



# Ranking spatial areas by risk of cancer: modelling in epidemiological surveillance

Pablo Fernández-Navarro<sup>1,2,3</sup>, Javier González-Palacios<sup>1,3</sup>, Mario González-Sánchez<sup>1,3</sup>, Rebeca Ramis<sup>1,2</sup>, Olivier Nuñez<sup>1,2</sup>, Francisco Palmí-Perales<sup>4</sup>, Virgilio Gómez-Rubio<sup>4</sup>

<sup>1</sup>Cancer and Environmental Epidemiology Unit, Department of Epidemiology of Chronic Diseases, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain; <sup>2</sup>Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública-CIBERESP), Madrid, Spain; <sup>3</sup>Bioinformatics and Data Management Group (BIODAMA), Department of Epidemiology of Chronic Diseases, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain; <sup>4</sup>Department of Mathematics, School of Industrial Engineering-Albacete, Universidad de Castilla-La Mancha, Albacete, Spain

**Contributions:** (I) Conception and design: P Fernández-Navarro, V Gómez-Rubio; (II) Administrative support: None; (III) Provision of study material or patients: P Fernández-Navarro; (IV) Collection and assembly of data: P Fernández-Navarro, J González-Palacios, M González-Sánchez; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Pablo Fernández-Navarro. Cancer and Environmental Epidemiology Unit, Department of Epidemiology of Chronic Diseases, National Center for Epidemiology (Pab. 12), Carlos III Institute of Health, Avda. Monforte de Lemos, 5, 28029 Madrid, Spain.

Email: pfernandezn@isciii.es.

**Background:** The representation and analysis of maps of events within a fixed time frame/period has been established as a basic tool for disease monitoring. In addition to having methods that can address the study of certain problems, the existence of criteria to discriminate relevant results is equally important. In chronic diseases such as cancer, monitoring the spatial distribution of mortality/morbidity in small areas through relative risk (RR) estimators is used frequently, but there is no clear strategy to discriminate which regions are important. Moreover, it usually requires substantial time for an effective surveillance or an advanced technical knowledge. The objectives of this study are to first establish a data analysis pipeline that allows users to make an initial screening for exploratory purposes so they can identify regions of interest in the context of chronic diseases monitoring and second to develop an R-Shiny application to implement this strategy in a straightforward way without requiring strong technical knowledge.

**Methods:** First, a pipeline of seven steps for ranking risk of disease spatial areas was developed taking into account relative and absolute risk estimators, using observed and expected cases in spatial units of a study region. Second, an R-Shiny application (RANKSPA, Ranking Spatial Areas) was developed to perform the pipeline. Third, we applied the pipeline using RANKSPA to simulated and real data of lung cancer municipal mortality 2005–2009 in Galicia (North-East of Spain), a region with 314 spatial units.

**Results:** There was a clear excess of mortality in the middle-east of the studied region using simulated data where a spatial mortality cluster is also located, existing 5 spatial units outside this cluster that occupy the top positions in the ranking generated by the application. From the total spatial units of the study region [314], only 14 had an excess of mortality whose posterior probabilities are greater than or equal to 80%. In addition, all the spatial tests implemented, with the exception of Moran's I test, were statistically significant. In the study of real data, a clear excess of mortality was observed in the east part of the study region where several of the spatial mortality clusters are also located. Moreover, there are three spatial units located outside these clusters that occupy the top positions in the ranking generated by the application. Eleven spatial units have an excess of mortality, with their posterior probability (PP) greater than or equal to 80%. All the spatial tests implemented were statistically significant. Both in simulated and real data, there was a positive correlation between absolute and relative measures. However, a greater dispersion was observed when these measures take the highest values.

**Conclusions:** The work presented shows a strategy of exploratory analyses to provide an initial assessment

of geographical patterns in disease risk, focused primarily on chronic diseases such as cancer. Furthermore, an R-Shiny application has been created to ease the implementation of this strategy without requiring substantial technical knowledge.

**Keywords:** Spatial analysis; ranking; epidemiological surveillance; ranking application; screening; cancer

Received: 17 February 2020; Accepted: 27 October 2020; Published: 30 November 2020.

doi: 10.21037/ace-20-15

**View this article at:** <http://dx.doi.org/10.21037/ace-20-15>

## Introduction

Public health surveillance includes a wide range of monitoring methods, being the spatial or spatio-temporal study of the diseases a reasonable starting point for any grounded health intervention. Likewise, strategies that employ or integrate different statistical methods could be relevant to best aid the task of this surveillance.

Spatial surveillance merges monitoring statistics for evidence of a change and spatial techniques which are often used to find or describe the extent of clustering across a map. Other ideas related to spatial surveillance is that of screening, which could be applied to populations as well as individuals (1), where the location of the public health event is as important as the fact that it occurred. Surveillance and screening carried out at an aggregate population level (e.g., municipality, census track, etc.) could be useful to detect exceeded limits based on observed or expected patterns, to know where the health incidents occur, and possibly allowing to anticipate where they will occur. All of these might trigger interventions designed, for example, to redirect health resources towards attempts to improve the health status of the population as well as it can be useful for prevention programs.

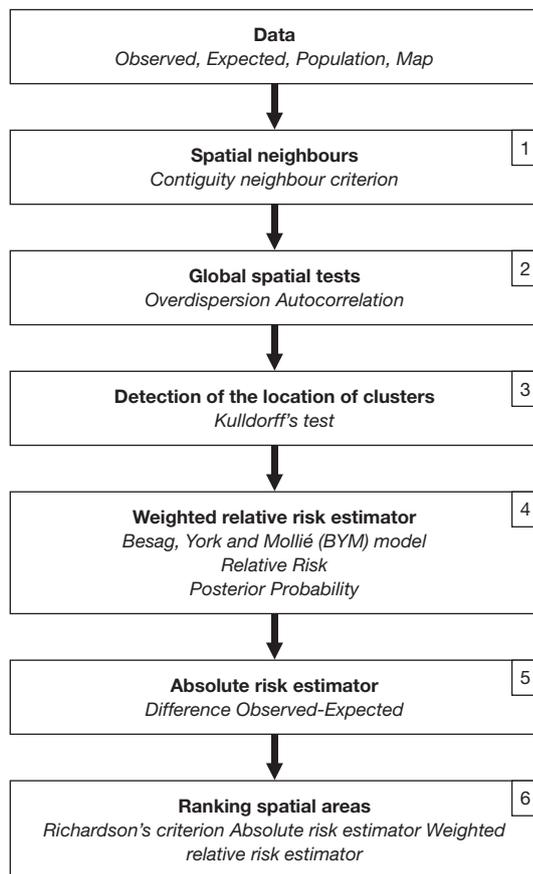
The representation and analysis of maps of events within a fixed time frame/period is established as a basic tool for disease monitoring. In this context, there is a wide range of methods that can be applied, such as disease mapping, spatial clustering or ecological analysis, each of them is appropriate to answer a specific question, and that usually requires advanced technical knowledge (programming, statistics, etc.) to be implemented. In that way, there are valuable applications or statistical packages such as “GeoDa” software (2) or “SSTCDapp” (3), and the R-Shiny web application “SpatialEpiApp” (4), that allow us to perform these methods in a simple way and without extensive technical knowledge.

