Modelling Mortality from Self-Reported Health and Disability Data: A Multi-Scale Report from Ireland, 2011-2016

Modelos de Mortalidad a Partir de Datos de Salud y Discapacidad Autopercebidos: Un Informe sobre Irlanda a Distintas Escalas, 2011-2016

Ronan Foley
Maynooth University, Ireland
ronan.foley@mu.ie


Enlace artículo/to link to this article: http://dx.doi.org/10.12795/rea.2018.i35.07

Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional
http://editorial.us.es/es/revista-de-estudios-andaluces
https://revistascientíficas.us.es/index.php/REA
Modelling Mortality from Self-Reported Health and Disability Data: A Multi-Scale Report from Ireland, 2011-2016

Modelos de Mortalidad a Partir de Datos de Salud y Discapacidad Autopercebidos: Un Informe sobre Irlanda a Distintas Escalas 2011-2016

Ronan Foley
Maynooth University, Ireland
ronan.foley@mu.ie

Recibido: 11 de diciembre, 2017
Revisado: 25 de enero, 2018
Aceptado: 26 de enero, 2018

Abstract
Relationships between mortality and morbidity are long established within geo-spatial modelling and GIS-based analyses. While there has always been a strong associational relationship between the two measures, this has been less tested at an aggregate areal scale than one might expect. From a geographical perspective this has been in part because access to data at meaningful spatial scales, especially for mortality, can be difficult.

This paper presents newly collected data from Ireland on mortality and maps cross-sectional associations with self-reported health and limiting long-term illness and disability conditions (LLTC) for the first time at an intermediate level geography. Data is also available for two different time-periods from administrative records and from five-yearly censuses. Mortality data was collected at a newly-created intermediate level geography (IA, n=410, average population=10,900) while the self-reported health/LLTC data was collected at a long-standing administrative scale (ED, n=3409, average population=1,350). Given there was a nested relationship between the two scales, redistricting techniques were used in GIS to enable direct comparisons. Mortality data was available for all deaths (SMR) and premature death (ASR), as well as for four different causes of death. Self-reported health was mapped in three ways; as a combined rate for poor health status; as a weighted health score and as a combined rate for LLTC. The associations were derived from correlation/regression modelling at the smaller IA scale. The results identified that the associations were statistically significant and of mixed magnitude, but had relatively low r-squared values. The associations were strongest for premature (under 75) mortality, while additional
correlations for cause of death were lower again. From this, we concluded that the self-reported health/LLTC statistics, while of some potential explanatory value, were not especially useful as predictive variable. Suggestions for improvement would be to weight the self-reported health data by age and additionally to take into account deprivation as a second explanatory factor operating within cross-sectional work. Finally, modelling at different spatial scales might act as a useful guide for comparative analysis in Andalucía and other regions of Spain, where spatial scales may be similar in terms of size and scale.

**Keywords:** mortality, morbidity, spatial patterns, modelling, Ireland.

**Resumen**
Desde hace mucho tiempo, se vienen realizando análisis geoespaciales utilizando herramientas SIG para establecer la relación entre mortalidad y morbilidad. A pesar que siempre ha existido una estrecha relación entre estas dos dimensiones, esta ha sido escasamente evaluada a escala agregada. Desde una perspectiva geográfica, esto no ha sido posible en parte debido a que el acceso a datos, tales como mortalidad, ha sido difícil para determinadas escalas espaciales. Este artículo presenta por primera vez, a un nivel geográfico intermedio, información y mapas recientes sobre mortalidad en Irlanda, asociándola de forma transversal con la salud auto-percibida, la limitación a causa de una larga enfermedad y el nivel de la discapacidad (LLTC). Los datos utilizados provienen de dos fuentes, registros administrativos y censos quinquenales y pertenecen a dos periodos diferentes temporales. Los datos de mortalidad se recopilaron en base a un nivel geográfico intermedio (IA, n = 410, población media = 10,900), mientras que los datos de salud auto-percibida/LLTC se recopilaron en base a una antigua escala administrativa (DE, n = 3409, población media = 1,350).

