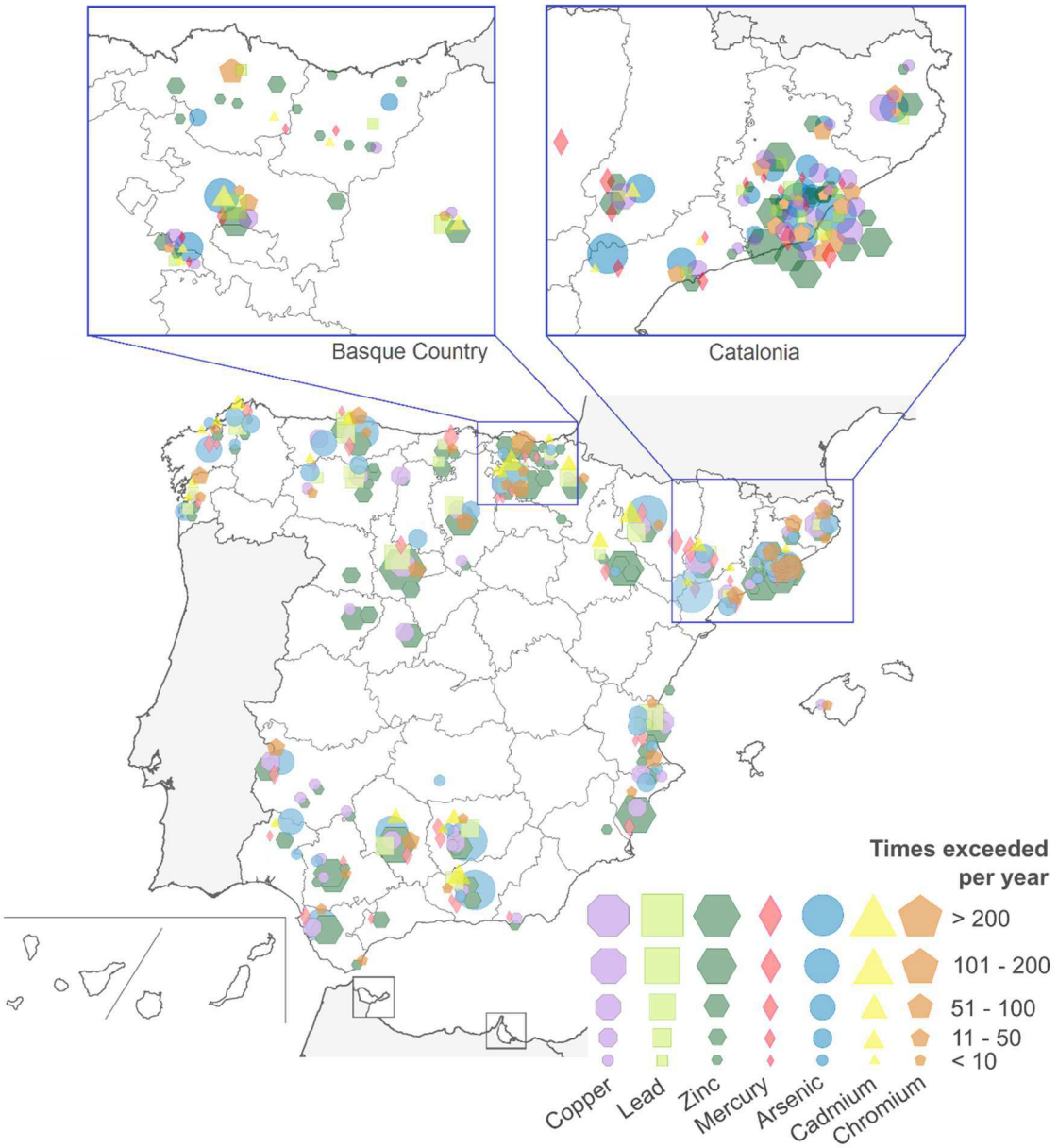


Figure 3. Number of times that heavy metal emissions exceeded the annual threshold by heavy metal and per site in 2007–2015. Only the emission points that in at least one year exceeded the maximum regulatory emission level are displayed.

Comparing the quantities of heavy metal emissions with the EU regulatory thresholds, we observed that zinc was the most common substance emitted exceeding its limit (100 kg per year per site), as shown in **Figure 3**. Across many emission points, the annual threshold of zinc was exceeded by more than 100, 200, or even 500 times. Cadmium and lead were the heavy metals least likely to exceed the regulatory thresholds, with exceeding values measured up to maximum 100 times per year per site as during the period 2007–2015.

6.4 Discussion

This study provides a detailed map of MND mortality in Spain by exploring possible associations between the MND and heavy metal exposure. The discovery of patterning of significant high and low mortality rates through the Spanish territory supports a non-randomly spatial distribution of MND. This spatial pattern is in line with and extends from other MND studies conducted at a provincial level [18,28]. The increased MND risk identified in certain municipalities close to heavy metal emissions suggests that environmental factors may contribute to MND etiology.

Much research supports the notion of heavy metals being harmful to health, with effects varying according to dose, duration, exposure conditions, bioavailability, and chemical species [11,30]. Despite different studies attempting to rank the health effects of specific heavy metals, a consensus has not been obtained. For example, the U.S. Department of Health and Human Services, Agency for Toxic Substances and Disease Registry (ATSDR), recognized arsenic, lead and mercury as being the heavy metals most harmful to human health, on the basis of a combination of frequency and potential exposure [31]. In relation to specific heavy metals, we found that municipalities exposed to lead had the highest increased MND risk (20.9%), while those exposed to mercury showed the lowest increased risk (12.4%) in reference to non-exposed areas.

Although pollutant emission data were available from 2001, we only used data from 2007; in the period 2001–2007, it was not necessary for emission points to report the exceeding quantities. The register was further improved in 2007, with more substances included and a better evaluation of activities occurring at the sites [32]. The strengths of this study are the potential extension of this method to include information on other substances and the receiving environment (e.g., air, soil) and an exhaustive analysis of the activity of sites and the measurements taken after the thresholds were exceeded. The database from which we extracted the heavy metal emissions data are also available for other 34 EU countries. The possibility of collecting national MND deaths registries, standardized and comparable across countries, and the study of heavy metal emissions in a continental context would provide a large-scale study with large-scale exposure heterogeneity. Identifying associations in a wider geographic context would be very useful for the European health policy planning and for further identifying the etiology of and risk factors for MND. Our paper only focuses on the likely effects of isolated metals on MND mortality and not on the synergistic or antagonistic effects of mixing various types of metals in water. For instance, there are several articles relating the severe effects in mice of metals like cadmium combined with fluoride dissolved in water [33]. These possible synergistic effects should be studied more deeply in the future.

There are limitations in this study. The mortality statistics from NSI did not provide individual information about personal habits, disease-related genetic history, or occupation. We did not have a direct measurement of individual exposure to heavy metals. Another limitation is that this registry does not include information about the possible movement of patients from small to larger municipalities, once diagnosed, for better treatment. This would imply that MND deaths could be inflated in big cities, while underestimated in small municipalities. These changes in the municipality of residence could result in less accuracy when exploring likely associations. However, the population-based and standardized mortality data used are a noticeable advantage because the registry has national coverage.

Exposure to metals is, in most cases, a long-term process; it would, therefore, be valuable to have reliable and standardized data about pollutant emissions available for longer time periods (i.e., decades). This information could be used to better assess and understand the association between neurodegenerative diseases and pollutant exposures. It should be noted that sites are not required to report quantity emissions when they do not exceed the set threshold; therefore, when we found a gap in a year information in one emission point in the studied period, it could mean two things: a released quantity lower than the limit or no emission produced. The absence of data seems to indicate that, in the event of heavy metal release, this would not involve a health hazard. Finally, it is important to note that this is emergent research that needs to be analyzed with greater precision in future studies, especially as regards the interaction of heavy metals with the rivers where they are released in terms of dissolution of substances in water and hydraulic flow of each river, and the study of multiple cross-sectional distances from the emission points.

While this study does not definitively link heavy metal exposure to MND, it identifies a valid approach and further lines of inquiry. Within a national context, this is a novel study which indicates characteristics of MND mortality and heavy metals that suggest further investigations at a local level, considering river characteristics and the behaviors of the different heavy metal in rivers.

6.5 Conclusions

In conclusion, we described the detailed mortality rates due to MND during 10 years in Spain, finding patterns that indicate higher rates than expected of MND in northern municipalities. Also, we explored heavy metal emissions in river basins, identifying emission points exceeding the recommended thresholds predominantly located in northern areas. Increased MND mortality risk in the exposed areas was shown when analyzing heavy

metals individually and together. This investigation underscores the value of combining geographic and epidemiologic techniques for the understanding of the patterning of relationships between rare diseases and environmental exposures.

Author Contributions: Conceptualization, F.E. and V.A.-F.; Data curation, G.S.-D. and V.A.-F.; Formal analysis, G.S.-D., H.B. and G.A.-M.; Funding acquisition, M.P.d.l.P. and V.A.-F.; Investigation, G.S.-D. and G.A.-M.; Methodology, F.E., H.B., G.A.-M. and V.A.-F.; Project administration, M.P.d.l.P. and V.A.-F.; Resources, M.P.d.l.P.; Software, G.S.-D.; Supervision, F.E., H.B. and V.A.-F.; Visualization, G.S.-D. and F.E.; Writing–original draft, G.S.-D.; Writing–review & editing, G.S.-D., F.E., H.B. and V.A.-F.

Funding: This research was supported by the Spanish Strategy Action for Health (AESI), project no. TPY1238/15. GS enjoys a research grant funded by the Spanish Ministry of Education, Culture and Sport (no. FPU14/03914).

Acknowledgments: Hannah Badland (BD) is supported by an RMIT Vice Chancellor's Senior Research Fellowship and is the 2018 Australian Health Promotion Association Thinker in Residence.

Conflicts of Interest: The authors have no conflict of interest to disclose.

6.6 References

1. Kiernan, M.C. Motor neuron disease in 2017: Progress towards therapy in motor neuron disease. Nat. Rev. Neurol. 2018, 14, 65–66. [Google Scholar] [CrossRef] [PubMed]

2. Mehta, P.; Kaye, W.; Raymond, J.; Wu, R.; Larson, T.; Punjani, R.; Heller, D.; Cohen, J.; Peters, T.; Muravov, O.; Horton, K. Prevalence of amyotrophic lateral sclerosis—United States, 2014. MMWR Suppl. 2018, 67, 216–218. [Google Scholar] [CrossRef] [PubMed]

3. Miller, R.G.; Appel, S.H. Introduction to supplement: The current status of treatment for ALS. Amyotroph. Lateral Scler. Frontotemporal Degener. 2017, 18.

4. Rosen, D.R.; Siddique, T.; Patterson, D.; Figlewicz, D.A.; Sapp, P.; Hentati, A.; Donaldson, D.; Goto, J.; O'Regan, J.P.; Deng, H.X. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 1993, 362, 59–62.

5. Strong, M.; Rosenfeld, J. Amyotrophic lateral sclerosis: A review of current concepts. Amyotroph. Lateral Scler. Other Motor Neuron Disord. 2003, 4, 136–143.

6. Kiernan, M.C.; Vucic, S.; Cheah, B.C.; Turner, M.R.; Eisen, A.; Hardiman, O.; Burrell, J.R.; Zoing, M.C. Amyotrophic lateral sclerosis. Lancet 2011, 377, 942–955.

7. Ingre, C.; Roos, P.M.; Piehl, F.; Kamel, F.; Fang, F. Risk factors for amyotrophic lateral sclerosis. Clin. Epidemiol. 2015, 7, 181–193.

