European Journal of Public Health, 2024, ckae104

https://doi.org/10.1093/eurpub/ckae104

Impact of the use of small-area models on estimation of attributable mortality at a regional level

Julia Rey-Brandariz [1,2], María I. Santiago-Pérez [3], Cristina Candal-Pedreira [1,2,*], Leonor Varela-Lema [1,2,4], Alberto Ruano-Ravina [1,2,4], Esther López-Vizcaíno [5], Carla Guerra-Tort [1], Jasjit S. Ahluwalia [6,7,8], Agustín Montes [1,2,4], Mónica Pérez-Ríos [1,2,4]

[1]Department of Preventive Medicine and Public Health, Universidade de Santiago de Compostela, Santiago de Compostela, Spain
[2]Consortium for Biomedical Research in Epidemiology and Public Health (CIBER en Epidemiología y Salud Pública/CIBERESP), Madrid, Spain
[3]Epidemiology Department, Directorate-General of Public Health, Galician Regional Health Authority, Santiago de Compostela, Spain
[4]Health Research Institute of Santiago de Compostela (Instituto de Investigación Sanitaria de Santiago de Compostela—IDIS), Santiago de Compostela, Spain
[5]Galician Statistics Institute, Santiago de Compostela, Spain
[6]Department of Behavioral and Social Sciences and Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, RI, United States
[7]Department of Medicine, Alpert Medical School, Brown University, Providence, RI, United States
[8]Legorreta Cancer Center, Brown University, Providence, RI, United States

*Corresponding author. Department of Preventive Medicine and Public Health, Medicine School, C/San Francisco s/n, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain. E-mail: cristina.candal.pedreira@rai.usc.es

Abstract

The objective of this study is to assess the impact of applying prevalences derived from a small-area model at a regional level on smoking-attributable mortality (SAM). A prevalence-dependent method was used to estimate SAM. Prevalences of tobacco use were derived from a small-area model. SAM and population attributable fraction (PAF) estimates were compared against those calculated by pooling data from three national health surveys conducted in Spain (2011–2014–2017). We calculated the relative changes between the two estimates and assessed the width of the 95% CI of the PAF. Applying surveys-based prevalences, tobacco use was estimated to cause 53 825 (95% CI: 53 182–54 342) deaths in Spain in 2017, a figure 3.8% lower obtained with the small-area model prevalences. The lowest relative change was observed in the Castile-La Mancha region (1.1%) and the highest in Navarre (14.1%). The median relative change between regions was higher for women (26.1%), population aged ≥65 years (6.6%), and cardiometabolic diseases (9.0%). The differences between PAF by cause of death were never greater than 2%. Overall, the differences between estimates of SAM, PAF, and confidence interval width are small when using prevalences from both sources. Having these data available by region will allow decision-makers to implement smoking control measures based on more accurate data.

Introduction

Smoking is one of the leading causes of death in developed countries, and to assess its impact at a population level, different indicators for monitoring have been developed. In Spain, there are several data sources that furnish information on smoking prevalence. Two of these are the National Health Survey (*Encuesta Nacional de Salud de España/ENSE*) [1] and the European Health Survey in Spain (*Encuesta Europea de Salud para España/EESE*) [2]. However, the capacity of national health surveys for furnishing prevalence of risk factors at a regional level is limited when it is estimated by sex and age. This is due to their sample size (approximately 23 000 persons/year in the National Health Survey and 22 000 persons/year in the European Health Survey in Spain), which does not allow for accurate estimates of regional tobacco use broken down by sociodemographic variables.

In Spain, the impact of tobacco use on mortality has been estimated, but these estimates have generally been made at a national level [3, 4]. In 2017, smoking-attributable mortality (SAM) was first estimated in the country's 17 autonomous regions by applying a prevalence-dependent method [5]. In order to obtain smoking prevalence at the autonomous regions level by sex and age group, with adequate precision, the information furnished by two National Health Surveys (2011–2012 and 2017) and one European Health Survey in Spain (2014) [2, 6] had to be analysed jointly. The drawback of this method is that the impact of pooling the three surveys on the finally estimated prevalence will vary depending on the evolution of the tobacco epidemic in each of the 17 autonomous regions. Hence, if the sex- and age-related prevalence in the respective autonomous regions is stable, the impact will be limited, but if there are changes in prevalence, these might be diluted as a consequence of the pooling of the surveys. As smoking prevalence is one of the data needed to estimate SAM, changes in this indicator will also affect estimates of attributed mortality. Thus, the problem with pooling surveys from different years is that they may be stabilizing increases and decreases in smoking prevalence, resulting in them overestimating or underestimating mortality attributed to tobacco use. In order to overcome this limitation, a small-area model was fitted, making it possible, for the first time in Spain, to obtain the prevalence of smokers, ex-smokers, and never-smokers for the 17 autonomous regions, by sex and five age groups [7]. This model is based on aggregated data from a single health survey and smoking-related auxiliary variables sourced from administrative records. These auxiliary variables included educational level, degree of urbanization, percentage population living on the coast or in the interior, occupation, and percentage of hospitalizations due to lung cancer or chronic obstructive pulmonary disease (COPD), among other variables [7].