Ante la falta de homogeneidad entre las dos escalas, se utilizaron técnicas de reordenación mediante SIG para realizar comparaciones directas. Los datos de mortalidad estaban disponibles para todas las muertes (SMR) y las muertes prematuras (ASR), para cuatro causas diferentes de mortalidad. La salud auto-percibida fue cartografiada de tres formas distintas: como una tasa combinada para el mal estado de la salud; como una puntuación ponderada de la salud y como una tasa combinada para LLTC. Las relaciones fueron derivadas desde un modelo de correlación/ regresión a una escala más pequeña IA. Los resultados han puesto de manifiesto que las relaciones eran estadísticamente significativas y de magnitud mixta, pero tenían valores r-cuadrados relativamente bajos. La relación fue más significativa para las muertes prematuras (menor de 75), mientras que las correlaciones relacionadas con la causa de muerte fueron bajas. A partir de esto, se llegó a la conclusión de que las estadísticas respecto a la salud auto percibida /LLTC, aunque poseen algún valor explicativo, no se pueden utilizar como variables predictores. Para mejorar los resultados en la realización de estudios transversales, se sugiere que la información sobre la salud auto-percibida se divida por grupos de edad y, además, se
tengan en cuenta la deprivación como un segundo factor explicativo. Finalmente, el modelado a diferentes escalas espaciales podría servir como una guía útil para realizar análisis comparativos en Andalucía y otras regiones de España, donde las escalas espaciales pueden ser similares en términos de tamaño y escala.

Palabras clave: mortalidad, morbilidad, patrones espaciales, diseñar, Irlanda.
1. INTRODUCTION

Relationships between mortality and morbidity are well established but are always open to further scrutiny within geo-spatial modelling and GIS-based analyses (Gatrell and Elliot, 2009). Typically, places that have high levels of morbidity tend to produce high levels of mortality, with the most common causal factors associated with low socio-economic status (SES), high levels of poverty and deprivation and additional spatial justice factors such as unhealthy and polluted environments (Boyle et. al., 2004; Brown et. al, 2017). In addition recent research on the developing impact of health behaviours in terms of physical activity, diet and growing obesity have also had a spatial component (Carroll-Scott et. al., 2013). While there has always been a strong associational relationship between the two measures, this has been less tested at an aggregate areal scale than one might expect (McIntyre, Ellaway and Cummins, 2002). This is in part based on wider theoretical discussions within social and health geographies about the relative importance of context (the place) or composition (the people in a place) in explaining spatial variations (Pickett and Wilkinson, 2015). Which of these factors is more important is still open to debate, but for those people who suggest that context matters more than composition, the focus of their research lies on identifying a fine enough scale of analysis to get as close to neighbourhood level as possible. In addition sensitivity to data protection means that individual level data cannot be used (Rigby et. al, 2017). From a geographical perspective this has been in part because access to data at meaningful spatial scales, especially for mortality, can be difficult (Foley and Kavanagh, 2016). This paper presents newly collected data from Ireland on mortality and, as a proxy for morbidity, self-reported health and limiting long term conditions (illness and disability, or LLTC) status. The aim of the piece is to identify whether any strong associations emerge to prove the relative importance of context.

At the heart of the research was a relatively straight-forward piece of geo-spatial modelling. This consisted of the mapping out and comparing of data on the geographical distribution of mortality, captured at a new intermediate scale of reporting, with the geography of proxy morbidity, as measured via Census data on self-reported health. The novelty of the paper is that this is the first time testing out associations of mortality and morbidity data in Ireland over two time periods, as the relevant data has not been available before now. Up to the last decade, getting information on the geography of mortality in Ireland had been more accurate in 1900 than in 2000 (Houghton, 2005). In general, mortality is formally reported through official routes, which includes information on specific causes of death. Other specific cause of death data, such as for cancer, are collected by separate agencies in Ireland such as the Cancer Registry, who produce regular maps of aggregated deaths by specific cancers for specific clusters of years and make these available publicly (https://www.ncri.ie/data/maps). In choosing a mortality over a morbidity route, the research was able to explore a number of key spatial issues, with a primary aim to establish statistical results on the associations between the two indicators at a
standardised areal scale. A further objective was to provide some additional nuanced results by cause of death to see if there were stronger relationships for these specific indicators. While a number of additional constraints were identified, at the heart of the work was the identification of the value of aggregated mortality data and its capacity to be modelled at an area-level spatial scale, notwithstanding some known modifiable areal unit problems (Schuurman et. al., 2007; Rigby et. al., 2017).