Furthermore, in spatial surveillance, apart from having

specific methods, it is equally important that there exists a criterion to discriminate relevant results. For example, in disease mapping, which usually involves Bayesian models to smooth the underlying risk estimates, the posterior probability (PP) that the relative risk (RR) is higher than 1 is used as a Bayesian method to assess high risk. Insofar as this indicator was concerned, it is usual to follow Richardson’s criterion (5), which recommends that probabilities above 0.8 should be deemed significant. Although this approach can improve the discrimination of the areas of interest based on the RR, in the context of public health and surveillance the methods should consider absolute measures related to, for example, the magnitude of the expected counts or other approaches to rank risk estimates with PP higher than 80%.

In chronic diseases, such as cancer, studying the spatial distribution of mortality or morbidity in small areas through RR estimators is very frequently used for monitoring (6-11). In this sense, cancer mortality or incidence atlases (12-15) are very useful tools. Nevertheless, on many occasions, the time required for conducting these analyzes or atlases is usually too long for effective surveillance or to develop interventions and address problems in time. In addition, there is no clear strategy to discriminate which regions are important, both from the point of view of relative estimators and absolute estimators important for public health.

Accordingly, the objectives of this work are: (I) to establish a pipeline of analysis that allows us to discriminate regions of interest in the context of chronic diseases (such as cancer) monitoring by studying the spatial distribution in small areas of mortality/morbidity, by means of relative and absolute risk estimators and the precision of them, to obtain a ranking of spatial units, thus being able to implement an initial spatial screening for exploratory purposes useful in epidemiological surveillance and public health and (II) to present the R-Shiny application that has been created to be able to implement this strategy in a simple way, fast and without a deep technical knowledge of the statistical and



**Figure 1** Steps of the pipeline for the ranking of spatial areas.

programming methods necessary to perform it.

## Methods

In order to achieve these objectives, several spatial analyses with the calculation of relative and absolute estimators were conducted to identify and to rank areas of risk of a disease. The study area is divided into several non-overlapping smaller regions and the analyses can start once a measure of observed and expected cases and the population is available for each of the study regions.

As a result, a pipeline of analysis including the strategies followed to rank risk areas has been established as indicated in *Figure 1* and described below. Finally, we show the Shiny web application for ranking spatial areas of risk in spatial surveillance “RANKSPA” to perform the pipeline and what results are obtained by applying this to simulated data as an example and to a real dataset of municipal lung cancer mortality.

### *Pipeline for ranking spatial areas*

The analyses that are presented in this work deal with modelling areal data (observed cases on areal entities with defined boundaries). In these analyses the spatial autocorrelation among areal entities should be taken into account. In that way, the first step in the pipeline is to create the neighbourhood structure using a standard neighbour criterion, the contiguity, where all touching polygons are neighbours. For this purpose, “poly2nb” function of the “spdep” R package could be applied (16,17).

As part of any spatial data analysis, a series of statistical tests should be applied to assess overdispersion, spatial autocorrelation and general spatial clustering. These analyzes constitute the second step and the results will enrich the ranking process that will be discussed later. First of all, the spatial heterogeneity of the risk measures is assessed, testing for global significant differences between observed and expected cases. The reason of a possible heterogeneity may be related to many different factors, such as the presence of different medical assistance systems or pollution sources in the areas. Two common tests of overdispersion are used in this step: the standard Chi-square test and the Pothoff-Whittinghill’s test [see the parametrization described in pages 348–349 of the Bivand *et al.* book (17), which consists of the use of a multinomial model in the bootstrap test and 999 replicates to compute the significance of the observed value of the test statistic]. Functions in the “DCluster” R package (18) “achisq.test” and “pottwhitt.test” could be used to perform them. Moreover, two common tests of global clustering, Moran’s I test (19) and Tango’s maximised excess events test (20,21), are also performed in this step to assess global spatial autocorrelation. “moranI.test” and “tango.test” functions in the “DCluster” R package (18) could be applied to perform these tests [see the parametrization described in pages 350–352 of the Bivand *et al.* book (17), which consists of the use of a negative binomial model in the bootstrap test and 999 replicates to compute the significance of the observed value of the test statistic].

In the third step of the proposed pipeline, a standard test [Kulldorff and Nagarwalla test (22)] for the detection of spatial clusters based on a circular window is applied. We selected this approach because it is a well-established method in this field, easy to implement and with a relatively low computational cost, although it has several limitations as will be discussed later. The results from this analysis will enrich, with relevant information for spatial surveillance,

the ranking of the risk areas that is the final result of the application of this pipeline. “opgam” function in the “DCcluster” R package could be applied to performed this test [see the parametrization described in pages 353–354 of the Bivand *et al.* book (17), which consists of the use of a negative binomial model in the bootstrap test, 99 replicates to compute the significance of the observed value of the test statistic, a significance level of 5% for the tests performed and a 15% fraction of the total population for the circles creation in the method]. To define the window in this detection of spatial clusters, the maximum fraction of the total population inside the cluster could be used. This parameter should be changed to assess if the location of clusters varies significantly. Moreover, only clusters with P values lower than 0.05 will be taken into account.

The fourth step consists of an initial calculation of local RR estimators of disease and the distribution of the PP that  $RR > 1$  (Bayesian way to assess high risks). To do this, based on the data to be analysed, conditional autoregressive model proposed by Besag, York and Mollié (BYM) (8,23,24) would be used. This model is based on fitting a Poisson spatial model with observed cases as the dependent variable, log-expected cases as offset, and two types of random effect terms which take the following into account: (I) contiguity (spatial autocorrelation term); and (II) non-spatial heterogeneity. For more details see the article of López-Abente *et al.* (8). As a tool for Bayesian inference, we recommend the use of Integrated nested Laplace approximations (INLA) (25). The reason for this approach is that it allows reliable results to be obtained in a reasonable time and at much lower computational cost, when compared to the more traditional Markov Chain Monte Carlo (MCMC) method (26–28). The “inla” function in the R-INLA package (29,30) (with the option of simplified Laplace estimation of the parameters) could be used to perform the models. Moreover, the default assumptions for priors of the functions included in the package are used. Finally, in this fourth step, with the estimators, RR and PP for each spatial area, we calculate a weighted RR (WRR) by multiplying each RR by its associated PP. In this way, a penalized RR measure by the precision of the estimation is obtained, and this could help to discard some non-significant results.

The step of the pipeline before ranking the spatial areas of risk, the fifth step, consisted in the calculation of an absolute measure of risk that is defined as the difference between observed and expected (DOE) cases. An indicator similar to the standardized mortality ratio (SMR) or the

standardized incidence ratio (SIR) that is the ratio between observed and expected cases but DOE could be directly more useful for health management aspects as it is expressed as a number of cases.

Finally, the last step of the pipeline, the sixth step, is the identification of those areas whose PP are greater than 80% [Richardson’s criterion (5)]. Firstly, these areas will be ranked according to the DOE and, secondly, according to the WRR. Once the areas have been ranked, they will be assigned a numerical value, starting from 1, to mark the highest level of importance. This means the highest DOE with the highest WRR. All the spatial areas where PP are lower than 80% will have the same rank value and this will be the highest numerical value assigned to the spatial areas of the first group plus one.

The final result of the pipeline is a table of ranked spatial areas attending to absolute and RR estimators and enriched with useful information of global and local clustering and spatial heterogeneity indicators to understand the ranking. This constitutes a rapid initial screening to identify areas of interest for public health.