8. Huisman, M.H.; de Jong, S.W.; van Doormaal, P.T.; Weinreich, S.S.; Schelhass, H.J.; van der Kooi, A.J.; de Visser, M.; Veldink, J.H.; van den Berg, L.H. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J. Neurol. Neurosurg. Psychiatry 2011, 82, 1165–1170.

9. Barnham, K.J.; Bush, A.I. Biological metals and metal-targeting compounds in neurodegenerative diseases. Chem. Soc. Rev. 2014, 43, 6727–6749.

10. Jaishankar, M.; Tsseten, T.; Anbalagan, N.; Mathew, B.B.; Beeregowda, K.N. Toxicity, mechanism and health effects of some heavy metals. Interdiscip. Toxicol. 2014, 7, 60–72.

11. Roos, P.M. Studies on Metals in Motor Neuron Disease. Available online: https://openarchive.ki.se/xmlui/handle/10616/41419 (accessed on 8 November 2018).

12. Järup, L. Hazard of heavy metal contamination. Br. Med. Bull. 2003, 68, 167–182.

13. Nragu, J.O. History of global metal pollution. Science 1996, 272, 223–224.

14. Praline, J.; Guennoc, A.; Limousin, N.; Hallak, H.; de Toffoi, B.; Corcia, P. Case report; ALS and mercury intoxication: A relationship? Clin. Neurol. Neurosurg. 2007, 109, 880–883.

15. Etabe, I.Z.; Contín, K.C.; Olalde, C.O.; Alonso, J.V. Release of lead and other metals from piping into drinking water in the Basque Country (Spain). Gac. Sanit. 2010, 24, 460–465.

16. Brochin, R.; Leone, S.; Phillips, D.; Shepard, N.; Zisa, D.; Angerio, A. The cellular effect of lead poisoning and its clinical picture. GUJHS 2008, 5, 1–8.

17. Campbell, A.M.; Williams, E.R.; Barltrop, D. Motor neuron disease and exposure to lead. J. Neurol. Neurosurg. Psychiatry 1970, 33, 877–885.

18. Santurtún, A.; Villar, A.; Delgado-Alvarado, M.; Riancho, J. Trends in motor neuron disease: Association with air lead levels in Spain. Neurol. Sci. 2016, 37, 1271–1275.

19. Kamel, F.; Umbach, D.M.; Hu, H.; Munsat, T.L.; Shefner, J.M.; Taylor, J.A.; Sandler, D.P. Lead exposure as a risk factor for amyotrophic lateral sclerosis. Neurodegener. Dis. 2005, 2, 195–201.

20. Bernard, A. Cadmium and its adverse effects on human health. Indian J. Med. Res. 2008, 138, 557–564.

21. Povedano, M.; Saez, M.; Martínez-Matos, J.; Barceló, M.A. Spatial assessment of the association between long-term exposure to environmental factors and the occurrence of Amyotrophic Lateral Sclerosis in Catalonia, Spain: A population-based nested Case-Control study. Neuroepidemiology 2018, 51, 33–49.

22. Gaetke, L.M.; Chow-Johnson, H.; Chow, C.K. Copper: Toxicological relevance and mechanisms. Arch. Toxicol. 2014, 88, 1929–1938.

23. Plum, L.M.; Rink, L.; Haase, H. The essential toxin: Impact of Zinc on human health. Int. J. Environ. Res. Public Health 2010, 7, 1342–1365.

24. Agency for Toxic Substances & Disease Registry (ATSDR): ToxFAQs for ZINC. 2006. Available online: https://www.atsdr.cdc.gov/toxprofiles/TP.asp?id=302&tid=54 (accessed on 22 October 2018).

25. Sejvar, J.J.; Holman, R.C.; Bresee, J.S.; Kochanek, K.D.; Schonberger, L.B. Amyotrophic lateral sclerosis mortality in the United States, 1979–2001. Neuroepidemiology 2005, 25, 144–152. [Google Scholar] [CrossRef] [PubMed]

26. Nakken, O.; Lindstrøm, J.C.; Tysnes, O.B.; Holmøy, T. Mortality trends of amyotrophic lateral sclerosis in Norway 1951–2014: An age-period-cohort study. J. Neurol. 2016, 263, 2378–2385.

27. Chiò, A.; Logroscino, G.; Traynor, B.J.; Collins, J.; Simeone, J.C.; Goldstein, L.A.; White, L.A. Global epidemiology of amyotrophic lateral sclerosis: A systematic review of the published literature. Neuroepidemiology 2013, 41, 118–130.

28. Alonso, V.; Villaverde-Hueso, A.; Hens, M.; Morales-Piga, A.; Abaitua, I.; de la Paz, M.P. Increase in motor neuron disease mortality in Spain: Temporal and geographical analysis (1990–2015). Amyotroph. Lateral Scler. 2011, 12, 192–193.

29. European Parliament. Regulation (EC) No 166/2006 of the European Parliament and of the Council of 18 January 2006 Concerning the Establishment of a European Pollutant Release and Transfer Register and Amending Council Directives 91/689/ECC and 96/61/EC; European Parliament: Brussels, Belgium, 2006.

30. Tchounwou, P.B.; Yedjou, C.G.; Patlolla, A.K.; Sutton, D.J. Heavy metal toxicity and the environment. In Molecular, Clinical and Environmental Toxicology; Springer: Basel, Switzerland, 2012.

31. Agency for Toxic Substances and Disease Registry. Substance Priority List-ATSDR. Available online: https://www.atsdr.cdc.gov/spl/ (accessed on 22 October 2018).

32. García-Pérez, J.; Boldo, E.; Ramis, R.; Pollán, M.; Pérez-Gómez, B.; Aragonés, N.; López-Abente, G. Description of industrial pollution in Spain. BMC Public Health 2007, 7, 40.

33. Wasana, H.M.; Perera, G.D.; Gunawardena, P.S.; Fernando, P.S.; Bandara, J. WHO water quality standards vs. Synergic effect(s) of fluoride, heavy metals and hardness in drinking water on kidney tissues. Sci. Rep. 2017, 7, 42516.

Chapter 7

Discussion

In this section, a joint discussion of the results shown in Chapters 3, 4, 5 and 6 will be conducted. First, considerations arising after the development of this study will be addressed, focusing mainly on RDs-related problems as well as on the hypothesis stated. Next, the results of each specific objective will be discussed. They will be followed by acknowledgment of the limitations found during the development of the thesis. Lastly, new research lines to orientate further studies will be drafted.

7.1 General considerations

The assessment of mortality has a great importance and impact in Public Health. Over the last decades, mortality has considerably decreased in developed countries at all levels and most of deceases are attributed to preventable conditions as cardiovascular diseases, respiratory diseases and cancer [Shofield *et al* 1991, Canudas-Romo *et al* 2008, Borrell and Joseph 2012, Debón-Aucejo *et al* 2012]. Such decrease has a lot to do with the role of preventive medicine and with action plans focused on the improvement of the health of populations [Rose *et al* 2008].

In the case of RDs, mortality is still high in most cases if compared with general mortality [Tan *et al* 2018]. These diseases have unrelated clinical, etiological, and physio-pathological characteristics and they evolve differently in affected patients. As a result, making joint action plans for RDs is more difficult, even though these disorders affect more than 3 million people in Spain, 40 million in Europe and about 400 million people around the world [de Vrueh *et al* 2013]. RD research presents an additional challenge, thus requiring a multidisciplinary approach [Posada *et al* 2008]. The inherent features of these diseases, such as the low number of cases, make epidemiological and geographic analyses harder when seeking out disparities in its indicators. Considering such disadvantageous situation, Valdez *et al* (2016) points out two different reasons that attempt to explain the lower number of studies of RDs in comparison with more frequent diseases: (1) patients are geographically dispersed, although they amount to a considerable quantity, and (2) the approaches based on the identification and removal of risk factors are not normally focused on RDs since most rare conditions have a genetic or congenital origin, which is most often irremovable.

This doctoral thesis has achieved the task of improving scientific understanding about RDs through the monitoring of mortality of some of them. Through the comprehensive combination of geographical and epidemiological methodologies, knowledge and awareness about mortality due to RDs has been increased. This first step is a valuable source of information that can help the Public Health workers in their task of reducing the mortality

attributed to these diseases. The contributions of this thesis to the existing knowledge on the subject are listed below according to the results obtained:

(1) We have presented RDs current problems by using a dual methodological perspective based on the combination of epidemiological analyses with GIS.
(2) The level of awareness about temporal trends for the selected RDs has been largely improved. The long-term monitoring across three decades is a strength of this thesis, since it provides an insight on deaths attributed to these diseases.
(3) We have been able to work with RDs epidemiological data at an innovative and highly detailed spatial level in Spain. Up to now, no population-based mortality studies at a national level for the RDs selected had been developed, and we have achieved results at districts (*comarcas)* level. We have even been able to increase the level of detail up to municipalities in some instances.
(4) Finally, by using the E-PRTR Registry, we have studied one of the plausible causes relating mortality of a selected RD and pollutant emissions into river basins. In so doing, we have introduced a broad field for future research about environmental-related studies of many RDs using epidemiological and geographic tools.

The results obtained verify the hypothesis proposed for this doctoral thesis. Thanks to the analyses we have carried out, differences in mortality attributed to HD, GPA and MND have been demonstrated at a temporal and geographical level. We have found geographical variability in mortality attributed to RDs despite the similar socio-economic structure and the existence of a National Health System across the country. In regards to the time-trend aspect, we can state a decreasing trend in mortality rates due to GPA, as we proposed in the hypothesis, but we have found the reverse trend for HD. The valuable information about geographic variability and temporal trends achieved in this doctoral thesis responds to the recent growth of mortality studies based on spatial comparisons of geographic units. These geographic mortality analyses together with temporal analyses, make the identification of inequalities easier in order to implement interventions to reduce mortality rates.

In order to discuss the outcomes obtained in greater detail, it is necessary to disaggregate the specific objectives that were proposed in Chapter 1.

7.2 Difficulties in the geographical and epidemiological analysis of RDs

Objective 1. *To identify the current problematic in the choice of geographic units of representation when working with RDs and to provide recommendations for the election of the most appropriate unit so as to obtain a better epidemiological and cartographic result.*

This objective has been mostly tackled in Chapter 3 as way of introduction to the issues arising in the RDs data processing.

The very definition of RDs implies a low number of cases and this leads to the need to aggregate individual data into geographic units due to privacy and confidentiality issues, as well as to the achievement of most powerful statistics outputs. In this sense, GITs and its application in health studies are shown as useful tools providing valuable and reliable results that can be focused on the observation of inequalities in RDs mortality. However, working with RDs data also adds complexities and issues that must be addressed from a geographical point of view. Such is the case of the MAUP and the different statistics results in each geographic unit analyzed.

Considering the need for data aggregation a starting point, identifying the number of cases and their geographical distribution was the first criteria of this study to choose the most adequate spatial unit of analysis for a specific disease.

The low number of cases in some specific areas and the use of short periods of time for the analysis have resulted in the excess of zeros and outlier results, something quite common in the epidemiological calculations for rare conditions [Arab 2015]. The challenge of choosing the most appropriate geographic unit with these three RDs has made us face a dilemma: either losing spatial detail but obtaining more inequalities in mortality, or achieving greater detail at risk of worse results in terms of mortality inequality. In SMR calculation for HD (the example of RD analyzed in the Chapter 3), municipalities, districts and provinces showed variability across the territory: the greater the level of spatial aggregation the lower the number of entities registering zero cases. In other words, when data are grouped in small geographic units (in size and population) results will be more unstable [Ugarte *et al* 2006].