Within Ireland, there are a number of administrative geographies against which data is collected, primarily via the 5-yearly Census. The most recent census years have been 2016, 2011, 2006 and 2002 (the delayed 2001 census). These shift and change over time but in 2011, the core collection date for this analysis, there were 34 Local Authorities (LA), the traditional cities and counties of Ireland with an average population of around 130,000 and a range from 32,000 to 527,000. Traditionally the main lower level geography had been the electoral division (ED, n=3409, average population= 1,350, range 75 to 36,000). This administrative unit has been in place since the mid-19th century, though its' significantly skewed population range makes it potentially problematic. In 2011, on the back of a series of national consultation exercises, a new neighbourhood level Small Area (SA) geography was introduced which had 18,488 units, a mean population of 250 and a narrow range of 70 to 1,940. In effect, especially in urban settings, this was a housing estate level geography. In developing this analysis, mortality data (a non-Census indicator) was collected and mapped at a newly created intermediate level geography (IA, n=410, average population=10,900) while the self-reported health and LLTC caring data was collected at the above three administrative scales, LA, ED and SA. The new IA level geography was developed by the team at the NCG, Maynooth University, using an aggregation programme, AZTool (Cockings et. al., 2011) to amalgamate SAs to a target population size of 10,000 based on a number of criteria around number thresholds, shape, social homogeneity and containment constraints (see Rigby et. al., 2017 for fuller details).

2. DATA COLLATION AND METHODS

In calculating the mortality data, Jan Rigby and colleagues were able to access individual level death records annually for the period 2006 to 2011 and map this to the new aggregated IA geography described above. Mortality data, from the Irish Central Statistics Office (CSO) and the General Record Office (GRO). There were a number of problems with the raw data that were tidied up at data preparation stage. Firstly, the data was cleaned as, rather surprisingly, several people were recorded as dying more than once. Secondly, the home address of the individual was geo-referenced, but the success rate for this process was relatively low (on average around 80%). This was due to a mix of poor recording but also the presence of a considerable number of duplicate addresses, especially in rural areas. With the arrival of new individual household postcodes in 2016, this process should be more speedy and accurate in future years. Finally, small number problems, even when aggregated to IA scale, meant that data
had to be aggregated across a five-year time period, which meant that longitudinal analysis was affected. Data was developed for all deaths (SMR) and premature deaths (ASR), as well as for four different causes of death, essentially the main ‘killers’ in Ireland, cancer, heart, cardiovascular and respiratory (Rigby et. al., 2017). Figure 1 below shows one representative map of this data based on the Age-Standardised Mortality rates for Under 75s – effectively premature mortality – for 2011 at the IA scale. The map shows a mixed pattern in rural areas, with lower levels in suburban areas, while the inset map for the Greater Dublin Area shows a sort of northeast to south-west divide, with the higher levels associated with a generally more deprived geography north of that line.

Figure 1. Age-Standardised Mortality (Under 75 years) in Ireland at IA Scale, 2011.
The other core data set used to indicate morbidity was taken from the 2011 Census, and used one newly revised variable, self-reported health, and from that data, a derived broad index, the Kavanagh-Foley Index of Wellbeing (KFIW) along with data on LLTC. For the purposes of this paper, the 2011 SA data was not used, but the other data was available at ED level for both indicators. The health question specifically asked for respondents to report on their perceived level of health. In 2011, respondents were asked to select from one of five responses - in effect a Likert scale - in addition to a not-stated class in the final data. The answers were listed as; very good – good – fair – bad – very bad. The amounts and levels of the former two classes were very high (around 85%), while those of the latter two classes were very low (around 2%), compared to other parts of the British Isles (Foley and Kavanagh, 2016). This was accounted for in part by the generally much younger average population age of the Republic of Ireland. Figure 2 below shows the ED level geography for an aggregated percentage of poor health (those who classified themselves as in fair, bad or very bad health) for 2011. The KFIW was calculated as follows; a weighting of 1 was added to the percenta...
Figure 2. Geography of Self-Reported ‘Poor’ Health at ED Scale, 2011.

(onto a new field) against each of the individual EDs that fell inside the associated IA boundary. These were non-overlapping in all cases, as the original IAs had been developed from clusters of SAs (Cockings et. al., 2011; Rigby et. al., 2017). The second step was then to run a dissolve (redistrict) command within the GIS, using the newly created IA field in the ED layer as a look-up, to match the lower-level geography to the upper-level one and to calculate appropriate aggregate statistics. We then had a set of comparable data for mortality and morbidity at the same scale for two different time periods. We were not able to create a full set as the mortality data was an aggregation across the time period from 2006 to 2011, but we did use that aggregated data for modelling at either end of that period, to at least begin to explore the relationships.
between this composite score and the other indicators for 2006 and 2011. This did raise issue of optimal data collection scales, which are still ongoing, until new mortality data for the period from 2014 to 2016 (currently being processed) becomes fully available.