Thus, the dual aim of this study was: first, to estimate SAM in 2017, along with its 95% confidence interval (95% CI), by sex and age group in the 17 autonomous regions, applying smoking prevalence based on a small-area model; and second, to compare these results against the previous SAM estimate obtained for the 17 autonomous regions in 2017 using prevalence derived from the pooling of three national health surveys.

Methods

Calculation procedure

To estimate SAM, we used a prevalence-dependent method based on the calculation of population attributable fractions (PAFs), in accordance with the STREAMS-P tool guidelines [8]. The PAF is based on knowledge of the prevalence of smokers, ex-smokers, and never-smokers, and the risk of smokers and ex-smokers to develop tobacco-related disease as compared with never smokers [9]. Once the PAFs have been calculated, these are then multiplied by the observed mortality to get the SAM. More detailed information on the calculation methodology can be found in a previous publication [5].

Data sources

Observed mortality was obtained from the National Statistics Institute of Spain for 2017, by sex, age group, and smoking-related causes of death according to the most recent Surgeon General's report [10]. Observed all-cause mortality deaths by sex, along with their International Classification of Diseases-10th Revision (ICD-10) codes, are shown in Supplementary Tables S1 and S2. The relative risks (RRs) for smokers and ex-smokers as compared with never smokers, by cause of death, sex, and age group, come from the follow-up of five US cohorts [10], which involved almost 1 million people across the period 2000–2010. The RRs used are listed in Supplementary Table S3. The prevalence of smokers, ex-smokers, and never-smokers by autonomous region, sex, and age group was estimated using a mixed multinomial logistic model with random area effects, applying data from the 2017 National Health Survey. This model relates aggregated data on smoking by autonomous region, sex, and age group (areas) with auxiliary information at an area level sourced from different administrative records. From now on, this model will be referred to as the small-area model. In both the nine editions of the National Health Survey published between 1987 and 2017 and the three editions of the European Health Survey in Spain published between 2009 and 2020, the question to define tobacco use was '*Do you currently smoke*?' A smoker was defined as a person who answered '*yes, I smoke daily*' or '*yes, I smoke, but not daily*'; an ex-smoker was defined as a person who answered '*I do not smoke, but I have smoked*'; and a never smoker was defined as a person who answered '*I do not smoke, nor have I ever smoked on a regular basis*'. The prevalence obtained with the small-area model is shown in Supplementary Table S4. More detailed information on the methodology of calculating the prevalence can be found in a previous study [7].

Ethical approval was not required as this study is not conducted on human participants including identifiable human material or identifiable data and no human intervention is performed.

Analysis

The PAFs and SAM were estimated for each autonomous region by sex, age group (35–64, 65 years and over), and large groups of causes of death (tumors, cardiometabolic, and respiratory diseases). The SAM and PAF estimates were accompanied by their 95% CIs. The 95% CI of the PAF was calculated using a naïve bootstrap procedure, and based on this, the CI of SAM was obtained. As a first step, we generated a bootstrap sample of 1000 replications of the prevalence of smokers, ex-smokers, and never smokers, and on the basis of this sample, then obtained a bootstrap sample of PAF, assuming that the RRs were measured without sampling error. The CI of the PAF was calculated using the Efron percentile method, so that the 95% CI limits are the 2.5th and 97.5th percentiles of the bootstrap distribution [11].

The SAM and PAF estimates based on pooled-survey prevalence [5] were compared with those obtained using prevalence from the small-area model [7], by reference to the percentage of relative change. The relative change was calculated on the basis of the PAF, taking survey-based estimates as reference:

$$RC = \frac{PAF_M - PAF_E}{PAF_E} * 100 = \left(\frac{PAF_M}{PAF_E} - 1\right) * 100$$

where RC is the relative change, PAF_M is the PAF calculated using prevalence from the small-area model, and PAF_E is the PAF calculated using survey-based prevalence.

To compare the precision of the two estimates, we used the width of the 95% CI of the PAFs, which were broken down by sex, age group, and cause of death via quartiles.

The comparison refers to 2017 as it is the year for which the SAM in the previous study was estimated. However, no changes are expected in the conclusions if the estimate refers to another year.

All calculations were performed using the Stata v16.1 computer software program.

Results

In 2017, 53 825 (95% CI: 53 182–54 342) persons died in Spain due to tobacco use when SAM was estimated using pooled survey-based prevalences. When SAM was estimated using prevalences derived from the small-area model, 55 847 (95% CI: 55 065–57 082) died, amounting to a difference of 3.8% between the two estimates. If the data are analysed by autonomous regions, then SAM estimated on the basis of prevalence from the small-area model rises in all autonomous regions except Andalusia, where it falls by 2.4%, going from 9469 (95% CI: 9177–9743) to 9243 deaths (95% CI: 8852–9646). In the remaining autonomous regions, the relative changes in SAM ranged from 1.1% in Castile-La Mancha (from 2488 [95% CI: 2350–2601] to 2516 [95% CI: 2430–2684]) to 14.1% in Navarre (from 716 [95% CI: 671–759] to 817 [95% CI: 782–862]) (Table 1). When SAM estimates in the autonomous regions were broken down by sex, age group, and cause of death, the highest relative changes were observed in women, with a median of 26.1% versus 0.8% in men and in the ≥65-year age group versus the 35–64 age group (6.6% vs. 0). By cause of death, the median relative change was 9.0% for cardiometabolic diseases, 4.0% for respiratory diseases, and 3.1% for tumors (Supplementary Table S5). SAM estimated using the small-area model prevalence and pooled survey-based prevalence by region, cause of death, and sex can be found in Supplementary Tables S6 and S7.