Such unstable results are also visible in the SMRs smoothing process. This type of indicator is important in the work of spatial epidemiology of a disease since it addresses the spatial dependency, which is a combination of the value of the spatial unit and the adjacent values [Lawson 2003 and 2006]. It is possible to avoid general results where the less populated geographic units are artificially dominating the general pattern by smoothing the death

data, thus creating more accurate risk estimation [Esnaola *et al* 2009]. After this process, this study can add some consecutive considerations in regards to our available geographic units:

(1) In the Spanish case, the average population of municipalities is smaller in the Northern half of the country, which entails a lower probability of RDs-related deaths. The less populated geographic units that register one or more cases have obtained an extreme value after the SMR calculation.
(2) For smoothing process purposes, the results in SMR of the geographic unit will be blurred since many neighbor entities do not have cases.
(3) The elimination of extreme values caused by the smoothing process leads to a loss of variability in the results of mortality indicators, especially at a municipal level.

In this doctoral thesis, MND was the RDs group (with its own code in ICD-10) accounting for the largest number of deaths in Spain. In this 10-year study period of these diseases, we have collected more than 9,000 cases in contrast to the other RDs (i.e. 500 GPA, 1,000 HD cases respectively), in spite of being based on longer study periods. Due to this disparity in the number of cases, the task of working with higher level of detail (municipalities) was chosen for MND while districts were picked for GPA and HD. This might obviously vary in terms of the number of cases of each disease. An aggrupation process of diseases might be even carried out when they share similar characteristics in order to increase the number of cases leading to analysis at a higher level of detail.

Working with RDs data is an additional challenge for researchers in the fields of epidemiology and geography. The election of the geographic unit of analysis should cater as much information as possible. In our study, variability in mortality indicators and spatial detail among the available aggregation units has been used for this purpose. This justifies our seeking for information that can be used by Public Health officers and policy makers to benefit society at large.

7.3 Time trends in mortality

Objective 2. *To evaluate the time trends on mortality attributed to RDs in Spain over three decades.*

The objectives proposed in this section have been developed in Chapters 4 and 5. By using rates and age at death, we have been able to analyze temporal trends of mortality, which has allowed us to find out if health policies have contributed to improve the health of the population affected. As a consequence, there is a great interest in understanding the longest possible extend of time trends. Thanks to our second objective, we have been able to deepen our knowledge on mortality trends for two RDs during 3 decades. Bearing in mind the different characteristics of each RD, we have observed different trends for HD and GPA.

The results obtained for HD can be summarized as a continuous rise in mortality during the period of study. Surprisingly, we have observed two trends for GPA disease: an initial period of growth and a later decline in mortality rates. While age-adjusted mortality rate due to HD experienced an annual significant increase of 3.44% during the whole period, two opposite trends were revealed in GPA disease: An annual increase of 20.58% between 1984 and 1992 and a 2.13% annual decline from 1992 onwards. In both diseases, similar age-adjusted mortality rates have been observed for men and women, the rate being slightly higher for men in HD. In addition to mortality rates trends, we have also assessed the evolution of average age at death for each disease. We have observed an increment in the average age at death due to HD from 45 to 62 years (0.59% per year) and also a significant increase for GPA during the analyzed period of time (0.78% per year).

These two RDs should be discussed separately in order to understand the trends in mortality rates and in the average age at death. In the case of HD, it is known that there are no effective treatments to stop the progression of the disease, but only palliative care to alleviate symptoms in patients. When diagnosed with HD, palliate care allow patients to have a life expectancy of 15 to 20 years [Foroud *et al* 1999, Solberg *et al* 2018]. Our study has only taken into account the mortality attributed to HD as underlying cause of death, which means that we could not consider other deaths associated to this condition but not recorded as such–suicide being one of those cases [Farrer 1984, Walker 2007, Cozaru *et al* 2016]. HD is not likely to be codified as underlying cause of death in the case of suicide of affected patients, which might result in an underestimation and even a decrease in HD mortality rates. However, we have not observed this trend as rate continues growing. It is also important to mention that a predictive test is available since the 1990s in order to make family planning easier [Rivera-Navarro *et al* 2015], which should have led to a decrease in mortality rates in the following decades. This having been said, some patients

that may develop HD reject to carry out the aforementioned predictive test due to varied range of reasons. Over the past years, the increasing mortality due to HD presented in our study makes us infer that Public Health services should make a greater investment in research, treatments and genetic counselling to help patients and families.

In the case of GPA disease, treatments have greatly improved during the past two decades. Easier diagnosis thanks to the ANCA test and the availability of drugs (cyclophosphamide, glucocorticoids and rituximab) to alleviate damage in organs affected can be listed among the reasons for this improvement. Such reasons have led to increase GPA curation and mortality rates [Phillip and Luqmani 2008]. The existing literature highlights a rise in the GPA incidence [Lutalo and D'Cruz 2014]. This means that more patients are diagnosed, which is not reflected in mortality rates as there is a delay in the age at death of affected patients.

Considering the results obtained in our study, we can assert that trends in mortality rates due to the RDs analyzed and the increase on age at death have changed in each disease thanks to improvements in diagnosis and drug and healthcare treatments. In any case, the mortality trends observed through the temporal analyses are a reflection of many factors that should be taken into account for proper interpretation.

7.4 Geographic patterns in mortality

Objective 3. *To identify geographic patterns in mortality attributed to RDs in order to observe variations in mortality among spatial units through the Spanish territory.*

This objective has been developed in Chapters 4, 5 and 6 of this doctoral thesis and it is the most transversal one as it analyzed the three RDs selected for our study. Mortality analyses at a geographic level are a good source to find health variations in the territory as it can be used to identify the areas requiring more intervention to reduce mortality. In the same manner, the geographic analysis units for epidemiological purposes have been chosen depending on the features of the available raw mortality data. For the first time in Spain, we have improved the geographical knowledge on the mortality of the three RDs at a high level of detail. In other words, we have carried out the analysis of HD at a district level, that of MND at municipal level and GPA at both district and municipal level. Such detailed analysis has helped us to prove that the three RDs present an important variability in mortality across the national territory. The nature of these diseases should be considered when analyzing the geographic differences observed as HD is classified as genetic whereas GPA disease and MND are non-genetic. Although in the later genetic susceptibility is

involved at a small percentage [Knight *et al* 2008, Kiernan *et al* 2011], the etiology remains unknown in the rest of cases.

The SMR results for HD were the most heterogeneous of the three diseases studied in this doctoral thesis. An unevenly distributed significant increased death risk was found across the Spanish territory, whereas significant lower death risk was found in some districts located in the center of the peninsula as well as in some districts located along the coast. When considering the most populated cities, Barcelona Metropolitan Area (MA) shows a significantly higher death risk than other densely populated areas such as Madrid or Valencia MAs. The smoothing process has led to reduce the SMR extreme values when taking into account neighboring geographic units. We have observed a similar pattern in both the Smoothed-SMR for HD and the SMR. Some districts are highlighted throughout the country but a geographical pattern cannot be established. In terms of the mortality pattern of HD, the great majority of cases have a genetic origin and new cases with no family background are uncommon [Rivera-Navarro *et al* 2015]. Interest in researching geographic patterns on HD focuses on the founder effect. The founder effect refers to regions or areas of concentration where it is possible to find small gene pool. In such areas, the mutated gene responsible for HD is predominantly present among the population [Provine 2004]. Geographically, a founder effect for HD has been revealed in other locations around the world such as Tasmania or around the Maracaibo Lake in Venezuela [Young 1986, Pridmore 1990, Paradisi *et al* 2008]. Thus, a cluster of cases due to a founder effect cannot be affirmed to exist in Spain according to our results. However, we can consider isolated areas with higher and lower death risk.

In the case of mortality attributed to MND, a more detailed comparison among geographic units has been achieved as we could analyze mortality at municipal level. Overall, we have observed less heterogeneity in the variation of SMR results in comparison to the other two diseases. Municipalities with significant lower death risk were found in the Southern part of Spain, while higher risks were observed across the country, but especially in the Northern part. A closer look allowed us to find a significantly higher death risk in Madrid, Valencia and Bilbao, while the risk was lower than expected in Seville. When comparing our results of SMR attributed to MND with the geographic patterns in other studies, several authors pointed out similar geographic trends in countries in the Northern hemisphere since incidence or mortality rates are higher in Northern areas [Noonan *et al* 2005, Fang *et al* 2009,]. Also, it is worth pointing out that Santartún *et al* (2016) found that MND mortality rates in Northern Spanish provinces were doubled than in the Southern ones. This kind of results leads us to think that mortality rates might be determined by a North-South geographical gradient, implying higher rates in the North while lower in the South.

However, Logroscino *et al* (2008) stated that such gradient cannot be a reason thorough enough to explain these differences to this day. In the case of our study, although most municipalities with higher SMR are in the North, it is not the scientific community possible to assert that a latitudinal gradient explains it since most municipalities explored do not present significant results. Still, the result obtained gives the scientific community clues about geographic patterns that are in line with other studies about MND mortality performed in Spain [Veiga-Cabo *et al* 1997, Alonso *et al* 2011, Santartún *et al* 2016]. In this sense, we can state that the geographical location of municipalities is not enough to explain mortality distribution. However, the possibility of an environmental risk in affected municipalities may provide another clue about the causes of MND mortality (see Objective 4).

The geographic analysis for GPA mortality has been addressed from two points of view. On the one hand, we have analyzed the distribution of mortality across the territory. On the other, the distribution of mortality has been related to the categorization of municipalities according to socio-demographic criteria and to latitude. According to the maps, the distribution of GPA-attributed mortality is heterogeneous. We have found a lower risk in districts located in the North and in the Southwest of Spain, while higher risks were observed in districts in the Center and Northeast. In that which concerns the most populated areas, Madrid and Barcelona MAs show a significantly lower risk, while Bilbao shows a higher risk. The study of the municipalities according to socio-demographic criteria (population and PWPS) shows that agro-urban municipalities have the greatest percentage of higher than expected SMR (almost the 5%) if compared to the other two categories. In terms of geographic location, we analyzed SMR results by grouping the deaths of municipalities located in the same horizontal strips. Even though the indicator was higher than 1 in the Northern strip, the results obtained were not significant. However, SMR were significantly higher for both sexes and for men in the Southern strip of Spain (meaning the South of Andalusia). The geographic focus of GPA mortality has been addressed from a categorization of municipalities (urbanity criteria and latitudinal location of municipalities) and at a district level for Smoothed-SMR. GPA mortality was also analyzed according to population criteria and percentage of workers in the primary sector in municipalities. For this purpose, we have performed a more complex methodology to characterize urban and rural municipalities, adding to these an extra category (agro-urban). These small cities, where the PWPS are still relevant, offer a challenge for future investigations in reference to risk factors associated to economical activities such as agriculture (pesticides) to be studied at a local level. In this sense, there have not been carried out studies to compare with our results at a national level. It was only reported a study in the Northwest of Spain showing a higher GPA incidence in urban areas, though classification criteria were not specified

[González-Gay *et al* 2003]. In addition, a comparison with prevalence data between urban and rural areas carried out in Germany also shows higher indicators for urban areas [Reinhold-Keller *et al* 2000]. Despite the scarce attention paid to rural-urban comparisons by RDs researchers, studies analyzing health outcomes in common diseases are becoming more frequent [Cyril *et al* 2013]. This growing interest might be explained by the increase on urbanization and chronic diseases associated to noise or air pollution [Allender *et al* 2008]. Moreover, several authors point out that GPA incidence in the North of Europe is higher than in the South. Unfortunately, examining this pattern by using mortality indicators is not possible due to a lack of studies. We expected a higher mortality rate for the Northern latitude strip in Spain, between 43º and 44ºN (corresponding to the municipalities approximately comprised in the regions in the North of the peninsula). In studies carried out for determining causes of GPA, vitamin D or the presence of PTPN22 R620W polymorphism might be linked with higher or lower GPA incidence [Gatenby *et al* 2009, Catanoso *et al* 2014].