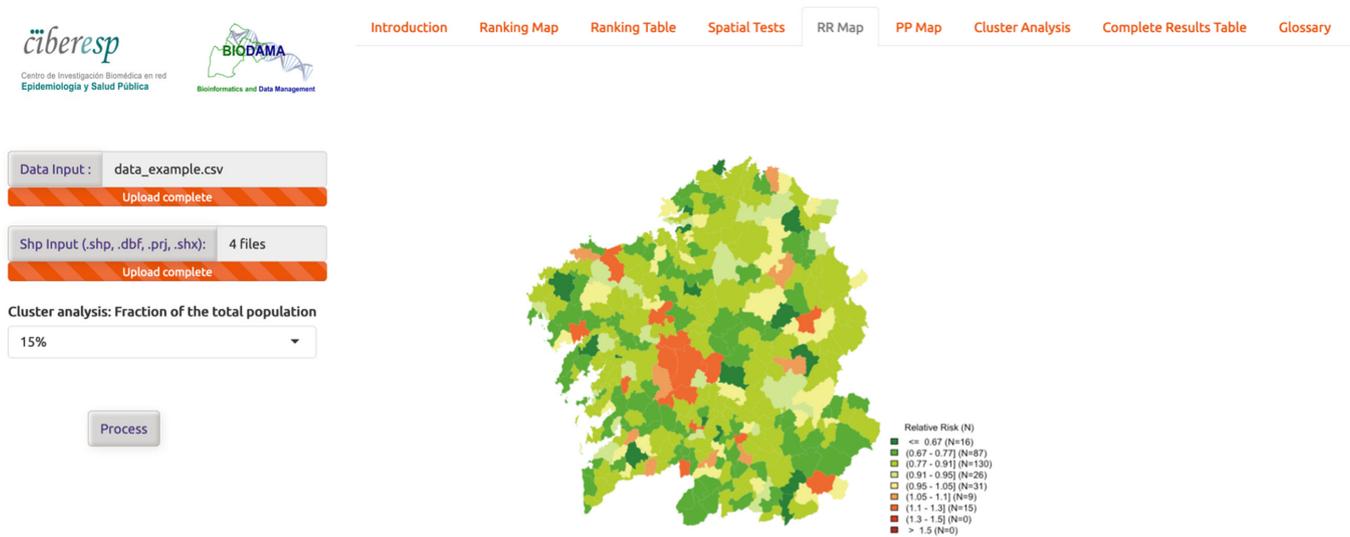
An R script (“R\_script.R”) is provided to perform the pipeline. Simulated data (“data\_example.xlsx”) of observed and expected cases in 314 spatial areas and a shapefile (“map\_example.zip”) are also provided for the application of the script. See “Code and data availability” section of the manuscript.

### **RANKSPA app**

In the same way other authors try to give useful applications for spatial analysis or data management (3,4) and allow users to apply statistical methods in a simple way and without extensive technical knowledge, we have created an R-Shiny web application called RANKSPA (“RANKing SPatial Areas”) which allows users to perform the pipeline described before. In that way, RANKSPA allows to obtain a ranking of spatial areas attending to absolute and relative disease risk. RANKSPA R-Shiny code is provided (see “Code and data availability” section of this manuscript) and it is also available online ready to be used at <https://biodama.shinyapps.io/rankspa/>.

RANKSPA consists of one page where: (I) on the left side, the user can upload the input files and select one of the parameters that is needed for the cluster analysis (“fraction of the total population”, see “Pipeline for ranking spatial areas” section of the manuscript); (II) on the right side of the page, there are nine tabs, where an application

## RANKSPA: Shiny web application for Ranking SPATial areas of risk in spatial surveillance



**Figure 2** RANKSPA application overview.

overview and the results of the different statistical analyses carried out and the maps created by the application can be visualized and downloaded. *Figure 2* shows the RANKSPA application, once the simulated data (see “Code and data availability” section of the manuscript) has been processed and specifically showing the tab corresponding to the map of RRs.

### Input data

First, we upload the data file by clicking the ‘Data input’ button and selecting the corresponding file. The file must be an Excel (XLSX format) or CSV file with the following columns:

- ❖ ID: identification code of the spatial areas;
- ❖ Population: total population in each spatial area;
- ❖ Obs: observed cases in each spatial area;
- ❖ Exp: expected cases in each spatial area.

Second, we upload the map file containing the areas of the region of study. The map file needs to be in shapefile format (files with extension .shp, .prj, .dbf, .shx). The ID area should be a unique identifier and should match with the ID area that is specified in the data. Map files can be uploaded by clicking the “Shp Input” button and selecting all the corresponding files.

After uploading the map and the data files we can specify one option related to the cluster analysis of the pipeline, the maximum fraction of the total population inside the cluster.

The default value is 15%.

### Start analysis

After uploading all the files and selecting the option for the cluster analysis, we click the “Process” button. When we do this, the application starts to perform the analysis of the pipeline.

### Introduction tab

An application overview is shown in this tab, including a “note” indicating the precaution to be taken in the interpretation and use of the results obtained from this application.

### Results tabs

The results of the analyses are shown in seven tabs. In the first one, “ranking map”, a map of the areas in the region of study is shown, which displays in colour only those areas that are relevant according to the ranking criteria of the pipeline describe before (areas with PP higher or equal to 80% and ranked first by the DOE and second by the WRR). Moreover, the colour scale used in the map (see the legend) corresponds to the sextiles of the absolute measures of risk (DOE) of the spatial areas. The numbers shown inside the spatial units correspond to the identification code ID (see description of the variables included in the ranking table tab). The application allows users to zoom in on the

map and also download it in pdf format.

In the second tab, “ranking table”, a table is shown containing the relative and absolute risk estimates for the coloured areas in the map of the previous tab as well as information about whether these areas are located within a spatial cluster (the application allows users to download the table in XLSX format). Specifically, the estimators and variables shown in the table are:

- ❖ Ranking: ranking position (number 1 means the highest score according to two criteria based on the excess of cases/deaths and a WRR, see below);
- ❖ Highest level of importance;
- ❖ ID: identification code of the spatial areas;
- ❖ Population: total population in each spatial area;
- ❖ Obs: observed cases in each spatial area;
- ❖ Exp: expected cases in each spatial area;
- ❖ SMR: standard mortality/morbidity ratio (observed/expected);
- ❖ Diff\_obs\_exp: difference between observed and expected (DOE) cases;
- ❖ RR: relative risk;
- ❖ lCre: lower limit of the RR credible interval;
- ❖ uCre: upper limit of the RR credible interval;
- ❖ PP: posterior probability (PP) that  $RR > 1$ ;
- ❖ cluster\_1 to cluster\_n: cluster membership (labels: cluster = belongs to a cluster; center = belongs to a cluster and is the central area).

Only the significant clusters by Kulldorff’s test ( $P$  value  $\leq 0.05$ ) are shown and the order of appearance in the table is established according to Kulldorff’s statistic, so that the cluster that present the highest value is the first (cluster\_1). There are some aspects related to multiple comparisons that are not taken into account in this section, see (31). The table is sorted based on the values of the ranking variable, but users can sort this table by any of the other variables such as, for example, RRs.

In the third tab, “spatial tests”, the results from the spatial tests described in the pipeline section of the manuscript are shown. The application displays exactly the output given by R software.

In the fourth and fifth tabs, “RR map” and “PP map” respectively, RRs (smoothed SMRs) and the distributions of PPs of having an  $RR > 1$  are mapped.