In terms of the PAF, the estimate for Spain overall went from 25.9% (95% CI: 25.6–26.2) using survey-based prevalence to 26.9% (95% CI: 26.5–27.5) using prevalence from the small-area model. The difference of the widths of the CIs was also higher for estimations based on the prevalence from the small-area model (0.95 vs. 0.54). In men, the most pronounced differences between PAFs were observed in Navarre, where the PAF ranged from 36.3% (95% CI 33.6–38.5) using survey-based prevalence to 38.9% (95% CI 37.2–40.6) using prevalence derived from the small-area model. In women, these differences were observed in La Rioja, where the PAF ranged from 7.3% (95% CI: 5.9–8.8) using survey-based prevalence to 11.0% (95% CI: 10.2–12.3) using small-area model prevalence (Table 2). Figure 1 shows the PAF estimates obtained using prevalence from the small-area model and the surveys, along with their 95% CIs, by autonomous region and cause of death. Except for

Table 1. Smoking-attributable mortality (SAM) overall and in men and women, using prevalences based on a small-area model and on pooled population surveys, accompanied by their 95% confidence intervals (95% CIs)

	Overall						Men						Women					
	Model-based prevalences			Pooled survey-based prevalences			Model-based prevalences			Pooled survey-based prevalences			Model-based prevalences			Pooled survey-based prevalences		
Autonomous region	SAM	95% CI		SAM	95% CI		SAM	95% CI		SAM	95% CI		SAM	95% CI		SAM	95% CI	
Andalusia	9243	8852	9646	9469	9176	9743	7810	7508	8111	8218	7970	8443	1433	1209	1676	1251	1100	1414
Aragon	1918	1827	2023	1805	1732	1880	1588	1495	1665	1550	1490	1612	329	298	397	256	215	295
Asturias, Principality of	1702	1630	1799	1616	1530	1692	1411	1340	1483	1357	1282	1425	291	262	347	258	219	298
Balearic Isles	1211	1165	1285	1173	1114	1227	947	897	1015	942	897	983	264	248	291	231	192	266
Canary Islands	2325	2190	2506	2298	2170	2408	1792	1695	1904	1831	1748	1906	533	440	684	467	370	552
Cantabria	792	754	844	755	722	785	637	606	678	635	607	660	154	136	183	119	103	136
Castile and Leon	3435	3282	3631	3256	3127	3374	2894	2769	3018	2856	2752	2957	541	475	694	400	338	468
Castile-La Mancha	2516	2430	2684	2488	2350	2601	2244	2143	2370	2260	2157	2342	272	254	368	228	150	318
Catalonia	8880	8619	9228	8283	8048	8513	7344	7079	7616	7129	6921	7320	1535	1451	1690	1154	1044	1274
Valencian Region	6183	5904	6523	5968	5754	6152	5012	4738	5260	5031	4855	5205	1171	1068	1380	936	836	1043
Extremadura	1623	1570	1727	1562	1502	1618	1424	1362	1491	1426	1379	1472	199	184	272	136	106	169
Galicia	3720	3589	3883	3618	3462	3770	3097	2964	3234	3099	2966	3227	622	577	698	519	435	605
Madrid Region	6664	6299	7054	6310	6075	6504	4869	4643	5092	4818	4675	4951	1795	1503	2126	1492	1331	1659
Region of Murcia	1565	1504	1651	1446	1366	1515	1358	1297	1433	1289	1218	1351	207	192	239	157	126	190
Navarre	817	782	862	716	671	759	635	608	663	593	549	629	182	156	218	123	102	148
Basque Country	2902	2752	3077	2738	2625	2850	2235	2117	2344	2208	2123	2298	667	555	821	529	449	607
Rioja, La	351	334	373	325	305	345	280	263	297	277	259	294	72	67	80	48	38	57

Table 2. Population attributable fractions (PAFs) in each autonomous region and their relative change (RC), overall and in men and women, using prevalences based on a small-area model and on pooled population surveys, accompanied by 95% confidence intervals (95% CIs)