This objective shows the variability in mortality existing in the Spanish territory regardless of the geographic unit chosen for each RD. However, we have not been able to establish a clear pattern for them. Explaining those disparities in mortality is a very complex task and as they can be attributed to differences in prevalence, disease characteristics, environmental factors or differences in regional health systems. For all this reasons, it seems difficult to compare the mortality among the three RDs analyzed in the study, even if using the same level of aggregation for HD and GPA. The greater level of detail in MND (municipalities) makes the comparison with the other two diseases even harder. In any case, the acknowledgement of the spatial distribution of RDs is a first step to reveal the results to the scientific community, which will mean a further exploration of the possible connection between RDs and environmental factors.

7.5 Associations between mortality and environmental factors

Objective 4. *To explore the association between RDs mortality and pollutant emissions using the existing scientific literature as a starting point.*

This last objective has been wholly developed in Chapter 6. We have focused on the geographic study of pollutant emissions and their possible connection with the mortality of a specific disease. We chose a group of RDs with unknown etiology and high suspicion of being influenced by environmental factors. MND is a group of RDs with suggested risk factors that have been analyzed in many studies [Ingre *et al* 2015, Oskarsson *et al* 2015], so it is a challenge for the research community to shed light on its etiology.

The results shown in Chapter 6 can be divided into two: (1) the description of heavy metals emitted into rivers, and (2) an exploration of associations between those heavy metals emitted and MND mortality. Concerning the heavy metals emitted between 2007 and 2016, we have verified that important quantities exceeding the thresholds set by the EU regulations were emitted into water [European Parliament 2006]. We have located these sites geographically and listed the activity they develop. The sites are mostly located in the regions of the Basque Country and Catalonia. The rivers we considered affected are therefore located in those regions, although there are other main rivers in the rest of the country have also been analyzed, such as Duero, Guadalquivir or Jucar. According to the kind of metal separately, zinc was the most common hazard substance, followed by cooper.

Concentration of heavy metals in the human body is involved in neurological disorders as MND. The possible effect of metals on ALS (the most common variant of MND) has been analyzed in numerous case-control studies, finding higher concentrations in patients with ALS in many occasions [Roos 2013]. Therefore, the study of potential risks and effects on health is a significant step towards the understanding of the disease and its possible triggering. Considering the association between mortality attributed to MND and heavy metals, our findings show that in the Spanish municipalities exposed, the mortality risk is higher than in non-exposed municipalities. In fact, the differences observed in the SMR indicator have been significantly higher when taking into account the heavy metals in water (considered both as a whole and separately). If compared to non-exposed municipalities, the increased risk in MND varied from 12.4% in mercury-exposed municipalities to 20.9% in lead-exposed municipalities. Although it was not by the same emission environment, a positive correlation between lead air levels and MND mortality was also found in another study carried out in Spain [Santartún *et al* 2016]. Though these results point at the possible influence of lead in higher MND mortality rates, it is not enough to infer a clear causal link. However, it has a great value in orienting researchers and policy makers towards possible health strategies.

The exploration of the connection between environmental hazards and the health of populations is a topic of growing interest. In this sense, there has traditionally been a worldwide effort to understand the burden disease using mortality indicators [GBD 2017 Mortality Collaborators 2018]. In many cases, these studies were aimed at finding out the effects of pollution or other environmental factors on health. At the same time, the burden disease attributable to hazards was analyzed by classing and ordering it according to an importance criteria [Briggs 2003]. Despite this, the difficulty in the quantification of the number of pollutants as well as how they affect health must be highlighted. Moreover, the minimum dose required to trigger disease symptoms and the time of exposure are also

difficult to assess [Järup 2003, Tchounwou *et al* 2012]. Increasing the knowledge on disease symptoms and on how the accumulation process of the possible hazard substances in the human body must be a priority for researchers and policy makers.

It is remarkable that studies relating environmental factors with other RDs are scarce to date. More than 80% of RDs have been attributed to genetic origins, but there is also a considerable quantity of them that are related to infections, allergies, degenerative, proliferative or teratogenic factors which have not been ascertained to date [EURORDIS 2005]. In that respect, apart from the RDs considered in this doctoral thesis, other studies to ascertain risk factors in other RDs whose origin are not genetic have been carried out. For instance, in the group of RDs known as pemphigus, there are studies that associate them with the exposure to chemical substances as pesticides or to workers that receive ultraviolet radiation [Wohl and Brenner 2003, Tirado-Sánchez *et al* 2006]. Another example would be the location of scleroderma patient clusters in several European countries which could be determined by an environmental factor, although it is still unknown [Mayes 2003].

Finally, geographical analysis might play a decisive role in research works seeking answers to questions involving places and RDs. The use of mortality as a reliable and trustable indicator offers many possibilities for the inclusion of external variables to obtain outcomes, leading to a better understanding of the causes of a disease. At the same time, environmental pollution studies are needed to raise awareness about associated risks, which would be achieved by creating and updating indicators. Such environmental indicators might enable a connection with health inequalities. If the influence of hazards in some diseases is proven, the risks and effects on health caused by them may be totally or partially avoidable since the solution would be prevention rather than using technologies or drugs in affected patients.

7.6 Limitations of the study

Our study presents the mortality attributed to the selected RDs, so we have only used the underlying cause of death included in the death registry. Our results may be enriched by considering the secondary cause of death from the death registry, or even by including other causes of death from patients affected by these RDs in our study. RDs visibility in information systems might be improved thanks to the use of more variables when carrying out an analysis. In some occasions, RD patients have another disease codified as the underlying cause of death when the die, the RD being considered the second cause. Such changes in codification might cause an underestimation of the number of diseases and, therefore, of mortality rates. Moreover, we can also point out three specific limitations to

our study connected to temporal and geographical analyses and to environmental associations.

The first limitation might affect time trend analyses of mortality attributed to HD and GPA. Some researchers have investigated the change in ICD codification between the ninth and tenth edition in 1999, and how this might affect the identification of these diseases. Cano-Serral *et al* (2006) analyzed a great amount of diseases to verify the level of agreement between both ICD versions, but their study did not verify our RDs selected. Despite this change in codification –as well as the rates and number of cases identified around the year of change–, trends do not show sharp shifts neither for HD nor for GPA between the years 1998 and 1999 in Spain.

The second of the limitations concerns the results at a geographical level. In this sense, death records follow a homogeneous standardized methodology throughout the national territory, but variations in death registries of each AC may happen. This is due to the fact that it is difficult to find the same level of accuracy in every death certificate [Ragonese *et al* 2004]. The NSI unifies and publishes all of them in a unique National Death Registry, even though regional governments are responsible of carrying out the first work of codification as well as of including the variables in the registry. This may produce slight differences among the certificates recorded by each region. Either way, the implementation of an automatic codification system in 2014 may help to increase the precision and the homogeneity throughout the country [Cirera-Suárez 2018].

Finally, in regards to the explorations of mortality associated with environmental factors, two issues should be mentioned. One of them refers to the limited information provided by the NSI death registry. Death certificates include scarce socio-economic information of individuals due to confidentiality and privacy issues [Kulhánová *et al* 2014]. Although patient registries, hospital data sources or surveys might include such socio-economic information, they are not national population-based registries, being this a disadvantage. Furthermore, the ecological fallacy may appear in association analysis when working with aggregated data. As previously discussed, this problem arises when associations that are meant to be found at individual level are found at other aggregation level instead [O'Sullivan and Unwin 2010]. In our association analyses, external variables have been taken into account, such as the categorization of municipalities in terms of urbanity vs rurality for GPA disease, or the exposure to heavy metal emissions according to municipality of death for MND. Even though we have found a positive association between metals and MND mortality, this does not mean that this association also exists at an individual level. We should be very careful when applying the results of individuals to the whole group, regardless the exposure of each patient. The ecological fallacy is an inherent

problem to ecological analysis with aggregated data. However, and even though it does not imply that results are not useful, the extrapolation to other levels must be cautiously done.

7.7 Next steps and future research lines

The research line of our study might continue in two directions in the immediate future. The first one would focus on deepening our knowledge on the plausible associations between low frequent diseases and environmental factors by adding more RDs to the study. The second line of research would imply the extension of the geographical framework to a European level. What is more, there are thousands of RDs which have not yet been analyzed at a geographic or temporal level in Spain. Increasing knowledge on these diseases (by calculating or updating epidemiological indicators) is crucial to understand the disease burden, to draft healthcare actions and to support investment for future treatments and research. Our research group (IIER-ISCIII and the University of Alcalá) is working in the elaboration of the *National Atlas of Mortality due to Rare Diseases*, which contains geographic and temporal information of groups of RDs, as well as specific examples of RDs. This Atlas, which will be published soon, will consist of a combination of explanatory texts, graphics with epidemiological results and maps. This work will enhance the visibility of RDs as it will be a rich source of information for affected patients, scientists and the society as a whole.

Furthermore, by including environmental factors in our research we have opened a relatively unexplored but broad field of research, can be seen as a complement to geographic and temporal analyses. Such plausible influences have a key role in the onset or development of non-genetic RDs. We have carried out an approach to the relationship between MND mortality and proximity to heavy metal emissions. However, investigations focused on seeking some RDs triggering factors may apply to a wide range of possible causes such as occupational risks (related with high linesmen and electricians, agriculture workers) or personal habits (smoking, low premorbid body mass index and a high fat intake) [Huisman 2015, Oskarsson *et al* 2015]. Some RDs have a high number of cases and a wider social repercussion, so analyzing environmental factors as a trigger is a challenge. Another important point is the role of data obtained by using remote sensing systems, which may provide interesting information to better understand health-related issues. Remote sensing data is used in a great deal of research focusing on infectious diseases such as malaria, dengue or Ebola [Stanforth *et al* 2016, Li 2016, Peckham and Sinha 2017]. Although the characteristics of rare conditions are not necessarily shared with the aforementioned diseases, they could also benefit from remote sensing tools aimed at

studying variables about possible environmental factors. For instance, an assessment of exposure to air pollution over large areas may be carried out by using different types of methodologies such as proximity-based assessments, line dispersion models or statistical interpolation between sparse monitoring networks [Sorek-Hamer *et al* 2017]. Thus, remote sensing methodologies are of support in both epidemiological and burden of diseases studies. Efforts from other research fields related to geography are needed to get new clues about RDs. Geographical analysis, along with its associated methodologies and tools, will play a key role in many of these investigations.

Finally, and as a proposal for a future research line, mortality studies not only focused on Spain but rather on the European continent would be a topic of keen interest. Studies at supranational level would be desirable by using harmonized mortality data both in time and codification in order to study health inequalities. These studies are required as many authors point out that incidence rates of many RDs are higher in the Northern countries of Europe than in the South. Using the data provided by the WHO, the IIER research group has started a line of temporal and geographic supranational studies about RDs mortality such as hereditary ataxia, MND and HD at a European level [Arias-Merino *et al* 2017]. Moreover, collecting data at municipal level could be especially valuable to aggregate them in administrative units with the same spatial detail –according to the availability of each country. For instance, Eurostat office has created subdivisions with different hierarchical-aggregated levels for all member countries within the EU. They are called Nomenclature of Territorial Units for Statistics (NUTS) and they are regulated by several thresholds of population [Eurostat 2018]. The NUTS2 or NUTS3 levels do not have such resolution as the district or the municipality unit, but the observation of geographic patterns thanks to epidemiological analyses may serve as a basis for further study in the geographic field focusing on specific areas.

Chapter 8

Conclusions

The aim of this doctoral thesis was to enhance the knowledge of mortality attributed to RDs. In order to do so, Geographic Information Systems and epidemiological analysis tools have been used in an integrating manner. This chapter presents the conclusion of the objectives stated in the introduction, each four of them having the purpose of verifying different elements of our hypothesis.