In the sixth tab, “cluster analysis”, the results from the cluster analysis are shown according to the criteria described in the pipeline section of the manuscript. On the left side of the tab, a table with the significant clusters detected using Kulldorff’s test ( $P$  value  $\leq 0.05$ ) is shown, and

the order of appearance in the table is established according to Kulldorff’s statistic, so that the cluster that presents the highest value appears first (cluster\_1). Specifically, the estimators and variables shown in the table are:

- ❖ Cluster: identification (as a numerical code) for the cluster;
- ❖ Size: number of spatial units included in the cluster;
- ❖ Statistic: value of the Kulldorff’s statistic;
- ❖ P value: P value from Kulldorff’s test.

Two maps are also shown in this tab, displaying in colour those spatial units belonging to the cluster 1 and 2 which are described in the table on the left side of the tab. The spatial units which correspond to the center of the shown clusters are coloured in a darker grey than the rest.

In the seventh tab, “complete results table”, a table is shown containing the relative and absolute risk estimates for all areas in the map as well as information about whether these areas are located within a spatial cluster (the application allows users to download the table in XLSX format). The estimators and variables shown are the same as those in the table in second tab.

Finally, all the maps are available support zooming. To zoom in, users must select the area pressing the left mouse button and then make a left double click over the area selected. To zoom out, users should make left double click over the map.

### Glossary tab

A description of the content of each of the application tabs is shown here.

### Example with simulated data

The simulated data (“data\_example.csv” and “map\_example.zip”) mentioned before is based on observed and expected cases in 314 spatial areas and a shapefile (see “Code and data availability” section of the manuscript). These data have been analyzed using RANKSPA application using the default value for the fraction of the total population parameter (15%). The results obtained are shown in the “Results” section of the manuscript.

### Example with real data

We have tested the pipeline on real data using the RANKSPA application. It consisted of individual death entries (observed cases) between 2005 and 2009 in men in Galicia [North-East of Spain, a region where there is a clear spatial pattern of lung cancer mortality (8)] corresponding to lung cancer (International Classification of Diseases,

10th Revision: C61), broken down by municipality (n=314). Population data (men) by town and age (18 age groups) in 2007, the midpoint of the study period, was obtained from the Spanish municipal Register. The National Statistics Institute [Instituto Nacional de Estadística (INE)] of Spain provided deaths and population data. The expected cases were calculated by multiplying the overall Spanish age-specific mortality rates (using lung cancer death entries between 2005 and 2009 in men for the whole country of Spain) for the 5-year study period by each town's person-years (2007 population multiplied by 5). Afterwards, SMRs were computed as the ratios of the observed to the expected deaths.

We selected lung cancer due to the high mortality and incidence of this chronic disease. Lung cancer remains the leading cause of cancer incidence and mortality, with 2.1 million new lung cancer cases (11.6% of the total cases) and 1.8 million deaths predicted in 2018 (18.4% of the total cancer deaths), representing close to 1 in 5 cancer deaths (32) which makes it possible to have reliable indicators. Moreover, lung cancer survival is poor, with a relative survival in Europe of 39% and 13% at 1 and 5 years since diagnosis, respectively (33). In that way, lung cancer mortality is a good indicator for monitoring this disease.

The results obtained using RANKSPA application to this data, using the default value for the fraction of the total population parameter, are shown in "Results" section of the manuscript.

The study was conducted in accordance with the Declaration of Helsinki, and ethical approval was not required due to the type of study and design carried out.

## Results

### *Simulated case study*

The results displayed in the tabs of RANKSPA application using simulated data provided (see "Code and data availability" section of the manuscript) are shown in *Figure 3*. The tables are shown in the "ranking table" and "complete results table" tabs are included in the [Tables S1,S2](#).

According to these results, there is a clear excess of mortality in the middle-west of the studied area (see *Figure 3C,D*) where a spatial mortality cluster is also located (see *Figure 3E*). Moreover, there are five spatial units (IDs: 65, 293, 152, 305 and 156) located outside this cluster that occupy the top positions in the ranking generated by the application (see *Figure 3A,B*).

According to the pipeline that implements the RANKSPA application, among the total spatial units of the study region [314], only 14 have excess mortality whose PP are greater than or equal to 80%. All spatial tests implemented with the exception of the Moran's I test of spatial autocorrelation provided statistically significant results ( $P<0.05$ ).

Finally, in *Figure 4* a scatterplot is shown displaying the relationship between DOE and WRR for the simulated data in each spatial unit, and indicating whether the PPs were above or below 80%, which was one of the criteria to be taken into account in the pipeline described above. It can be seen that there is in general a positive correlation between absolute (DOE) and relative (WRR) measures. However, a greater dispersion is observed when these measures take the highest values.

### *Real case study*

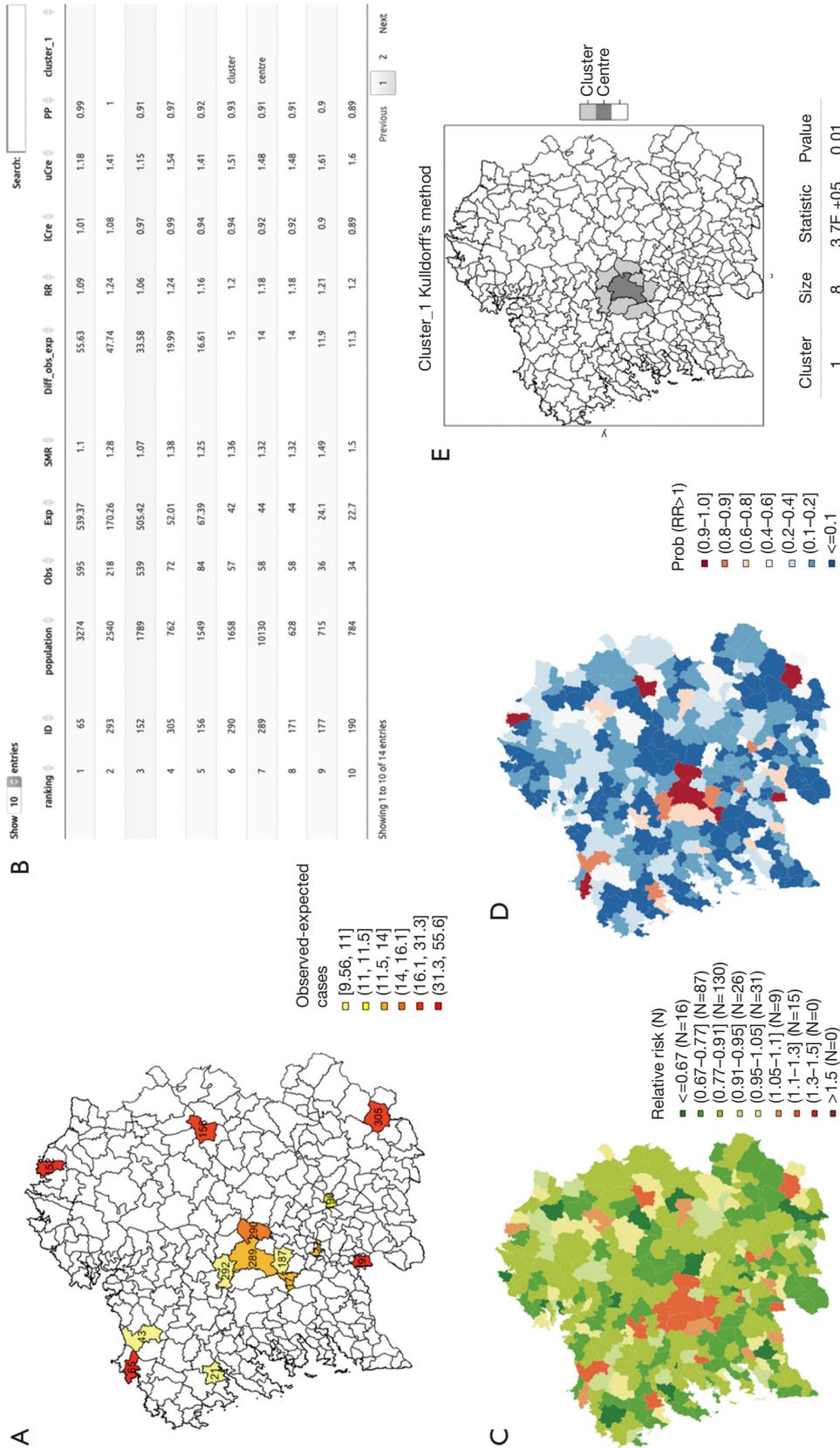
The results displayed in the tabs of RANKSPA application using real data are shown in *Figure 5*. According to them, there is a clear excess of mortality in the west of the region studied (see *Figure 5C,D*) where several of the spatial mortality clusters are also located (see *Figure 5E*). There are three spatial units (IDs: 279, 53 and 33) located outside these clusters that occupy the top positions in the ranking generated by the application (see *Figure 5A,B*).