	Overall							Men							Women						
	Model-based prevalences			Pooled survey-based prevalences				Model-based prevalences			Pooled survey-based prevalences				Model-based prevalences			Pooled survey-based prevalences			
Autonomous region	PAF	95% CI		PAF	95% CI		% RC	PAF	95% CI		PAF	95% CI		% RC	SAM	95% CI		SAM	95% CI		% RC
Andalusia	25.3	24.2	26.4	25.9	25.1	26.6	−2.4	38.3	36.8	39.8	40.3	39.1	41.4	−5.0	8.9	7.5	10.4	7.7	6.8	8.7	14.6
Aragon	25.5	24.3	26.9	24.1	23.1	25.0	6.2	37.7	35.5	39.6	36.8	35.4	38.3	2.5	10.0	9.0	12.0	7.8	6.5	9.0	28.7
Asturias, Principality of	25.8	24.7	27.3	24.5	23.2	25.7	5.3	39.2	37.2	41.1	37.7	35.6	39.5	4.0	9.7	8.8	11.6	8.6	7.3	10.0	12.6
Balearic Isles	29.4	28.3	31.2	28.5	27.0	29.8	3.3	40.7	38.5	43.6	40.5	38.5	42.2	0.5	14.7	13.8	16.2	12.9	10.7	14.9	14.5
Canary Islands	30.5	28.7	32.8	30.1	28.4	31.5	1.2	40.2	38.1	42.8	41.1	39.3	42.8	−2.1	16.7	13.8	21.5	14.7	11.6	17.3	14.2
Cantabria	27.4	26.1	29.2	26.1	25.0	27.2	4.9	38.8	36.9	41.3	38.6	36.9	40.2	0.3	12.4	10.9	14.7	9.6	8.3	11.0	29.3
Castile and Leon	24.2	23.1	25.6	22.9	22.0	23.7	5.5	36.4	34.8	37.9	35.9	34.6	37.2	1.3	8.6	7.6	11.1	6.4	5.4	7.5	35.0
Castile-La Mancha	25.3	24.4	26.9	25.0	23.6	26.1	1.1	39.5	37.7	41.7	39.8	38.0	41.2	−0.7	6.3	5.9	8.6	5.3	3.5	7.4	19.3
Catalonia	28.4	27.6	29.5	26.5	25.8	27.3	7.2	41.4	39.9	42.9	40.2	39.0	41.3	3.0	11.4	10.8	12.5	8.6	7.7	9.4	33.0
Valencian Region	27.5	26.2	29.0	26.5	25.6	27.4	3.6	39.6	37.4	41.6	39.8	38.4	41.1	−0.4	11.9	10.9	14.0	9.5	8.5	10.6	25.1
Extremadura	27.0	26.1	28.8	26.0	25.0	26.9	3.9	41.6	39.8	43.5	41.7	40.3	43.0	−0.1	7.7	7.1	10.5	5.3	4.1	6.6	45.7
Galicia	23.0	22.2	24.0	22.4	21.4	23.3	2.8	35.5	34.0	37.1	35.5	34.0	37.0	−0.1	8.4	7.8	9.4	7.0	5.8	8.1	20.0
Madrid Region	30.4	28.7	32.2	28.8	27.7	29.7	5.6	41.1	39.2	43.0	40.7	39.5	41.8	1.1	17.8	14.9	21.1	14.8	13.2	16.5	20.3
Region of Murcia	27.1	26.1	28.6	25.0	23.7	26.2	8.2	41.2	39.4	43.5	39.1	37.0	41.0	5.4	8.4	7.7	9.6	6.3	5.1	7.7	32.0
Navarre	29.0	27.8	30.6	25.4	23.8	26.9	14.1	38.9	37.2	40.6	36.3	33.6	38.5	7.2	15.4	13.2	18.4	10.4	8.6	12.5	47.4
Basque Country	28.5	27.0	30.2	26.9	25.8	28.0	6.0	38.7	36.7	40.6	38.2	36.8	39.8	1.2	15.1	12.6	18.6	12.0	10.2	13.8	26.1
Rioja, La	23.8	22.6	25.2	22.0	20.6	23.3	8.2	33.8	31.8	35.9	33.5	31.4	35.6	0.8	11.0	10.2	12.3	7.3	5.9	8.8	50.8

The RC was calculated, taking survey-based prevalences as reference.

Andalusia, the PAFs were always higher with the small-area model prevalence, as was also the case for SAM rates. The negative differences between PAFs never exceeded one percentage point, and the positive differences were generally below two percentage points. The greatest difference was observed for the PAF for respiratory diseases in Navarre, which went from 48.9% (95% CI: 46.0–51.1) to 53.8% (95% CI: 52.0–55.6), when prevalence from the small-area model was applied.