Objective 1: *To identify the current problematic in the choice of geographic units of representation when working with RDs and to provide recommendations for the election of the most appropriate unit so as to obtain a better epidemiological and cartographic result.*

The MAUP-related issue is present in RDs analysis and their mortality data. More specifically, the scale effect is visible when choosing different geographic units. Thus, the election of spatial aggregation level depends on the number of cases of the studied variable, as well as on the topological characteristics of the geographic units when carrying out a smoothing spatial process.

In the Spanish case, districts (*comarcas*) seem to be the most advantageous geographic unit for RDs analysis, as it achieves a balance between the number of inhabitants in each entity, the number of neighbor entities (used for further smoothing process), the detail in spatial information and the number of deaths per entity.

Objective 2: *To evaluate the time trends on mortality attributed to RDs in Spain over three decades.*

Different time trends in mortality for the two selected RDs have been observed. Such differences result from the characteristics of each disease: diagnosis, evolution and severity, as well as drug therapy and healthcare.

Mortality rates due to HD have increased over the last 30 years in spite of the existence of an available genetic test that was expected to help decrease this trend.

Mortality rates due to GPA experienced a sharp increment until 1992, and a slight decrease from then on.

An increment in the average age at death for both diseases has been proven, which could be related to better treatments (in the case of GPA), or to the ageing in the general population (in the case of HD).

Objective 3: *To identify geographic patterns in mortality attributed to RDs in order to observe variations in mortality among spatial units through the Spanish territory.*

Explaining geographic patterns of the selected RDs has been hard. However, we have been successful in identifying spatial inequalities in the death risk attributed to them.

Spatial variability in mortality due to HD, GPA and MND has been observed across the Spanish territory. Mortality varies according to the characteristics of the diseases analyzed, to differences in prevalence and to patient healthcare.

Variation in the distribution of death rates has been observed for GPA disease according to the latitude of municipalities, to urban vs agro-urban and rural classification, as well as to district analysis. However, we did not find similar results in other countries so far.

Although the findings in this thesis about the geographic distribution of HD show differences within the country, they are not thorough enough to establish a cluster of cases based on genetic origin.

A spatial pattern in death risk due to MND has been observed at a municipal level: municipalities with significantly increased death risk are mostly located in the North, while it is significantly lower in the Southern ones.

Objective 4: *To explore the association between RDs mortality and pollutant emissions using the existing scientific literature as a starting point.*

E-PRTR Registry is presented as a powerful open-and-free data source about pollutant emissions in every European country, offering plenty of possibilities in geographical and epidemiological studies.

Industrial complexes emitting heavy metals to river basins are mostly located in the North of Spain. They present different activities and in some cases they exceed significantly the defined threshold.

By associating MND and heavy metal exposure, death risk is proven to be higher in the exposed areas than in the non-exposed ones. Such increase takes place both at a general

level and once the type of metal is analyzed, being lead and zinc the ones that show higher increased risks.

Final considerations

The above stated conclusions show that this thesis enhances knowledge on mortality attributed to RDs. Geography has been proved to be an essential discipline to understand the distribution of RDs mortality, despite the difficulties arising from the low number of cases and the necessity of spatial aggregation because of privacy and confidentiality issues. Moreover, GIS are an effective tool in epidemiological studies to seek environmental factors that could be related with the RDs etiology.

Focusing on mortality and from a global perspective, epidemiological calculations and geographic analyses have increased current knowledge on GPA disease, HD and MND, although further and deeper investigations are necessary to explain the inequalities and variations observed. The relationship between risk factors and RDs has been analyzed from a geographical approach, which may give clues about the causes of these diseases. The results obtained might be useful to develop preventive measures that are aimed at minimizing risks.

The value of this PhD dissertation lies in the possibility of replicating the study in other RDs. The results presented can be very useful in implementing prospective health policies aimed at decreasing the impact of these rare conditions in Spain.

Bibliography

Allender S, Foster C, Hutchinson L & Arambepola C. (2008). Quantification of urbanization in relation to chronic diseases in developing countries: A systematic review. *J Urban Health, 85*(6): 938-951.

Alonso V, Villaverde-Hueso A, Hens M, Morales-Piga A, Abaitua I & de la Paz MP. (2011). Increase in motor neuron disease mortality in Spain: Temporal and geographical analysis (1990-2015). *Amyotroph Lateral Scler, 12*(3): 192-198.

Alonso V, Sánchez-Díaz G, Villaverde-Hueso A, de la Paz MP & Bermejo-Sánchez E. (2018). A nationwide registry-based study on mortality due to rare congenital anomalies. *Int J Environ Res Public Health, 15*(8): pii:E1715.

Anselin L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science, 89*: 3-25.

Anthamatten P & Hazen H. (2011). *An Introduction to the Geography of Health, first edition*. Routledge, London.

Arab A. Spatial and spatio-temporal models for modelling epidemiological data with excess zeros. (2015). *Int J Environ Res Public Health, 12*(9): 10536-10548.

Aránguez-Ruiz E, Soto MJ, Iriso A, de Paz C & García JF. (2005). Sistema de información geográfica y salud pública en la Comunidad de Madrid. *Serie Geográfica, 12*: 137-146.

Aránguez-Ruiz E, Arribas M, Aránguez J & Ordoñez JM. (2012). *Salud y territorio: Aplicaciones prácticas de los Sistemas de Información Geográfica para la salud ambiental*. Sociedad Española de Sanidad Ambiental, Madrid.

Arbia G. (1989). *Spatial data configuration in statistical analysis of regional economic and related problems*. Kluwer, Dordrecht.

Arias-Merino G, Sánchez-Díaz G, Villaverde-Hueso A, Posada de la Paz M & Alonso-Ferreira V. (2017). Mortality statistics and their contribution to improving the knowledge of rare diseases epidemiology: The example of Hereditary Ataxia in Europe. In: Posada de la Paz M, Taruscio D & Groft S. (Eds.) *Rare Diseases Epidemiology: Update and Overview*. Springer, pp. 521-533.

Arias-Merino G. (2017). *Mortalidad debida a enfermedades raras en Europa: enfermedad de Huntington, ataxias hereditarias y enfermedades de la motoneurona*. PhD Thesis. Universidad Complutense, Spain.

Arizo-Luque V, Botella P, Abellán C, Martínez-Beneito MA, García-Blasco MJ, Amador-Iscla A, Bosch S & Zurriaga O. (2005). *Aproximación a las enfermedades raras en la Comunidad Valenciana, periodo 1999-2003*. Conselleria de Sanitat, Valencia.

Aymé S, Bellet B & Rath A. (2015). Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. *Orphanet J Rare Dis, 10*: 35.

Bates GP, Dorsey R & Gusella JF. (2015). Huntington disease. *Nat Rev Dis Primers, 1*: 15005.

Beale L. (2016). Effective use of GIS for spatial epidemiology. In: Kanaroglou P & Delmelle E. (Eds.) *Spatial Analysis in Health Geography*. Routledge, pp. 15-31.

Benach J, Yasui Y, Borrell C, Rosa E, Pasarín MI, Benach N, Español E, Martínez JM, Daponte A. (2001). *Atlas de mortalidad en áreas pequeñas en España 1987-1995*. Universitat Pompeu Fabra, Barcelona.

Benach J, Martínez JM, Borrell C, Pasarín MI, Yasui Y, Vergara M, Buxó M, Muntaner C, Daponte A & Ocaña R. (2007). *Estudio geográfico de la mortalidad en España, Análisis de tendencias temporales en municipios o agregados de municipios*. Fundación BBVA, Bilbao.

Benach J, Martínez JM, Martín JC, Box M, Vergara M, Yasui Y, Cleries R, Borrel C, Español E & grupo MEDEA. (2013). *Atlas de mortalidad en municipios y unidades censales de España (1984-2004)*. Fundación BBVA, Bilbao.

Besag JE, York J & Molliè A. (1991). A Bayesian image restoration with two applications in spatial statistics. *Ann Inst Statist Math, 43*: 1–59.

Boelaert M, Arbyn M & Van der Stuyft P. (1998). Geographical information systems (GIS), gimmick or tool for health district management? *Trop Med Int Health, 3*(3): 163-165.

Bonita R, Kjellström T, Beaglehole R & World Health Organization. (2006). *Basic Epidemiology*. World Health Organization, Geneva.

Borrell LN & Joseph SP. (2012). Health disparities: The role of preventive medicine in closing the gap. *Prev Med, 55*(6): 569-570.

Bosque-Sendra J. (1997). *Sistemas de Información Geográfica*. Rialp, Madrid.

Botella-Rocamora P, Zurriaga O, Posada de la Paz M, Martínez-Beneito MA, Bel E, Robustillo A, Ramalle-Gomara E, Durán-Plá E & Sánchez-Porro P. (2006). *Atlas Nacional Provincial de Enfermedades Raras 1999-2003*. Repier, Madrid.

Bowling A. (1997). *Research methods in health: Investigating health and health services*. Open University Press, Philadelphia.

Briggs D. (2003). Environmental pollution and the global burden of disease. *Br Med Bull, 68* (1): 1–24.

Brömer R. (2000). The first global map of the distribution of human diseases: Friedrich Schnurrer's 'Charte über die geographische Ausbreitung der Krankheiten' (1827). *Medical History Supp, 20*: 176-185.

Brown T, Andrews GJ, Cummins S, Greenhough B, Lewis D & Power A. (2018). *Health Geographies: A Critical Introduction*. Wiley Blackwell, Hoboken and Chichester.

Burton I. (1963). The quantitative revolution and theoretical geography. *Canadian Geographer, 7*: 151-162.

Cano-Serral G, Perez G, Borrell C & COMPARA Group. (2006). Comparability between ICD-9 and ICD-10 for the leading 133 causes of death in Spain. *Rev Epidemiol Sante Publique, 54*(4): 355-365.

Canudas-Romo V, Glei D, Coelho E & Boe C. (2008). Mortality Changes in the Iberian Peninsula in the last decades of the Twentieth Century. *Population, 63*(2): 319-344.

Carter D. (2016). El desarrollo de la geografía médica: una reseña de tendencias actuales. *Población & Sociedad, 23*(2): 207-220.

Catanoso M, Macchioni P, Boiardi L, Manenti L, Tumiati B, Cavazza A, Luberto F, Pipitone N & Salvarani C. (2014). Epidemiology of granulomatosis with polyangiitis (Wegener's granulomatosis) in Northern Italy: A 15-year population-based study. Semin. *Arthritis Rheum, 44*: 202–207.

Cauvin C, Escobar F & Serradj A. (2010). *Thematic Cartography and Transformations, volume 1*. John Wiley & Sons, London.

Cebrecos A, Domínguez-Berjón MF, Duque I, Franco M & Escobar F. (2018). Geographic and statistic stability of deprivation aggregated measures at different spatial units in health research. *Applied Geography, 95*: 9-18.

Chiò A, Logroscino G, Traynor BJ, Collins J, Simeone JC, Goldstein LA & White LA. (2013). Global epidemiology of amyotrophic lateral sclerosis: A systematic review of the published literature. *Neuroepidemiology, 41*: 118-130.

Cirera-Suárez L. (2018). Primer bienio de estadísticas de mortalidad con el codificador automático Iris de causas de muerte. *Gac Sanit, 32*(1): 5-7.

Cozaru GC, Aşchie M, Mitroi AF, Poinăreanu I & Gorduza EV. (2016). Ethical and genetic aspects regarding presymptomatic testing for neurodegenerative diseases. *Rev Med Chir Soc Med Nat Iasi, 120*(1): 15-22.

Cromley E & McLafferty S. (2011). *GIS and Public Health, second edition*. The Guilford Press, New York.