Of the total spatial units of the study region [314], only 11 have excess mortality whose PP are greater than or equal to 80%. All spatial tests implemented provided statistically significant results ( $P<0.05$ ).

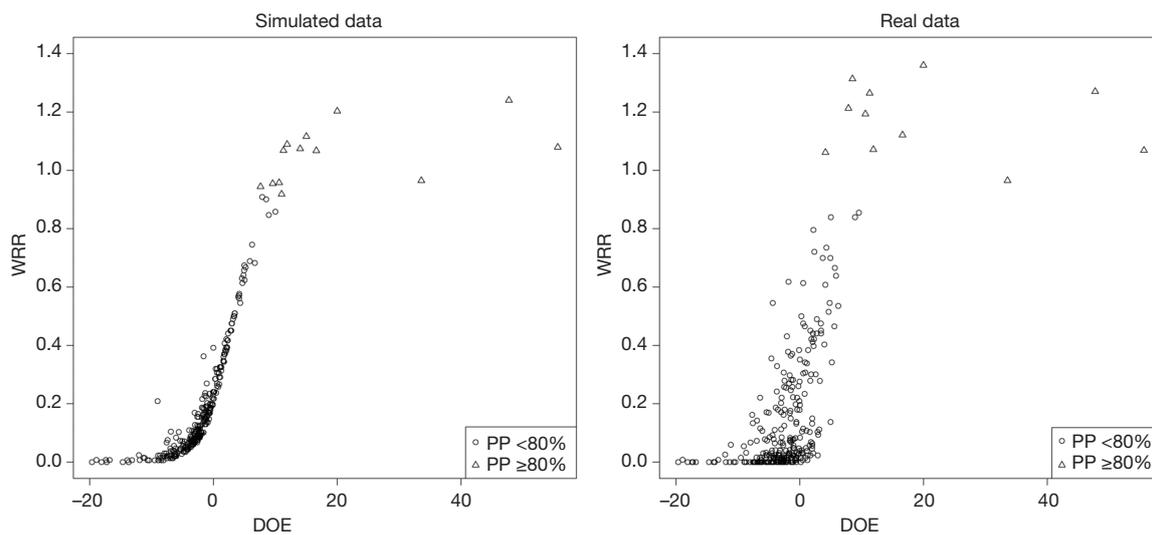
Finally, in *Figure 4* a scatterplot is shown displaying the relationship between DOE and WRR also for real data in each spatial unit, and indicating whether the PP are above or below 80%, which was one of the criteria to be taken into account in the pipeline described above. It can be seen, as it was the case with the simulated data, that there is in general a positive correlation between absolute (DOE) and relative (WRR) measures. However, there is some dispersion when these measurements take the highest values.

## Discussion

In this paper, we present a pipeline of analysis that could allow us to discriminate regions of interest in the context of chronic diseases monitoring (like cancer monitoring) by studying the spatial distribution in small areas of mortality/morbidity. Taking relative and absolute risk estimators



**Figure 3** Results displayed in the tabs of RANKSPA application using simulated data. (A) Ranking map tab; (B) ranking table tab; (C) RR map tab; (D) PP map tab; (E) cluster analysis tab. RR, relative risk; PP, posterior probability.



**Figure 4** Scatterplots displaying the relationship between WRR and the DOE cases in simulated and real data, indicating when the PPs were above or below 80%. WRR, weighted relative risk; DOE, difference between observed and expected; PP, posterior probability.

and the uncertainty about them into account, it obtains a ranking of spatial units, thus being able to implement an initial spatial screening for exploratory purposes useful in epidemiological surveillance and public health. Moreover, we introduce the RANKSPA app, an R-Shiny web application to implement this strategy in a simple way, fast and without a deep technical knowledge of the statistical and programming methods necessary to perform it. It also serves as an exploratory tool for spatial analysis since it enables visualization of maps and the detection of clusters.

As already indicated in the Introduction section, in the context of public health and surveillance of chronic diseases such as cancer, the representation and analysis of maps of events (disease mapping) within a fixed time frame/period is a basic tool and the existence of criteria to discriminate relevant results is important. Furthermore, in many cases, this process should not only take into account relative measures, but also absolute measures of risk. The shown pipeline tries to incorporate and integrate different aspects already known in the field of space surveillance, combining these risk measures.

Here we have focused on the DOE cases and the WRR to rank areas. However, this is a first approach and other measures that take into account the size of the region could be used as the DOE will likely depend on the population of the areas. For this reason, other relative measures that take into account the size of the areas (such as Chi-square distances between observed and expected) and their

associated uncertainty could be considered when ranking the areas. Our preliminary assessment of such measures indicates that the ranking is very similar to that provided by the DOE.

On the other hand, the RANKSPA app, that is an open-source tool implemented using R, Shiny and incorporating the functions from several R packages, is easy to use and allows health researchers to perform all the analyses of the pipeline without the need of having advanced statistical or programming skills. As the methods for the initial screening that the pipeline implements are very common in surveillance, researchers that need to perform more complex analyses should use other statistical packages or other more complete and complex applications like SpatialEpiApp (4).

In relation to the application of the RANKSPA App to data, it is worth mentioning that the extent of the utility of the results it provides may depend largely on spatial heterogeneity. When this is very large or there is no spatial heterogeneity, it can be expected that the discrimination of areas of interest through the described pipeline is not so easy to interpret or useful at all. Moreover, in which ranges of heterogeneity may be useful will also depend on the spatial distribution of the disease. Therefore, the results obtained when using the RANKSPA application must be carefully evaluated taking into account all these aspects. For reasons of time and space in this work no results are presented analyzing other possible situations. Further studies are necessary to address other types of scenarios



where changes in the pipeline could surely be found and proposed to provide a useful response in these contexts.

In relation to the proposed pipeline, some assumptions have been made regarding the neighbourhood between spatial units and the priors in spatial models that, although they are among the most used in studies about the spatial distribution of diseases such as cancer, may not be the most suitable in certain contexts. For example, rare diseases for which the observed cases are very few or even zero in most areas. Therefore, the results obtained must also be carefully evaluated taking into account all these aspects.