In general, the width of the CI of the PAF was greater for estimates based on the prevalence from the small-area model, but the differences were small. Hence, the median width of the CI in the case of cancer was 2.3 for survey-based prevalence and 2.9 for model-based prevalence, with an interquartile range (IQR) of 0.3 and 0.7, respectively. For cardiometabolic diseases, the median went from 1.9 (IQR: 0.5) to 2.3 (IQR: 0.7) when the model-based prevalence was used. For respiratory diseases, it remained practically the same with values of 3.8 (IQR: 1.9) and 3.7 (IQR: 0.5), respectively, but with considerably lower variability. Broken down by sex, the median width of CI in men was 3.0 (IQR: 1.0) for survey-based prevalence and 4.0 (IQR: 0.8) for model-based prevalence, with the equivalent values for women being 2.6 (IQR: 1.3) and 3.0 (IQR: 1.4), respectively. Among subjects under the age of 65 years, the median went from 2.9 (IQR: 1.0) with survey-based prevalence to 4.4 (IQR: 1.1) with model-based prevalence. In the population aged 65 years and over, it went from 2.4 (IQR: 0.8) with survey-based prevalence

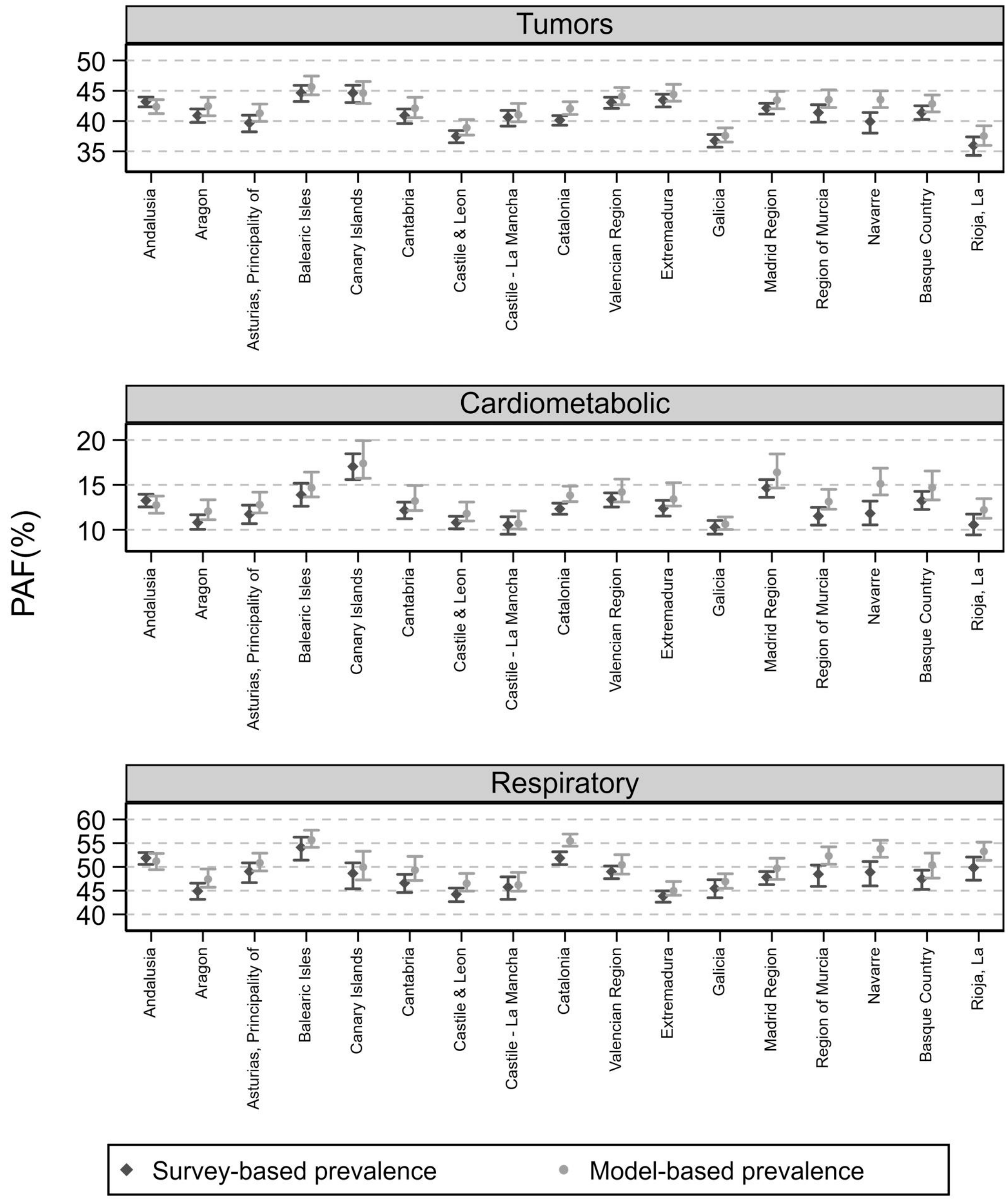

Figure 1. Population attributable fractions (PAFs) in each autonomous region by cause of death estimated using prevalences based on pooled population surveys and on a small-area model, accompanied by their 95% confidence intervals.

to 2.9 (IQR: 0.5) with model-based prevalence (Fig. 2 and Supplementary Table S8).

Discussion

When prevalences derived from a small-area model were applied, tobacco use was associated with 55 847 deaths in Spain in 2017. This estimate is 3.8% higher than that observed when prevalences from the three pooled health surveys are applied. The changes in SAM vary among the autonomous regions, and in some regions, these changes can be as much as 14%. The median of the relative change between regions was higher for women (26.1%), persons aged ≥65 years (6.6%), and cardiometabolic diseases (9.0%). With the sole exception of Andalusia, the estimated SAM and PAFs are higher when prevalences from the small-area model are used, and this same effect is seen in the width of the 95% CI of the PAF.