Cruz-Coke R. (1999). Hippocratic philosophy. *Rev Med Chil, 127*(5): 611-614.

Curtin LR & Kelin RJ. (1995). Direct standardization (age-adjusted death rates). *Healthy People 2000 Stat Notes, 6*: 1-10.

Cyril S, Oldroyd JD & Renzaho A. (2013). Urbanisation, urbanicity, and health: a systematic review of the reliability and validity of urbanicity scales. *BMC Public Health, 13*: 513.

Debón-Aucejo AM, Martínez-Ruiz F & Montes F. (2012). Temporal evolution of some mortality indicators: Application to Spanish data. *North American Actuarial Journal, 16*(3): 364-367.

Decreto Legislativo 1/2006, de 27 de diciembre, del Gobierno de Aragón.

Deffontaines P. (1926). La géographie humaine de Jean Brunhes. *Annales de Géographie, 35*(195): 268-271.

Dicker RC, Coronado F, Koo D & Parrish RG. (2006). *Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics, third edition*. CDC, Atlanta.

Dorey J, Trigo P, García de Yébenes J & Toumi M. (2012). M07 Burden of Huntington's disease in Spain. *J Neurol Neurosurg Psychiatry, 83*: A48.

Dummer TJ. (2008). Health geography: supporting public health policy and planning. *CMAJ, 178*(9): 1177-1180.

Eagleson S & Escobar F. (2003). Administrative boundary design in support of SDI objectives. In: Williamson I, Rajabifard A & Fenney ME (Eds.) *Developing Spatial Data Infrastructures – From concept to reality*. CRC Press, PP. 249-263.

Eagleson S, Escobar F & Williamson I. (2010). Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. *Int J Geogr Inf Sci, 17*(2): 99-118.

Eicher CI & Brewer CA. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science, 28*(2): 125-138.

Ekestern E & Lebhart G. (2005). Long-term monitoring of the mortality trend of Huntington's disease in Austria. *Eur J Epidemiol, 20*: 169-172.

Emch M, Root E & Carrel M. (2017). *Health and medical geography, fourth edition*. Guilford Press, New York.

Escobar F, Green J, Waters E & Williamson I. (1999). Enabling Geographic Information Systems for the Public Health sector. A proposal for the research agenda. In: *Proceedings GEOMED'99, Second International Workshop on Geography and Medicine*. Elsevier, Paris.

Escobar F, Díaz-Castillo C, Salado-García MJ, Rodríguez-Espinosa VM & Gómez-Delgado M. (2005). Caracterización de los riesgos ambientales para la salud: puesta a punto de un SIG en dos distritos del sureste de Madrid. *Serie Geográfica, 12*: 11-37.

Esnaola S, Montoya I, Calvo M, Ibáñez B, Audicana C, Ruiz R, Aldasoro E & Martín U. (2009). Desigualdades socioeconómicas en la mortalidad en el País Vasco y sus capitales: un análisis de áreas geográficas pequeñas. *Estudios geográficos, 70*(267): 443-462.

Decision No 1295/1999/EC of the European Parliament and of the Council of 29 April 1999 adopting a programme of Community action on rare diseases within the framework for action in the field of public health (1999–2003). (1999). Commission of the European Communities, Brussels.

European Pollutant Release and Transfer Register (E-PRTR). Accessed: July 2019, https://prtr.eea.europa.eu/#/home

EURORDIS. (2005). *Rare Diseases: Understanding this Public Health Priority*. Accessed: July 2019, https://www.eurordis.org/sites/default/files/publications/princeps_document-EN.pdf

Eurostat & European Commission. (2009). *Health statistics – Atlas on mortality in the European Union*. Office for Official Publications of the European Communities, Luxembourg.

Eurostat (2013). *Revision of the European Standard Population – Report of Eurostat's task force*. Publications Office of the European Union, Luxembourg.

Eurostat. (2018). *Methodological manual on territorial typologies*. Publications Office of the European Union, Luxembourg.

Fang F, Valdimarsdóttir U, Bellocco R, Ronnevi LO, Sparén P, Fall K & Ye W. (2009). Amyotrophic lateral sclerosis in Sweden, 1991–2005. *Arch Neurol, 66*: 515-519.

Farrer LA. (1986). Suicide and attempted suicide in Huntington disease: implications for preclinical testing of persons at risk. *Am J Med Genet, 24*(2): 305–311.

Fernández J. (2011). *Caracterización de las Comarcas Agrarias en España*. Ministerio de Medio Ambiente y Medio Rural y Marino, Madrid.

Fletcher-Lartey SM & Caprarelli G. (2016). Application of GIS technology in public health: success and challenges. *Parasitology, 143*(4): 401-415.

Flowerdew R. (2011). How serous is the modifiable areal unit problem for analysis of English census data? *Popul Trends, 145*: 102-114.

Foroud T, Gray J, Ivashina J & Conneally PM. (1999). Differences in duration of Huntington's disease based on age at onset. *J Neurol Neurosurg Psychiatry, 66*(1): 52-56.

Garfield S. (2013). *En el mapa. De cómo el mundo adquirió su aspecto*. Taurus, Barcelona.

Gatenby PA, Kucas RM, Engelsen O, Ponsobony AL & Clements M. (2009). Antineuthrophil cytoplasmic antibody-associated vasculitis: Could geographic patterns be explained by ambient ultraviolet radiation? *Arthritis Rheum, 61*: 1417–1424.

Gatrell AC & Elliot SJ. (2014). *Geographies of Health: an Introduction*. Wiley, Hoboken.

GBD 2017 Mortality Collaborators. (2018). Global, regional, and national age-sex-specific mortality and life expectancy, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet, 392*(10159): 1684-1735.

Glover J, Tennant S & Woollacott T. (2005). The Social Health Atlas: a policy tool to describe and monitor social inequality and health inequality in Australia. *Serie Geográfica, 12*: 103-122.

Gómez-Barroso D, Prieto-Flores ME, Mellado A & Moreno A. (2015). Análisis espacial de la mortalidad por enfermedades cardiovasculares en la ciudad de Madrid, España. *Rev Esp Salud Publica, 89*: 27-37.

González-Gay MA, García-Porrua C, Guerrero J, Rodríguez-Ledo P & Llorca J. (2003). The Epidemiology of the Primary Systemic Vasculitides in Northwest Spain: Implications of the Chapel Hill Consensus Conference Definitions. *Arthritis Rheum, 49*: 388–393.

Groft SC & Posada de la Paz M. (2017). Rare Diseases: joining mainstream research and treatment based on reliable epidemiological data. In: Posada de la Paz M, Taruscio D & Groft S. (Eds.) *Rare Diseases Epidemiology: Update and Overview*. Springer, pp. 3-21.

Haining R. (2003). *Spatial data analysis – Theory and practice*. Cambridge University Press, Cambridge.

Hippocrates, Thucydides, Clifton F & of Ephesus S. (1752). *Hippocrates upon air, water, and situation: upon endemical diseases and upon pronosticks, in acute cases especially*. Printed for Whiston J & White B, London.

Huismann M. (2015). *Exogenous risk factors for amyotrophic lateral sclerosis*. PhD Thesis. Utrecht University, Netherlands.

Iglesias Díaz M. (1886). *Programa razonado de Geografía Médica de España*. Real Academia de Medicina, Madrid.

IGN. *Instituto Geográfico Nacional*. Accessed: July 2019, http://www.ign.es/web/ign/portal.

Ingre C, Roos M, Piehl F, Kamel F & Fang F. (2015). Risk factors for amyotrophic lateral sclerosis. *Clin Epidemiol, 7*: 181-193.

INE. *Instituto Nacional de Estadística. Estadística de Defunciones según Causa de Muerte*. Accessed: July 2019, http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176780&menu=metodologia&idp=1254735573175

INE. *Instituto Nacional de Estadística. Número de municipios por provincias, comunidades autónomas e islas*. Accessed: July 2019, http://www.ine.es/daco/daco42/codmun/cod_num_muni_provincia_ccaa.htm

Inskip H, Beral V, Fraser P & Haskey J. (1983). Methods for age-adjustment of rates. *Stat Med, 2*(4): 455-466.

Järup L. (2003). Hazards of heavy metal contamination. *Br Med Bull, 68*:167-182.

Jori G. (2013). El estudio de la salud y la enfermedad desde una perspectiva geográfica: temas, enfoques y métodos. *Biblio 3W: Revista bibliográfica de geografía y ciencias sociales, 28*(1029).

Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, Burrell JR & Zoing MC. (2011). Amyotrophic lateral sclerosis. *Lancet, 377*: 942-955.

Kiernan MC. (2018) Motor neuron disease in 2017: Progress towards therapy in motor neuron disease. *Nat Rev Neurol, 14*: 65-66.

Kim HJ, Fay MP, Feuer EJ & Midthune DN. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med, 19*: 335-351.

Knight A, Sandin S & Askling J. (2008). Risks and relative risks of Wegener's granulomatosis among close relatives of patients with the disease. *Arthritis Rheum, 58*(1): 302-307.

Kulhánová I, Bacigalupe A, Eikemo TA, Borrell C, Regidor E, Esnaola S, Mackenbach JP & Eurothine consortium. (2014). Why does Spain have smaller inequalities in mortality? An exploration of potential explanations. *European Journal of Public Health, 24*(3): 370–377.

Laffly D & Handschumacher P. (2005). Spatial expressions of sleeping sickness in Ivory Coast and Guinea Conakry remote sensing and spatial analysis contributions to map human African trypanosomiasis risk area. *Serie Geográfica, 12*: 83-102.

Lai PC, So FM & Chan KW. (2008). *Spatial epidemiological approaches in disease mapping and analysis*. CRC Press, Boca Raton.

Last JM. (2001). *A dictionary of epidemiology*. Oxford University Press, Oxford.

Lawson AB, Brwone WJ & Vidal-Rodeiro CL. (2003). *Disease mapping with WinBUGS and MLwiN*. John Wiley & Sons, Chichester.

Lawson AB. (2006). *Statistical methods in spatial epidemiology, second edition*. John Wiley & Sons, Chichester.

Le Cam Y. (2014). A hidden priority: the paradox of rarity (EURORDIS perspective). *Expert Opinion on Orphan Drugs, 2*(11):1123-125.

Ley 6/1987, de 4 de abril de la Organización Comarcal de Cataluña.

Li ZC, Roux E, Dessay N, Girod R, Stefani A, Nacher M, Moiret A & Seyler F. (2016). Mapping a knowledge-based Malaria hazard index related to Landscape using remote sensing: application to the cross-border area between French Guiana and Brazil. *Remote Sens, 8*(4): 319.

Liebenberg E, & Demhardt IJ. (2012). *History of Cartography: International Symposium of the ICA Commission, 2010*. Springer, New York.

Lochmüller H, Toorent I, Farnell J, Le Cam Y, Jonker AH, Lau L, Baynam G, Kaufmann P, Dawkins H, Lasko P, Austin C, Boycott K & on behalf of the IRDiRC Consortium Assembly. (2017). The International Rare Diseases Research Consortium: Policies and Guidelines to maximize impact. *Eur J Hum Genet, 25*(12): 1293-1302.

Logroscino G, Traynor BJ, Hardiman O, Chio' A, Couratier P, Mitchell JD, Swingler RJ, Beghi E & EURALS. (2008). Descriptive epidemiology of amyotrophic lateral sclerosis: new evidence and unsolved issues. *J Neurol Neurosurg Psychiatry, 79*(1): 6–11.

López-Abente G, Pollán M, Escolar A, Errezola M, Abraira V. (2001). *Atlas de mortalidad por cáncer y otras causas en España, 1978-1992*. Instituto de Salud Carlos III, Madrid.