3. RESULTS

The results of the modelling are presented below in two forms, one statistical, the other visual. Firstly, the statistical associations with mortality at IA level for the dependent variables, poor self-reported health, KFIW index score and high levels of LLTC are listed in Table 1. These include correlations and regressions for overall mortality (ASR and MSR) as well as for individual causes of death, though the data for LLTC is more restricted due to small number constraints. Overall the scores for the ASRs and MSRs produce the best results, with the strongest relationships evident for premature mortality against poor self-reported health (0.417, 0.437) but even for all age mortality the associations are significant and reasonably large (0.342, 0.336). The relationships with the KFIW Score are weaker and in the cases of all age ASR and MSR, not significant, while the measures are lower for premature mortality (0.204, 0.215). The results for LLTC, somewhat surprisingly, are weaker again and we feel this may also be due to an artefact in the aggregation of that data, where the combination of a range of conditions and severity levels have the effect, in data terms of cancelling each other out. In the case of the individual cause of death data, the associations are relatively low, though with some variations. Pearson’s Correlation values (statistically significant) between the dependent variables and heart disease were the highest (0.353, 0.335, 0.211) with the next highest associations recorded against cancers (0.294, 0.261) and respiratory disease (0.288, 0.257). In the case of these latter two causes of death, the LLTC data was not significant. Statistics for the associations with stroke were not statistically significant and were the lowest of the set for health and KFIW measures (0.134, 0.126).

Secondly, in Figure 3 below, one can see a visual representation of the two data sets for which there seemed to be the strongest association, SMR Under 75 and Poor SR Health. While the statistics for the broad mortality data showing up some strong associations, these are less immediately evident through the maps, especially for some the mid-western and midlands rural areas and around some of the smaller cities.

A better visualisation of the relationship is shown in Figure 4, a closer zoom in to the Greater Dublin Area, which accounts for about one third of the national population. While the broad northeast/southwest divide remained visible for both indicators, there were some significant variations. These applied especially in some of the suburban areas to the north-west of the map. These were areas with young populations, yet had relatively high levels of premature mortality. They were also areas
Table 1. Statistical Associations at IA level, 2011. (*Statistically significant).

<table>
<thead>
<tr>
<th>Independent</th>
<th>Dependent</th>
<th>Poor SR Health</th>
<th>KFIW Index</th>
<th>High LLTC rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR All</td>
<td>Pearson's Correlation</td>
<td>0.342*</td>
<td>0.134</td>
<td>0.228*</td>
</tr>
<tr>
<td>ASR Under 75</td>
<td>Pearson's Correlation</td>
<td>0.417*</td>
<td>0.204*</td>
<td>0.216*</td>
</tr>
<tr>
<td>SMR All</td>
<td>Pearson's Correlation</td>
<td>0.336*</td>
<td>0.126</td>
<td>0.237*</td>
</tr>
<tr>
<td>SMR Under 75</td>
<td>Pearson's Correlation</td>
<td>0.437*</td>
<td>0.215*</td>
<td>0.219*</td>
</tr>
<tr>
<td>Mean ASR Respiratory Disease</td>
<td>Pearson's Correlation</td>
<td>0.288*</td>
<td>0.257*</td>
<td>0.085</td>
</tr>
<tr>
<td>Mean ASR Stroke</td>
<td>Pearson's Correlation</td>
<td>0.134</td>
<td>0.126</td>
<td>0.091</td>
</tr>
<tr>
<td>Mean ASR Heart Disease</td>
<td>Pearson's Correlation</td>
<td>0.353*</td>
<td>0.335*</td>
<td>0.211*</td>
</tr>
<tr>
<td>Mean ASR Cancer</td>
<td>Pearson's Correlation</td>
<td>0.294*</td>
<td>0.261*</td>
<td>0.049</td>
</tr>
<tr>
<td>ASR All</td>
<td>r-squared</td>
<td>0.117*</td>
<td>0.018</td>
<td>0.052*</td>
</tr>
<tr>
<td>ASR Under 75</td>
<td>r-squared</td>
<td>0.174*</td>
<td>0.041*</td>
<td>0.047*</td>
</tr>
<tr>
<td>SMR All</td>
<td>r-squared</td>
<td>0.113*</td>
<td>0.016</td>
<td>0.056*</td>
</tr>
<tr>
<td>SMR Under 75</td>
<td>r-squared</td>
<td>0.191*</td>
<td>0.046*</td>
<td>0.058*</td>
</tr>
<tr>
<td>Mean ASR Respiratory Disease</td>
<td>r-squared</td>
<td>0.083*</td>
<td>0.066*</td>
<td>0.007</td>
</tr>
<tr>
<td>Mean ASR Stroke</td>
<td>r-squared</td>
<td>0.018</td>
<td>0.016</td>
<td>0.008</td>
</tr>
<tr>
<td>Mean ASR Heart Disease</td>
<td>r-squared</td>
<td>0.124*</td>
<td>0.112*</td>
<td>0.044*</td>
</tr>
<tr>
<td>Mean ASR Cancer</td>
<td>r-squared</td>
<td>0.086*</td>
<td>0.068*</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Figure 3. Premature Mortality and Self-Reported Poor Health (IA scale, Ireland).