The real data used of lung cancer mortality represent a good example where the results may be useful in monitoring this disease. There is some spatial heterogeneity, with a clear and located spatial pattern. However, at the highest values of the risk measures, the ranking of the space units depends very much on which measure of risk (absolute as implemented directly in the RANKSPA application) or relative (the table shown in the RANKSPA app allows this visualization quickly) is taken into account first.

Finally, we have tried to design both the pipeline and the application to allow for a quick and easy-to-interpret exploratory analysis. We plan to incorporate additional methodological approaches (such as spatio-temporal analysis, cut-off combinations for posterior probabilities, other types of neighbourhoods, model settings, ...) or other technical aspects related to the flexibility of data loading and interaction with the information provided, to enhance the work presented here.

As a general conclusion, the work presented shows a strategy of analysis to implement an initial screening of disease risk for exploratory purposes that allows to discriminate regions of interest by studying the spatial distribution in small areas, focused primarily on chronic diseases such as cancer, where the proposed methods are commonly used in their monitoring. Moreover, an R-Shiny application (RANKSPA) that has been created to be able to implement this strategy in a simple way, fast and without a deep technical knowledge of the statistical and programming methods necessary to perform it. However, due to the limitations of the implemented methodology, the results obtained should be treated with caution, and in no case be used as the only method for decision-making. See, for example (34), for a recent review of sound methods for spatial epidemiology and the detection of regions of high risk. Nevertheless, for an initial exploratory objective and/or to obtain in a quick way information in the context of health management, RANKSPA may be more useful than

other applications.

#### *Code and data availability*

- ❖ Pipeline R-code: R\_script.R (<https://doi.org/10.6084/m9.figshare.11635908>)
- ❖ RANKSPA R-Shiny code: RANKSPA.zip (<https://doi.org/10.6084/m9.figshare.11635935>)
- ❖ Simulated data: data\_example.csv (<https://doi.org/10.6084/m9.figshare.11467749>)
- ❖ Shapefiles: map\_example.zip (<https://doi.org/10.6084/m9.figshare.11467755>)

#### *Software availability*

<https://biodama.shinyapps.io/rankspa/>

#### **Acknowledgments**

*Funding:* The study was partially supported by research grants from the Spanish Health Research Fund (FIS PI17CIII/00040), by grants PPIC-2014-001-P and SBPLY/17/180501/000491, funded by Consejería de Educación, Cultura y Deportes (Junta de Comunidades de Castilla-La Mancha, Spain) and FEDER, grant PID2019-106341GB-I00 from Ministerio de Ciencia e Innovación (Spain) and grant MTM2016-77501-P, funded by Ministerio de Economía y Competitividad (Spain). FPP has been supported by a Ph.D. scholarship awarded by the University of Castilla-La Mancha (Spain).

#### **Footnote**

*Provenance and Peer Review:* This article was commissioned by the Guest Editors (Peter Baade and Susanna Cramb) for the series “Spatial Patterns in Cancer Epidemiology” published in *Annals of Cancer Epidemiology*. The article has undergone external peer review.

*Data Sharing Statement:* Available at <http://dx.doi.org/10.21037/ace-20-15>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/ace-20-15>). The series “Spatial Patterns in Cancer Epidemiology” was commissioned by the editorial office without any funding or sponsorship. The authors have no other conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), and ethical approval was not required due to the type of study and design carried out.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Lawson AB, Kleinman K. Spatial and syndromic surveillance for public health. Chichester: John Wiley & Sons; 2005:288.
2. Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to spatial data analysis. *Geogr Anal* 2006;38:5-22.
3. Adin A, Goicoa T, Ugarte MD. Online relative risks/rates estimation in spatial and spatio-temporal disease mapping. *Comput Methods Programs Biomed* 2019;172:103-16.
4. Moraga P. SpatialEpiApp: A Shiny web application for the analysis of spatial and spatio-temporal disease data. *Spat Spatiotemporal Epidemiol* 2017;23:47-57.
5. Richardson S, Thomson A, Best N, et al. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* 2004;112:1016-25.
6. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect* 2004;112:998-1006.
7. Lawson AB, Biggeri A, Böhning D, et al. editors. Disease mapping and risk assessment for public health. 1st ed. Chichester: Wiley; 1999:502.
8. López-Abente G, Aragonés N, Pérez-Gómez B, et al. Time trends in municipal distribution patterns of cancer mortality in Spain. *BMC Cancer* 2014;14:535.
9. Rodríguez-Sánchez L, Fernández-Navarro P, López-Abente G, et al. Different spatial pattern of municipal prostate cancer mortality in younger men in Spain. *PLoS One* 2019;14:e0210980.
10. Baltrus P, Malhotra K, Rust G, et al. Identifying county-level all-cause mortality rate trajectories and their spatial distribution across the United States. *Prev Chronic Dis* 2019;16:E55.
11. Jiang F, Chu J, Chen X, et al. Spatial distribution and clusters of pancreatic cancer mortality in Shandong Province, China. *Sci Rep* 2019;9:12917.
12. Duncan EW, Cramb SM, Aitken JF, et al. Development of the Australian Cancer Atlas: spatial modelling, visualisation, and reporting of estimates. *Int J Health Geogr* 2019;18:21.
13. López-Abente G, Ramis R, Pollán M, et al. Atlas municipal de mortalidad por cáncer en España 1989-1998. Madrid: Área de Epidemiología Ambiental y Cáncer del Centro Nacional de Epidemiología, ISCIII; 2007.
14. Pickle LW; National Center for Health Statistics (U.S.). Atlas of United States mortality. Hyattsville: National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Dept. of Health and Human Services; 1996. Available online: <http://catalog.hathitrust.org/api/volumes/oclc/36045997.html>
15. Boyle P, Smans M. Atlas of Cancer Mortality in the European Union and the European Economic Area, 1993-1997. Available online: <https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Atlas-Of-Cancer-Mortality-In-The-European-Union-And-The-European-Economic-Area-1993-1997-2008>
16. Bivand RS, Wong DWS. Comparing implementations of global and local indicators of spatial association. *TEST* 2018;27:716-48.
17. Bivand RS, Pebesma E, Gómez-Rubio V. Applied spatial data analysis with R. New York: Springer; 2013:415.
18. Gómez-Rubio V, Ferrándiz-Ferragud J, López-Quílez A. Detecting clusters of disease with R. *J Geogr Syst* 2005;7:189-206.
19. Moran PA. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17-23.
20. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat Med* 1995;14:2323-34.
21. Tango T. A test for spatial disease clustering adjusted for multiple testing. *Stat Med* 2000;19:191-204.
22. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med* 1995;14:799-810.
23. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 1991;43:1-20.
24. López-Abente G, García-Gómez M, Menéndez-Navarro A, et al. Pleural cancer mortality in Spain: time-trends