López-Abente G, Ramis R, Pollán M, Aragonés M, Pérez-Gómez B, Gómez-Barroso D, Carrasco JM, Lope V, García-Pérez J, Boldo E & García-Mendizábal M. (2007). *Atlas municipal de mortalidad por cáncer en España, 1989-1998*. Instituto de Salud Carlos III, Madrid.

Lutalo PM & D'Cruz DP. (2014). Diagnosis and classification of granulomatosis with polyangiitis (aka Wegener's granulomatosis). *J Autoimmun, 48-49*: 94-98.

Lyseen AK, Nohr C, Sorensen EM, Gudes O, Geraghty EM, Shaw NT & Bivona-Tellez, C. (2014). A review and framework for categorizing current research and development in health related Geographic Information Systems (GIS) studies. *Yearb Med Inform, 9*: 110-124.

Mahr AD, Neogi T & Merkel PA. (2006). Epidemiology of Wegener's granulomatosis: Lessons from descriptive studies and analyses of genetic and environmental risk determinants. *Clin Exp Rheumtol, 24*(suppl 41): S82-S91.

Manley D. (2014). Scale, Aggregation, and the Modifiable Areal Unit Problem. In: Fischer M & Nijkamp P. (Eds.) *Handbook of Regional Science*. Springer, pp. 1157-1171.

Martínez-Morillo M, Grados D, Naranjo-Hans D, Mateo L, Holgado S & Olivé A. (2012). Granulomatosis con poliangeítis (Wegener). Descripción de 15 casos. *Reumatol Clin, 8*(1): 15-19.

Mayes MD. (2003). Sclerodermia epidemiology. *Rheum Dis Clin North Am, 29*(2): 239-254.

Miller RG & Appel SH. (2017). Introduction to supplement: The current status of treatment for ALS. *Amyotroph Lateral Scler Frontemproal* Degener, *18*(supp 1): 1-4.

Mira-Escolano MP, Cano F, Maceda LA, Sánchez-Escámez A, Sequer C, Serrano A, Abellán MV, Ciller P & Palomar-Rodríguez JA. (2018). *Prevalencia de enfermedades raras en la Región de Murcia 2015*. Consejería de Salud, Murcia.

Morgenstern H & Thomas D. (1993). Principles of study design in environmental epidemiology. *Environ Health Perspect, 101*(suppl 4): 23-38.

Nagy C, Juhász A, Papp Z & Beale L. (2013). Hierarchical spatio-temporal mapping of premature mortality due to alcoholic liver disease in Hungary, 2005-2010. *Eur J Public Health, 24*(5): 827-833.

Noonan CW, White MC, Thurman D & Wong LY. (2005). Temporal and geographic variation in United States motor neuron disease mortality, 1969–1998. *Neurology, 64*: 1215-1221.

Nykiforuk CI & Flaman LM. (2011). Geographic Information Systems (GIS) for health promotion and public health: A review. *Health Promot Pract, 12*(1): 63-73.

Ocaña-Riola R. (2010). Common errors in disease mapping. *Geospat Health, 4*(2): 139-154.

Olaya V. (2016). *Sistemas de Información Geográfica*. CreateSpace Independent Publishing Platform.

Olivera A. (1992). *Geografía de la Salud*. Síntesis, Madrid.

Openshaw S & Taylor PJ. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. In: Wrigley N. (Ed.) *Statistical application in the spatial sciences*. Pion, London, pp. 127-144.

Openshaw S & Taylor PJ. (1981). The modifiable areal unit problem. In: Wrigley N & Bennet RJ. (Eds.) *Quantitative Geography: a British view*. Routledge & Kegan Paul, London, pp. 60-69.

Openshaw S. (1984). *The modifiable areal unit problem*. Geo Books, Norwick.

Orphanet Report Series - *List of rare diseases and synonyms listed in alphabetical order – January 2019*. Accessed: July 2019, http://www.orpha.net/orphacom/cahiers/docs/GB/List_of_rare_diseases_in_alphabetical_order.pdf

Orth M & Schwenke C. (2011). Age-at-onset in Huntington disease. *PLoS Curr, 29*(3): RRN1258.

Oskarsson B, Horton K & Mitsumoto H. (2015). Potential Environmental Factors in Amyotrophic Lateral Sclerosis. *Neurol Clin, 33*(4): 877-888.

O'Sullivan D & Unwin D. (2010). *Geographic Information Analysis, second edition*. John Wiley & Sons, Hoboken.

Paradisi I, Hernández A & Arias S. (2008). Huntington disease mutation in Venezuela: age of onset, haplotype analyses and geographic aggregation. *J Hum Genet, 53*(2): 127-135.

Parenteau M & Sawada MC. (2011). The modifiable areal unit problem (MAUP) in the relationship between exposure to NO2 and respiratory health. *Int J Health Geog, 10*: 58.

Peckham R & Sinha R. Satellites and the new war on infection: tracking Ebola in West Africa. *Geoforum, 80*: 24–38.

Pfeiffer D, Robinson T, Stevenson M, Stevens K, Rogers D & Clements A. (2008). *Spatial analysis in Epidemiology*. Oxford University Press, Oxford.

Piantadosi S, Byar DP & Green SB. (1988). The ecological fallacy. *Am J Epidemiol, 127*(5): 893-904.

Phillip R & Luqmani R. (2008). Mortality in systemic vasculitis: a systematic review. *Clin Exp Rheumatol, 26*: S94-104.

Posada M, Izquierdo-Martínez M, Ferrari-Arroyo MJ, Avellaneda-Fernández A, Andrés-Copa P & Martín-Arribas C. (2002). Plan de Acción de la UE y del Estado Español sobre enfermedades de baja prevalencia. *Boletín del Real Patronato sobre* Discapacidad, 53: 25-29.

Posada M, Martín-Arribas C, Ramírez A, Villaverde A & Abaitua I. (2008). Enfermedades raras. Concepto, epidemiología y situación actual en España. *Ann Sis San Navar, 32* (suppl 2): 9-20.

Posada M, Villaverde-Hueso A, Alonso V, János S, Zurriaga O, Pollán M & Abaitua-Borda I. (2010). Rare diseases epidemiology research. In: Posada de la Paz M & Groft S. (Eds.) *Rare Diseases Epidemiology*. Springer, Dordrecht, pp. 17-39.

Posada M, Alonso V & Bermejo E. (2016). *Enfermedades raras*. Catarata, Madrid.

Pridmore SA. (1990). The large Huntington's disease family in Tasmania. *Med J Aust, 153*: 593-595.

Pringsheim T, Wiltshire K, Day L Dykeman J, Steeves T & Jette N. (2012). The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. *Mov Disord, 27*(9): 1083-1091.

Provine WB. (2004). Ernst Mayr: Genetics and Speciation. *Genetics, 167*(3): 1041-1046.

PRTR España. *Registro Estatal de Emisiones y Fuentes Contaminantes*. Accessed: July 2019, http://www.prtr-es.es/.

Pupo R, Michel J, Díaz-Rojas Y, Rojas-Rodríguez Y, Rodríguez-Batista Y & Núñez-Arias E. (2013). Actualización en enfermedad de Huntington. *Correo Científico Med, 17*: 546–557.

Ragonese P, Filipini G, Salemi G, Beghi E, Citterio A, D'Alessandro R, Marini C, Monsurrò MR, Aiello I, Morgante L, Tempestini A, Fratti C, Ragno M, Pugliatti M, Epifanio A, Testa D & Savettieri G. (2004). Accuracy of death certificates for amyotrophic lateral sclerosis varies significantly from north to south of Italy: implications for mortality studies. *Neuroepidemiology, 13*(1-2): 73-77.

Ramalle-Gomara E, González MA, Perucha M, Quinones C, Lezaun ME & Posada de la Paz M. (2007). Huntington's disease mortality in Spain in the period 1981-2004. *Rev Neurol, 45*(2): 88-90.

Ramirez L. (2005). Las tecnologías de la información geográfica aplicadas a la planificación territorial sanitaria. *Serie Geográfica, 12*: 57-82.

Ramos-Arroyo MA, Moreno S & Valiente A. (2005). *J Neurol Neurosug Psychiatry, 76*(3): 337-342.

Regidor E, de Mateo S, Rodríguez C & Gutiérrez-Fisac JL. (1993). Evaluación de la significación estadística y cálculo del intervalo de confianza de la razón de mortalidad estandarizada. *Gac Sanit, 7*(38): 237–243.

Regulation (EC) No 166/2006 of the European Parliament and of the Council of 18 January 2006 Concerning the Establishment of a European Pollutant Release and Transfer Register and Amending Council Directives 91/689/ECC and 96/61/EC. (2006). European Parliament, Brussels.

Reinhold-Keller E, Zeidler A, Gutfleisch, J, Peter HH, Raspe HH & Gross WL. (2000). Giant cell arteritis is more prevalent in urban than in rural populations: results of an epidemiological study of primary systemic vasculitides in Germany. *Rheumatology (Oxford), 39*: 1396-1402.

Rivera-Navarro J, Cubo E & Mariscal N. (2015). Analysis of the reason for non-uptake of predictive testing for Huntington's Disease in Spain: A qualitative study. *J Genet Couns, 24*(6): 1011-1021.

Roos PM. (2013). *Studies on Metals in Motor Neuron Disease*. PhD Thesis. Karolinska Institutet, Sweden.

Rose G, Khaw K-T & Marmot M. (2008). *Rose's Strategy of Preventive Medicine*. Oxford University Press, Oxford.

Santartún A, Villar A, Delgado-Alvarado M & Riancho J. (2016). Trends in motor neuron disease: association with latitude and air lead levels in Spain. *Neurol Sci, 37*(8): 1271-1275.

Santos-Preciado JM. (2004). *Sistemas de Información Geográfica*. UNED, Madrid.

Schoenbach V & Rosamond W. (1999). *Understanding the fundamentals of epidemiology: an evolving text*. Chapel Hill, North Carolina.

Schofield R, Reher H & Bideau A. (1991). *The decline of mortality in Europe*. Claredon Press, Oxford.

Shaw NT & McGuire SK. (2017). Understanding the use of geographical information systems (GISs) in health informatics research: a review. *J Innov Health Inform, 24*(2): 228-233.

Sipilä JO, Hietala M, Siitonen A, Pälvärinta M & Majamaa L. (2015). Epidemiology of Huntington's disease in Finland. *Parkinsonism Relat Disord, 21*(1): 46-49.

Slocum TA, McMaster RB, Kessler FC & Howard HH. (2005). *Thematic cartography and geography visualization*. Pearson Prentice Hall, Upper Saddle River.

Snow J. (1855). *On the Mode of Communication of Cholera*. John Churchill, London.

Solberg K, Filkukova P, Frich JC & Billaud-Feragen KJ. (2018). Age at death and causes of death in patients with Huntington Disease in Norway in 1986–2015. *J Huntingtons Dis, 7*(1): 77-86.

Solís JC. (2001). Las topografías médicas: revisión y cronología. *Asclepio: Revista de historia de la medicina y de la ciencia, 53*(1): 213-244.

Sorek-Hamer M, Just AC & Kllog I. (2016). The use of satellite remote sensing in epidemiological studies. *Curr Opin Pediatr, 28*(2): 228-234.

Stanforth A, Moreno-Madriñán MJ & Ashby J. (2016). Exploratory analysis of dengue fever niche variables within the rio Magdalena watershed. *Remote Sens, 8*(9): 770.

Steenland K & Deddens JA. (1997). Design and analysis of studies in environmental epidemiology. In: Steenland K & Savitz D. (Eds.) *Topics in environmental epidemiology*. Oxford University Press, New York.

Subdirección General de Estadísticas de la Población. (2011). *Proyecto de los Censos Demográficos 2011*. Instituto Nacional de Estadística, Madrid.