http://dx.doi.org/10.12795/rea.2018.i35.07
Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional

188
of high deprivation as measured by two different deprivation indices in Ireland and here it was the relative inequality of the area, which combined with age to identify them as problematic. The difference between the different indicators in both Figures 3 and 4 can also be explained in part by the fact that the self-reported health data (the right-hand map) is not full age-standardised. This was in part due to the large number of observations and the difficulties of applying a global modelling method like age-standardisation, to such a local scale. Nonetheless, and as noted by national health agencies, age and deprivation continue to account disproportionately for both high levels of mortality and morbidity across the country, no matter what scale they are measured at (Rigby et. al., 2017).

4. DISCUSSION

In terms of the original aim of the paper, to show the relative importance of context in explaining associations between mortality and morbidity, we would have to say the results suggest that, for spatially aggregated Irish data, this is, at best, a limited association. As noted in the previous paragraph, the importance of wider structural factors, especially associated with poor SES, low levels of employment and the more compositional ways in which poverty and deprivation shape people’s health would seem to be equally and potentially more important. Nevertheless, explanations are almost always a mix of both composition and context and we would argue our work may be of value to identify the relative importance of both factors and might be comparable to similar work in the British Isles and indeed, in other parts of Europe including Andalucía. The ability to map both mortality and morbidity at a meaningful geographical scale might be considered routine in some places, but was innovative for...
Ireland. This was especially the case when the mapping uncovered strong relationships with poverty and deprivation and received much public coverage in Ireland when they were originally released in 2014 (Cullen, 2014).

A second conclusion from the analysis concluded that the self-reported health statistic, while of some potential explanatory value, was not perhaps as fully useful as a predictive variable when aggregated up at national level. This was especially the case when modelling putative associations with indicators of mortality. These were also weakest for some specific causes of death, though again the effects of MAUP (aggregating data from ED to IA scale) and the lack of age-standardisation were complicating factors. Suggestions for improvement would be to weight the self-reported health data by age and additionally to take into account deprivation as a second explanatory factor operating within cross-sectional work (Cairns, Curtis and Bambra, 2012). To date this has not been modelled, as there is a strong and well-established separate relationship between the mortality data and deprivation (Rigby et. al, 2017). However, preliminary explorations of the relationships between self-reported health and deprivation scores have been relatively weak at ED level so it may not provide any additional explanation as it currently stands (Schuurman et. al., 2007; Foley and Kavanagh, 2016).

One surprise was our expectation of the LLTC category as being a good predictor of mortality, or at least a variable with relatively high levels of association, as noted elsewhere in the literature (Boyle, et. al., 1999, 2004). The data used was an amalgamation of individual questions that. As noted above, included both prevalence and levels of severity. We think that the amalgamation of all answers into a single variable has introduced an effective data artefact. Given this the first time this data has been modelled in this way, it has been very helpful to identify its poor performance, at least in contextual work. A way forward would be to request specific data on sub-sets of those questions, especially those on physical and emotional health and chronic illness and model these as individual variables, to identify whether these are more effective single variable predictors (CSO, 2015). To do this would involve a lengthy data request and secure environment work and this would be work for future research.

A final important policy finding of the work was the potential for the IA level geographies, a scale never used before for the collection of health data, to be a potentially valuable in-between scale for future health data research. Sitting as it did between ED and local authority scale, but with sufficient levels of aggregation for more detailed work (clusters of 10,700 people), it has the potential to be a valuable reporting scale, even for anonymised personal data in the future, provided the correct protocols on safe data havens are followed (Moran, 2016). Obviously, the causal factors for both mortality and morbidity extend well beyond their spatial and contextual patterning, but nonetheless, modelling at different spatial scales might act as a useful starting point for comparative analysis and to view those patterns on the ground. Equally such mapping maybe open up scope for comparative work in
Andalucía (such as at District and potentially, *Zonas Básicas de Salud* level) and other regions of Spain, where spatial scales may be similar in terms of size and scale.

**REFERENCES**