In other countries like Canada, United States, Poland, China, Portugal, Germany, and Belgium, SAM has been estimated, as in Spain, at a regional level. Most of the studies [12–16] applied a method that uses lung cancer mortality rates to obtain SAM, an approach that is used as an alternative when prevalence figures are unavailable [17, 18]. Some studies have used a method based on smoking prevalences, such as some conducted in the United States [19], whose data source was a single survey which does not guarantee regional representativeness. Other studies have, however, used different alternatives. A study undertaken in British Columbia (Canada) [20] reported that the limited sample size of the survey in some age groups had to be resolved by calculating a weighted mean smoking prevalence from three surveys conducted in consecutive years. Furthermore, in another study conducted in 589 towns and cities in Belgium [21], the limited sample size of the health surveys was, as in our study, overcome by applying a small-area model to obtain prevalences at a municipal level. However, the precision of the estimates obtained with respect to other data sources, such as population surveys, was not analysed.

Knowledge of smoking prevalences and SAM at a regional level is fundamental for some reasons. First, because this provides a more detailed view of the evolution of the smoking epidemic in the Spanish population. This can help to better understand why the prevalence of consumption and the figures of SAM is higher in

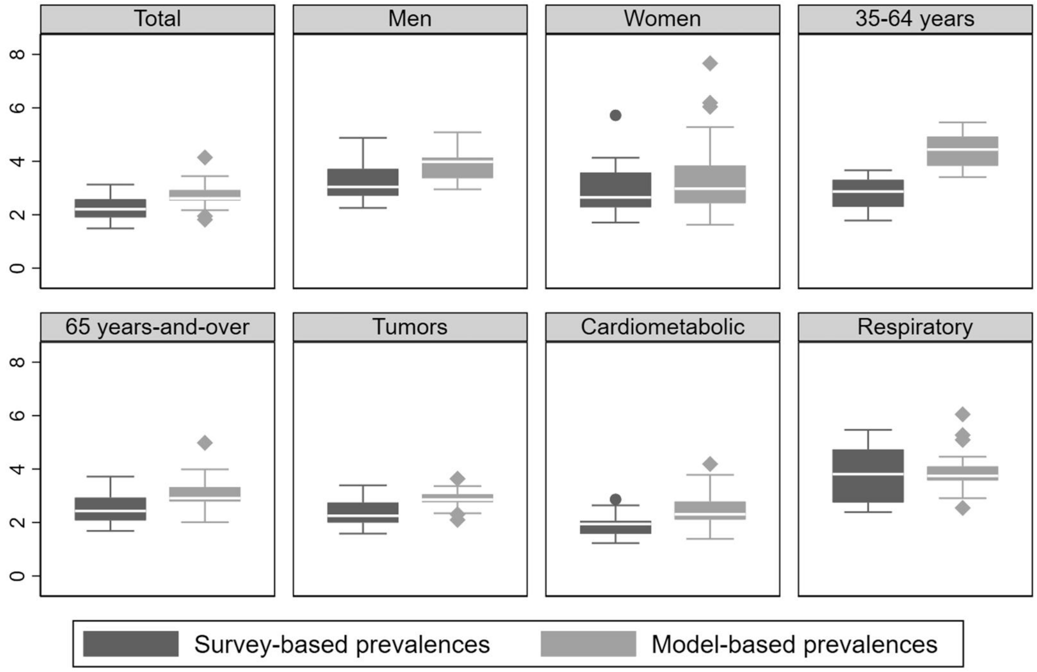

Figure 2. Width of confidence intervals of population attributable fractions calculated using prevalences based on pooled population surveys and a small-area model, overall and by sex, age group, and cause of death.

some autonomous regions than in others. On the other hand, the availability of this information permits the design of prevention measures appropriate to the needs of each region. In Spain, autonomous regions have the authority to develop public health policies at the local level. Therefore, having this data available facilitates autonomous region decision-makers to implement new tobacco control measures by basing their decisions on more accurate data.

While regional data are essential, they tend not to be available as the conduct of these surveys is expensive, largely driven by a larger sample size. The use of a small-area model is therefore postulated as an alternative, making it possible to estimate tobacco use prevalences for smokers, ex-smokers, and never smokers. A small-area model was applied in previous studies to obtain smoking prevalences at a local level [22–24]; however, several methodological aspects vary among them. For example, Srebotnjak *et al.* [25] used two independent models for each sex, which may be appropriate for estimating smoking prevalence in Spain because the evolution of smoking prevalence differs between men and women. However, the Santiago-Pérez *et al.* [7] model used for this study offered greater precision when both sexes were included. On the other hand, there are other studies in which prevalences were calculated for the categories of smoker and non-smoker [24], whereas in the model used by Santiago-Pérez *et al.* [7], prevalences were obtained for smokers, ex-smokers, and never smokers, which is essential for the calculation of SAM. Moreover, these models can be used to estimate the prevalences of other risk factors such as alcohol [24, 26], obesity [27], or overweight [28]. However, it should be noted that in a study that compared the use of these models to estimate prevalences of tobacco and alcohol consumption, it was indicated that this model works better in estimating prevalences of tobacco use [24]. These advantages must be weighed against the disadvantage; it should be noted that the complexity associated with developing the model and, in some cases, the difficulty of access to auxiliary variables can be obstacles.