- and updating of predictions up to 2020. *BMC Cancer* 2013;13:528.
25. Rue H, Martino S, Copin N. Approximate Bayesian Inference for latent Gaussian models using integrated nested laplace approximations. *J R Stat Soc Ser B* 2009;71:319-92.
  26. Carroll R, Lawson AB, Faes C, et al. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spat Spatiotemporal Epidemiol* 2015;14-15:45-54.
  27. Rue H, Riebler A, Sørbye SH, et al. Bayesian computing with INLA: a review. *Annu Rev Stat Appl* 2017;4:395-421.
  28. De Smedt T, Simons K, Van Nieuwenhuysse A, et al. Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models. *Arch Public Health* 2015;73:O2.
  29. The R-Inla Package. Available online: <http://www.r-inla.org/download>
  30. Gómez-Rubio V. Bayesian inference with INLA. Boca Raton: CRC Press, 2020.
  31. Gómez-Rubio V, Moraga P, Molitor J, et al. DClusterM: Model-based detection of disease clusters. *J Stat Softw* 2019;90:1-26.
  32. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
  33. De Angelis R, Sant M, Coleman MP, et al. Cancer survival in Europe 1999-2007 by country and age: results of EURO CARE--5-a population-based study. *Lancet Oncol* 2014;15:23-34.
  34. Lawson AB, Banerjee S, Haining RP, et al. editors. *Handbook of Spatial Epidemiology*. 1st ed. New York: Chapman and Hall/CRC; 2016.

doi: 10.21037/ace-20-15

**Cite this article as:** Fernández-Navarro P, González-Palacios J, González-Sánchez M, Ramis R, Nuñez O, Palmí-Perales F, Gómez-Rubio V. Ranking spatial areas by risk of cancer: modelling in epidemiological surveillance. *Ann Cancer Epidemiol* 2020;4:10.

## Supplementary

**Table S1** Table shown in the “ranking table” tab of RANKSPA application using non-real data

Ranking	ID	Population	Obs	Exp	SMR	Diff_obs_exp	RR	ICre	uCre	PP	cluster_1
1	65	3,274	595	539.37	1.1	55.63	1.09	1.01	1.18	0.99	
2	293	2,540	218	170.26	1.28	47.74	1.24	1.08	1.41	1	
3	152	1,789	539	505.42	1.07	33.58	1.06	0.97	1.15	0.91	
4	305	762	72	52.01	1.38	19.99	1.24	0.99	1.54	0.97	
5	156	1,549	84	67.39	1.25	16.61	1.16	0.94	1.41	0.92	
6	290	1,658	57	42	1.36	15	1.2	0.94	1.51	0.93	Cluster
7	289	10,130	58	44	1.32	14	1.18	0.92	1.48	0.91	Centre
8	171	628	58	44	1.32	14	1.18	0.92	1.48	0.91	
9	177	715	36	24.1	1.49	11.9	1.21	0.9	1.6	0.9	
10	190	784	34	22.7	1.5	11.3	1.2	0.89	1.6	0.89	
11	292	3,272	61	50	1.22	11	1.12	0.88	1.39	0.82	Cluster
12	187	49,807	61	50	1.22	11	1.12	0.88	1.39	0.82	Cluster
13	43	14,689	46	35.39	1.3	10.61	1.14	0.87	1.47	0.84	
14	21	3,671	33	23.44	1.41	9.56	1.15	0.85	1.54	0.83	

SMR, standardized mortality ratio; RR, relative risk; PP, posterior probability.

Table S2 Table shown in the "complete results table" tab of RANKSPA application using non-real data