Tan JA, Choi HK, Xie H, Sayre EC, Esdaile JM, Aviña-Zubieta JA. (2018). All-cause and cause-specific mortality in Granulomatosis with Polyangiitis: A Population-based Study. *Arthritis Care Res (Hoboken), 71*(1): 155-163.

Tavakolpour S & Alesaedi S. (2019). Rituximab as an effective and probably safe treatment for granulomatosis with polyangiitis (Wegener's Granulomatosis). *Int Inmmunopharmacol, 70*:67-68.

Tchounwou PB, Yedjou CG, Patlolla AK & Sutton DJ. (2010). Heavy metals toxicity and the environment. In: Luch A. (Ed.) *Molecular, Clinical and Environmental Toxicology, vol 1*. Springer, Basel.

Timmis A, Townsend N, Gale C, Grobbee R, Maniadakis N, Flather M, Wilkins E, Wright L, Vos R, Bax J, Blum M, Pinto F, Vardas P & ESC Scientific Document Group. (2018). European Society of Cardiology: Cardiovascular Disease Statistics 2017. *Eur Heart J, 39*(7): 508-579.

Tirado-Sánchez A, Ponce RM, Montes de Oca G & León G. (2006). Pénfigo vulgar. Estudio epidemiológico y análisis de posibles factores de riesgo de mortalidad. *Dermatología Rev Mex, 50*(2): 50-53.

Titos-Gil S, Arizo-Luque V & Palomar-Rodríguez JA. (2011). *Aproximación a las enfermedades raras en la Región de Murcia 2002-2009*. Consejería de Sanidad y Política Social, Murcia.

Tomlinson R. (1973). *A technical description of the Canada Geographic Information System*. Environment Canada, Ottawa.

Ugarte MD, Ibáñez B & Militino AF. (2006). Modelling risks in disease mapping. *Stat Methods Med Res, 15*(1): 21-35.

United Nations. (1991). *Handbook of Vital Statistics Systems and Methods, Volume 1: Legal, Organisational and Technical Aspects, United Nations Studies in Methods, Glossary, Series F, No. 35*. United Nations, 1991.

Urteaga L. (1980). Miseria, miasmas y microbios. Las topografías médicas y el estudio del medio ambiente en el siglo XIX. *Geocrítica, Cuadernos Críticos de Geografía Humana, 29*.

Valdez R, Ouyang L & Bolen J. (2016). Public health and rare diseases: Oxymoron no more. *Prev Chronic Dis, 13*: E05.

Velleman PF. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. *J Am Stat Assoc, 75*: 609-615.

Veiga-Cabo J, Almazán-Isla J, Sendra-Gutiérrez JM & de Pedro-Cuesta J. (1997). Differential features of motor neuron disease mortality in Spain. *Int J Epidemiol, 26*: 1024-1032.

Verhasselt Y. (1993). Potentialities of geography of health, “Elisabeth Cass” Memorial lecture. *Arch Public Health, 51*: 481-486.

Villaverde-Hueso A, Sánchez-Díaz G, Molina-Cabrero FJ, Gallego E, Posada de la Paz M & Alonso-Ferreira V. (2019). Mortality Due to Cystic Fibrosis over a 36-Year Period in Spain: Time Trends and Geographic Variations. *Int J Environ Res Public Health, 16*: 119.

de Vrueh R, Baekelandt E & de Haan J. (2013). Background Paper 6.19 rare diseases: update on 2004 background paper written by van Weely S & Leufkens HGM. In: *Priority medicines for Europe and the world: A public health approach to innovation*. WHO, Geneva.

Walker FO. (2007). Huntington's disease. *Semin Neurol, 27*(2): 143-150.

Williams-Pickle L, Mungiole M, Jones GK & White AA. (1996). *Atlas of United States Mortality*. National Center for Health Statistics, Hyattsville.

Wohl Y & Brenner S. (2003). Pemphigus in Israel - An Epidemiologic Analysis of Cases in Search of Risk Factors. *Isr Med Asssoc J, 5*(6): 410-412.

Wong D. (2009). The modifiable areal unit problem (MAUP). In Fotheringham AS (Ed.) *The SAGE handbook of spatial analysis*. Springer, Dordrecht, pp. 95-112.

Wright JK. (1936). A method of mapping densities of population: with Cape Cod as an example. *The Geographical Review, 26*(1): 103-110.

WHO. (1946). *Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946*. World Health Organization, Geneva.

WHO. (1967). *Twentieth World Health Assembly. Part I, Resolutions and decisions, annexes*. World Health Organization, Geneva.

WHO. (1977). *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*. World Health Organization, Geneva.

WHO. (2008). *Atlas of health in Europe, 2nd edition*. World Health Organization Regional Office for Europe, Copenhagen.

WHO. (2011). *International statistical classification of diseases and related health problems*. World Health Organization, Geneva.

Young AB, Shoulson I, Penney JB, Starosta-Rubinstein S, Gomez F, Travers H, Ramos-Arroyo MA, Snodgrass SR, Bonilla E & Moreno H. (1986). More Huntington's disease in Venezuela: neurologic features and functional decline. *Neurology, 36*(2): 244-299.

Annex

Other publications related to this dissertation

Scientific papers

Llamosas-Falcón L, Bermejo-Sánchez E, Sánchez-Díaz G, Villaverde-Hueso A, Posada de la Paz M & Alonso-Ferreira V. (2019). Tetralogy of Fallot in Spain: a nationwide registry-based mortality study across 36 years. *Orphanet J Rare Dis, 14*(1): 79. doi: 10.1186/s13023-019-1056-y.

Villaverde-Hueso A, Sánchez-Díaz G, Posada de la Paz M & Alonso-Ferreira V. (2019). Mortality Due to Cystic Fibrosis over a 36-Year Period in Spain: Time Trends and Geographic Variations. *International J Environ Res Public Health, 16*, 119. doi: 10.3390/ijerph16010119.

Ortega-Torres A, Sánchez-Díaz G, Villaverde-Hueso A, Posada de la Paz M & Alonso-Ferreira V. (2018). Mortalidad atribuida a telangiectasia hemorrágica hereditaria y variabilidad geográfica en España (1981-2016). *Gac Sanit*. In press. doi: 10.1016/j.gaceta.2018.08.004.

Alonso-Ferreira V, Sánchez-Díaz G, Villaverde-Hueso A, Posada de la Paz M & Bermejo-Sánchez E. (2018). A Nationwide Registry-Based Study on Mortality Due to Rare Congenital Anomalies. *International J Environ Res Public Health, 15*, 1715. doi: 10.3390/ijerph15081715.

Book chapters

Arias-Merino G, Sánchez-Díaz G, Hens M, Villaverde-Hueso A, Posada de la Paz M & Alonso V (2017) Mortality statistics and their contribution to the improvement of the rare diseases epidemiology: an example for Hereditary Ataxia in Europe. (2017). In: Posada de la Paz M, Taruscio D & Groft S. (Eds.) Rare Diseases Epidemiology: Update and Overview. Springer, pp. 521-533. doi: 10.1007/978-3-319-67144-4_28.

Communications in scientific conferences

XXXVI Reunión Científica de la Sociedad Española de Epidemiología (11-14 September 2018, Lisbon, Portugal). Arias-Merino G, Sánchez-Díaz G, Villaverde-Hueso A, Posada de la Paz M & Alonso-Ferreira V. (2018). Evidencia de cambios en los patrones de mortalidad por sarcoidosis en Europa, 2003-2014. *Gac Sanit, 32*(Espec Congr):300-335.

9th European Conference on Rare Diseases & Orphan Products (10-12 May 2018, Vienna, Austria). Alonso-Ferreira V, Villaverde-Hueso A, Arias-Merino G, Sánchez-Díaz G & Posada de la Paz M. (2018). Rare Diseases in Spain: a nationwide registry-based mortality study. *Orphanet J Rare Dis, 13* (suppl 2):167. doi: 10.1186/s13023-018-0895-2.

IX Reunión Anual CIBERER (12-14 March 2018, Castelldefels, Spain). Sánchez-Díaz G, Arias-Merino G, Villaverde-Hueso A, Morales-Piga A, Posada de la Paz M & Alonso-Ferreira V. (2018). Epidemiology of Granulomatosis with Polyangiitis mortality in Spain: Spatial and temporal dynamics.

14th International Conference on Urban Health (26-29 September 2017, Coimbra, Portugal). Sánchez-Díaz G, Alonso-Ferreira V, Arias-Merino G, Villaverde-Hueso A, Posada de la Paz M & Escobar F. (2017). A cartographic approach to analysing Granulomatosis with Polyangiitis mortality in urban and rural municipalities.

X Reunión Anual CIBERER (23-24 March 2017, San Lorenzo del Escorial, Spain). Arias-Merino G, Sánchez-Díaz G, Hens M, Villaverde-Hueso A, Posada de la Paz M & Alonso-Ferreira V. (2017). Cambios en la Mortalidad por Enfermedad de Huntington en Europa 2001-2012.

VIII Congreso Internacional de Medicamentos Huérfanos y Enfermedades Raras (2-4 February 2017, Seville, Spain). Sánchez-Díaz G, Escobar F, Arias-Merino G, Villaverde-Hueso A, Posada de la Paz M & Alonso V. (2017). Atlas Nacional de Mortalidad debida a Enfermedades Raras.

11th annual ICORD meeting (19-22 October 2016, Cape Town, South Africa).

Sanchez Diaz G, Arias Merino G, Villaverde Hueso A, Hens Perez MJ, Posada de la Paz M & Alonso Ferreira V. Epidemiology of Granulomatosis with Polyangiitis mortality in Spain: Spatial and temporal dynamics. (2016). *Rare Diseases and Orphan Drugs, 4*(suppl 1): 1-39.

Zurbano-San-Emeterio A, Sánchez-Díaz G, Arias-Merino G, Posada de la Paz M, Escobar-Martínez FJ & Alonso-Ferreira V. (2016). Design and implementation of an interactive mapping tool to represent the distribution of Rare Disease-related in Spain. *Rare Diseases and Orphan Drugs, 4*(suppl 1): 1-39.

Arias-Merino G, Sánchez-Díaz G, Villaverde-Hueso A, Hens M, Posada de la Paz M & Alonso V. Mortality due to hereditary ataxia: time trends and spatial differences in Europe. (2016). *Rare Diseases and Orphan Drugs, 4*(suppl 1): 1-39.

Alonso V, Villaverde-Hueso A, Arias-Merino G, Sánchez-Díaz G & Posada de La Paz M. Congenital anomalies in Spain: a population-based mortality study (2016). *Rare Diseases and Orphan Drugs, 4*(suppl 1): 1-39.

XXXIV Reunión Anual de la Sociedad Española de Epidemiología y XI Congresso da Associação Portuguesa de Epidemiologia (14-16 September 2016, Seville, Spain). Arias-Merino G, Sánchez-Díaz G, Hens M, Villaverde-Hueso, Abaitua I, Posada M & Alonso V. (2017). Mortalidad por enfermedad de Huntington en Europa. *Gac Sanit, 31*(Espec Congr): 24-347.

5º International Meeting on Spastic Paraparesis and Ataxias, ASG / Spatax International Conference (23-25 June 2016, Paris, France). Arias-Merino G, Sánchez-Díaz G, Villaverde-Hueso A, Hens M, Abaitua I, Posada de la Paz M & Alonso V. (2016). Mortality due to hereditary ataxia: Time trends and spatial differences in Europe.

IX Reunión Anual CIBERER (7-8 March 2016, Castelldefels, Spain). Sánchez-Díaz G, Alonso V, Arias-Merino G, Villaverde-Hueso A, Hens M, Morales-Piga A, Abaitua I & Posada de la Paz M. (2016). Estudio geográfico y temporal de la mortalidad debido a la enfermedad de Huntington en España (1984-2013).