The pooling of several databases conducted within a short time of one another is another alternative for obtaining smoking prevalences at a regional level [7], though this alternative is subject to limitations. It should be borne in mind that the impact of pooling several surveys on the prevalence that is finally estimated may be variable. If prevalences remain stable, this impact will be minimal. However, if prevalences vary between surveys, prevalence data that have undergone changes could be stabilized. For instance, one of the prevalences used in this study came from pooling three surveys separated by six-year gap between the first (2011) and the last (2017) survey. During the period 2011-2017, smoking prevalence among men decreased in all autonomous regions, while among women, smoking prevalence also decreased in some regions and remained stable in others [29]. The fact that prevalences in 2011 were slightly higher than those in 2017 might give rise to a situation where prevalences derived from the pooling of surveys may be slightly higher than those based on the small-area model, which uses prevalence data pertaining to a single year, in this case, 2017. This may be one of the reasons for the differences in the SAM estimates found in this study. In addition, the pooling of surveys can give rise to other doubts, such as how many surveys should be pooled, what sample size should be achieved, or how great a gap there can be between the pooled surveys. These doubts highlight the fact that pooling a number of surveys does not follow an established methodology of calculation and could thus be researcher-dependent.

Analysis of the results showed that, at a national level, the differences in the SAM estimates obtained on using model-based or pooled-survey-based prevalences are not important. Yet, when SAM is estimated at a regional level, the estimates display variation in specific groups, such as women, the population aged 65 years and over, and cardiometabolic diseases. In women, variability in the SAM figures may be related with the fact that smoking prevalences are low [1, 2], which renders prevalences more unstable and prone to greater variability. In the population aged 65 years and over, the explanation may also be related with the fact that prevalences of smokers above this age tend to decline [1, 2]. Indeed, the group of women aged 75 years and over registers very low smoking prevalences in all autonomous regions.

This study has limitations related to the estimation method, the data sources, and the use of a small-area model to obtain prevalences of tobacco use. The prevalence-dependent method does not respect the causality criterion of temporality proposed by Bradford Hill because the prevalences of tobacco use (cause) and observed mortality (effect) are from the same year. Therefore, the time required for tobacco use to cause a death associated with its use is not respected. This limitation associated with the use of the prevalence-dependent method may result, in the case of Spain, in an underestimation of the real impact of tobacco use on mortality. This is due to the fact that the prevalence of tobacco use in Spain has been decreasing since 1987 among men and since 2001 among all age groups [29] so that when using prevalences contemporaneous with mortality, these would be lower than those of two or three decades ago. A previous study that compared the use of this method with another that approximates the latency period of tobacco use with respect to lung cancer by using mortality rates found that the differences were not important at the national level. However, larger differences were found in subgroups where smoking prevalence was lower, such as women or people aged over 65 years of age [30]. Regarding data sources, this study used RRs from five US cohorts given that Spanish cohort studies assessing the risk of tobacco-related diseases in smokers and ex-smokers versus never smokers are not available. However, by using these RRs, we are assuming that the evolution of the tobacco epidemic is the same in Spain as in the United States. Smoking prevalences derived from surveys are used, which may result in concealment of consumption and underestimation of prevalences. However, previous studies observed that in adults the degree of agreement between self-reported consumption and biological measurements is high [31]. Among the limitations associated with the use of a small-area model are those related with the obtaining of auxiliary variables. Access to such data is often very complicated, as is finding data for the years for which the estimate of prevalences is sought. In addition, there is the difficulty of these data being available at the level of geographic breakdown for which prevalences are to be estimated. Furthermore, it must be borne in mind that the quantity of auxiliary variables and their relationship with the outcome variable (smoking prevalence) by reference to the geographic setting may affect the model's goodness-of-fit and its predictive capacity.

Overall, the differences between the estimates of SAM, PAF, and the width of the CI are small, when estimates based on prevalences from a small-area model are compared with those based on prevalences from pooled population surveys. Small-area models could be a valid alternative to the pooling of surveys in cases where prevalence-dependent methods are applied to estimate SAM at a regional level, especially in countries in which prevalence may be changing. In addition, the small-area models only require data from a single health survey, and the precision in the estimates of attributed mortality is good.

Acknowledgements

This paper forms part of the work leading to the doctoral thesis of Julia Rey Brandariz, who has received a University Teacher-Training (*Formación de Profesorado Universitario/FPU*) grant (reference number FPU20/00926) from the Spanish Ministry of Universities.