Ranking	ID	Population	Obs	Exp	SMR	Diff_obs_exp	RR	ICre	uCre	PP	cluster_1
1	65	3,274	595	539.37	1.1	55.63	1.09	1.01	1.18	0.99	
2	293	2,540	218	170.26	1.28	47.74	1.24	1.08	1.41	1	
3	152	1,789	539	505.42	1.07	33.58	1.06	0.97	1.15	0.91	
4	305	762	72	52.01	1.38	19.99	1.24	0.99	1.54	0.97	
5	156	1,549	84	67.39	1.25	16.61	1.16	0.94	1.41	0.92	
6	290	1,658	57	42	1.36	15	1.2	0.94	1.51	0.93	Cluster
7	289	10,130	58	44	1.32	14	1.18	0.92	1.48	0.91	Centre
8	171	628	58	44	1.32	14	1.18	0.92	1.48	0.91	
9	177	715	36	24.1	1.49	11.9	1.21	0.9	1.6	0.9	
10	190	784	34	22.7	1.5	11.3	1.2	0.89	1.6	0.89	
11	292	3,272	61	50	1.22	11	1.12	0.88	1.39	0.82	Cluster
12	187	49,807	61	50	1.22	11	1.12	0.88	1.39	0.82	Cluster
13	43	14,889	46	35.39	1.3	10.61	1.14	0.87	1.47	0.84	
14	21	3,671	33	23.44	1.41	9.56	1.15	0.85	1.54	0.83	
15	1	2,246	5	6.07	0.82	-1.07	0.83	0.55	1.24	0.19	
15	2	4,766	9	11.54	0.78	-2.54	0.81	0.55	1.17	0.14	
15	3	11,942	1	3.31	0.3	-2.31	0.77	0.49	1.17	0.11	
15	4	8,978	2	5.63	0.36	-3.63	0.74	0.48	1.11	0.07	
15	5	1,422	28	27.43	1.02	0.57	0.94	0.69	1.26	0.34	
15	6	3,449	5	4.21	1.19	0.79	0.89	0.58	1.35	0.29	
15	7	3,112	19	25.99	0.73	-6.99	0.78	0.56	1.06	0.06	
15	8	2,401	8	13.43	0.6	-5.43	0.75	0.51	1.07	0.06	
15	9	929	226	235.04	0.96	-9.04	0.95	0.84	1.08	0.22	
15	10	928	1	3.67	0.27	-2.67	0.75	0.48	1.16	0.1	
15	11	432	8	10.97	0.73	-2.97	0.8	0.54	1.16	0.12	
15	12	1,864	5	5.52	0.91	-0.52	0.85	0.55	1.27	0.21	
15	13	2,514	12	18.31	0.66	-6.31	0.75	0.52	1.06	0.05	
15	14	1,603	5	5.52	0.91	-0.52	0.85	0.55	1.27	0.21	
15	15	1,784	3	11.93	0.25	-8.93	0.64	0.42	0.94	0.01	
15	16	2,479	12	18.39	0.65	-6.39	0.75	0.52	1.06	0.05	
15	17	4,812	19	14.16	1.34	4.84	1.05	0.74	1.47	0.61	
15	18	2,775	7	8.06	0.87	-1.06	0.84	0.56	1.24	0.19	
15	19	3,087	2	8.32	0.24	-6.32	0.68	0.44	1.01	0.03	
15	20	549	5	4.56	1.1	0.44	0.88	0.57	1.33	0.23	
15	22	4,342	4	4.13	0.97	-0.13	0.85	0.55	1.3	0.27	
15	23	1,600	1	3.55	0.28	-2.55	0.76	0.48	1.16	0.1	
15	24	5,645	4	4.58	0.87	-0.58	0.84	0.54	1.27	0.2	
15	25	13,455	7	14.41	0.49	-7.41	0.7	0.47	1	0.03	
15	26	2,386	10	7.97	1.25	2.03	0.95	0.64	1.4	0.4	
15	27	5,145	11	7.65	1.44	3.35	1	0.88	1.47	0.5	
15	28	43,735	15	20.04	0.75	-5.04	0.8	0.56	1.1	0.09	
15	29	8,542	6	9.49	0.63	-3.49	0.77	0.52	1.13	0.09	
15	30	253	8	13.53	0.59	-5.53	0.75	0.51	1.07	0.05	
15	31	5,646	9	19.68	0.46	-10.68	0.67	0.46	0.94	0.01	
15	32	538	5	8.32	0.6	-3.32	0.77	0.51	1.14	0.1	
15	33	113,830	6	9.22	0.65	-3.22	0.78	0.52	1.15	0.1	
15	34	416	19	20.7	0.92	-1.7	0.88	0.63	1.2	0.21	
15	35	2,898	2	8.71	0.23	-6.71	0.68	0.44	1	0.02	
15	36	2,785	12	10.72	1.12	1.28	0.93	0.64	1.34	0.35	
15	37	1,903	2	4.84	0.41	-2.84	0.76	0.49	1.15	0.1	
15	38	954	9	9.21	0.98	-0.21	0.88	0.59	1.28	0.25	
15	39	1,495	6	6.58	0.91	-0.58	0.85	0.56	1.27	0.22	
15	40	2,499	7	7.54	0.93	-0.54	0.86	0.57	1.27	0.23	
15	41	1,015	9	6.99	1.29	2.01	0.95	0.63	1.41	0.4	
15	42	11,108	13	10.01	1.3	2.99	0.99	0.68	1.43	0.48	
15	44	1,032	2	10.45	0.19	-8.45	0.64	0.42	0.94	0.01	
15	45	1,492	13	10.27	1.27	2.73	0.98	0.67	1.41	0.46	
15	46	3,737	10	13.09	0.76	-3.09	0.81	0.55	1.15	0.12	
15	47	707	9	10.16	0.89	-1.16	0.85	0.57	1.23	0.2	
15	48	1,136	10	18.63	0.54	-8.63	0.7	0.49	0.99	0.02	
15	49	815	177	173.74	1.02	3.26	1	0.87	1.15	0.5	
15	50	3,652	8	5.83	1.37	2.17	0.95	0.63	1.43	0.41	
15	51	13,019	2	5.4	0.37	-3.4	0.75	0.48	1.13	0.08	
15	52	6,897	5	5.64	0.89	-0.64	0.84	0.55	1.27	0.21	
15	53	35,080	21	18.76	1.12	2.24	0.97	0.69	1.34	0.43	
15	54	1,729	0	1.87	0	-1.87	0.77	0.48	1.2	0.13	
15	55	835	18	21	0.86	-3	0.85	0.6	1.16	0.15	
15	56	2,615	10	9.44	1.06	0.56	0.9	0.61	1.32	0.3	
15	57	577	6	12.95	0.46	-6.95	0.7	0.47	1.01	0.03	
15	58	2,748	3	4.4	0.68	-1.4	0.81	0.52	1.23	0.16	
15	59	17,639	7	14.57	0.48	-7.57	0.7	0.47	1	0.03	
15	60	2,757	4	12.89	0.31	-8.89	0.65	0.43	0.95	0.01	
15	61	15,375	15	16.51	0.91	-1.51	0.87	0.61	1.22	0.21	
15	62	3,526	22	28.39	0.78	-6.39	0.8	0.58	1.08	0.07	
15	63	6,137	39	46.41	0.84	-7.41	0.84	0.64	1.06	0.08	
15	64	1,461	49	55.9	0.88	-6.9	0.87	0.68	1.09	0.12	
15	66	638	17	34.27	0.5	-17.27	0.64	0.46	0.86	0	
15	67	1,171	7	8.56	0.82	-1.56	0.83	0.55	1.22	0.17	
15	68	9,208	9	11.08	0.81	-2.08	0.83	0.56	1.19	0.15	
15	69	3,401	15	21.59	0.69	-6.59	0.77	0.54	1.06	0.05	
15	70	666	22	17.87	1.23	4.13	1.02	0.73	1.41	0.56	
15	71	725	19	16.69	1.14	2.31	0.97	0.69	1.35	0.43	
15	72	1,123	30	31.44	0.95	-1.44	0.91	0.67	1.2	0.25	
15	73	3,228	5	11.19	0.45	-6.19	0.71	0.47	1.03	0.04	
15	74	5,754	9	12.44	0.72	-3.44	0.79	0.54	1.14	0.11	
15	75	684	1	4.01	0.25	-3.01	0.74	0.47	1.14	0.09	
15	76	5,154	5	5.5	0.91	-0.5	0.85	0.56	1.27	0.22	
15	77	2,700	5	5.01	1	-0.01	0.87	0.57	1.31	0.25	
15	78	2,095	0	2.64	0	-2.64	0.75	0.47	1.15	0.09	
15	79	879	41	58.4	0.7	-17.4	0.74	0.57	0.94	0.01	
15	80	6,579	2	5.21	0.38	-3.21	0.75	0.48	1.13	0.09	
15	81	3,877	5	8.03	0.62	-3.03	0.78	0.51	1.15	0.11	
15	82	1,537	14	16.63	0.84	-2.63	0.84	0.59	1.17	0.15	
15	83	1,215	32	33.82	0.95	-1.82	0.9	0.68	1.19	0.24	
15	84	2,374	15	7.4	2.03	7.6	1.18	0.8	1.73	0.8	
15	85	3,678	3	5.04	0.59	-2.04	0.79	0.51	1.19	0.13	
15	86	3,280	26	21.33	1.22	4.67	1.04	0.75	1.4	0.59	
15	87	1,085	6	6.3	0.95	-0.3	0.86	0.57	1.28	0.23	
15	88	1,136	5	7.52	0.66	-2.52	0.79	0.52	1.17	0.12	
15	89	754	16	17.02	0.94	-1.02	0.88	0.62	1.23	0.23	
15	90	2,119	29	28.64	1.01	0.36	0.94	0.69	1.25	0.34	
15	91	1,472	8	11.63	0.69	-3.63	0.79	0.53	1.13	0.1	
15	92	1,732	6	7.07	0.85	-1.07	0.84	0.55	1.24	0.19	
15	93	2,184	3	5.55	0.54	-2.55	0.78	0.5	1.17	0.11	
15	94	1,551	2	9.01	0.22	-7.01	0.67	0.44	0.99	0.02	
15	95	1,081	21	20.45	1.03	0.55	0.93	0.67	1.27	0.33	
15	96	857	7	9.12	0.77	-2.12	0.82	0.55	1.19	0.15	
15	97	1,406	11	6.42	1.71	4.58	1.05	0.7	1.56	0.6	
15	98	4,437	2	12.42	0.16	-10.42	0.61	0.4	0.9	0.01	
15	99	665	0	0.92	0	-0.92	0.8	0.5	1.26	0.17	
15	100	2,402	17	15.28	1.11	1.72	0.95	0.67	1.33	0.39	
15	101	2,306	23	25.5	0.9	-2.5	0.87	0.63	1.18	0.19	
15	102	1,625	6	8.21	0.73	-2.21	0.81	0.54	1.19	0.14	
15	103	1,894	9	5.02	1.79	3.98	1.03	0.68	1.55	0.55	
15	104	2,776	9	7.4	1.22	1.6	0.93	0.63	1.38	0.37	
15	105	1,353	2	3.12	0.64	-1.12	0.81	0.52	1.24	0.17	
15	106	1,678	23	34.33	0.67	-11.33	0.74	0.54	0.98	0.02	
15	107	1,231	33	49.79	0.66	-16.79	0.72	0.54	0.93	0.01	
15	108	1,608	2	5.52	0.36	-3.52	0.74	0.48	1.12	0.08	
15	109	2,234	9	11.92	0.76	-2.92	0.81				