Author contributions

J.R.B. has participated in the conceptualization, planification, methodology, formal analysis, writing of the individual draft, and visualization. M.I.S.P. has participated in the conceptualization, methodology, formal analysis, and writing—review and editing. C. C.P. has participated in the conceptualization, planification, methodology, and writing—review and editing. L.V.L. has participated in the conceptualization, methodology, and writing—review and editing. A.R.R. has participated in the conceptualization, methodology, and writing—review and editing. E.L.V. has participated in the formal analysis and writing—review and editing. C.G.T. has participated in the formal analysis, methodology, and writing—review and editing. J.S.A. has participated in the methodology and writing—review and editing. A.M. has participated in the planification and writing—review and editing. M.P.R. has participated in the conceptualization, formal analysis, methodology, supervision, project administration, and writing—review and editing.

Supplementary data

Supplementary data are available at *EURPUB* online.

Conflict of interest

J.S.A. was funded in part by P20GM130414, an NIH-funded Center of Biomedical Research Excellence (COBRE). J.S.A. received sponsored funds for travel expenses as a speaker for the 2021 and 2022 annual GTNF conference. J.S.A. serves as a consultant, receives compensation, and has equity in Qnovia, a start-up company developing a prescription nicotine replacement product for FDA approval. The rest of the authors declare that they have no conflict of interest.

Funding

This study has been funded by Instituto de Salud Carlos III (ISCIII) through the project 'PI19/00288' and co-funded by European Union.

Data availability

The derived data generated in this research will be shared on reasonable request to the corresponding author.

Key points

- The results suggest that using prevalences from a small-area model to estimate smoking-attributed mortality is a good alternative to have data at the regional level according to sex, age, and cause of death.
- The differences found were small, although the largest differences were observed in groups in which the prevalence of tobacco use is smaller (women and population aged ≥65 years).
- The availability of these estimates allows improved surveillance and monitoring of the tobacco epidemic in a population in order to design and implement the most appropriate preventive measures at a more local level.

References

1 Ministerio de Sanidad. *Encuesta Nacional de Salud de España*. https://www.sanidad.gob.es/estadEstudios/estadisticas/encuestaNacional/home.htm (16 March 2023, date last accessed).

2 Ministerio de Sanidad. *Encuesta Europea de Salud en España*. https://www.sanidad.gob.es/estadEstudios/estadisticas/EncuestaEuropea/home.htm (16 March 2023, date last accessed).

3 Rey-Brandariz J, Pérez-Ríos M, Santiago-Pérez MI *et al.* Trends in smoking-attributable mortality in Spain: 1990-2018. *Eur J Public Health* 2022;**32**:919–25.

4 Rey-Brandariz J, Pérez-Ríos M, Santiago-Pérez MI *et al.* Smoking-attributable mortality in Spain: a systematic review. *Adicciones* 2023;**35**:185–96.

5 Rey J, Pérez-Ríos M, Santiago-Pérez MI *et al.* Smoking-attributable mortality in the autonomous communities of Spain, 2017. *Rev Esp Cardiol (Engl Ed)* 2022;**75**:150–8.

6 Ministerio de Sanidad. *Encuesta Nacional de Salud de España 2017*. 2018. https://www.mscbs.gob.es/estadEstudios/estadisticas/encuestaNacional/encuesta2017.htm (21 March 2023, date last accessed).

7 Santiago-Pérez MI, López-Vizcaíno E, Pérez-Ríos M *et al.* Small-area models to assess the geographical distribution of tobacco consumption by sex and age in Spain. *Tob Induc Dis* 2023;**21**:63. https://doi.org/10.18332/TID/162379.

8 Pérez-Ríos M, Rey-Brandariz J, Galán I *et al.* Methodological guidelines for the estimation of attributable mortality using a prevalence-based method: the STREAMS-P tool. *J Clin Epidemiol* 2022;**147**:101–10.

9 Levin M. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum* 1959;**9**:531–41.

10 U.S. Department of Health and Human Services. *The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, 2014.

11 Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press, 1997. https://doi.org/10.1017/CBO9780511802843.

12 Fihel A, Muszyńska M. The regional variation in tobacco smoking—attributable mortality in Poland, 2006-2010. *Przegl Epidemiol* 2015;**69**:87–92.

13 Muszyńska MM, Fihel A, Janssen F. Role of smoking in regional variation in mortality in Poland. *Addiction* 2014;**109**:1931–41.

14 Forjaz G, Bastos J, Castro C *et al.* Regional differences in tobacco smoking and lung cancer in Portugal in 2018: a population-based analysis using nationwide incidence and mortality data. *BMJ Open* 2020;**10**:e038937. https://doi.org/10.1136/BMJOPEN-2020-038937.

15 Grigoriev P, Klüsener S, van Raalte A. Quantifying the contribution of smoking to regional mortality disparities in Germany: a cross-sectional study. *BMJ Open* 2022; **12**:e064249.

16 Xu Z, Qi F, Wang Y *et al.* Cancer mortality attributable to cigarette smoking in 2005, 2010 and 2015 in Qingdao, China. *PLoS One* 2018;**13**:e0204221. https://doi.org/10.1371/JOURNAL.PONE.0204221.

17 Ezzati M, Lopez AD, Rodgers A, Murray CJL. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva, Switzerland: World Health Organization 2004.

18 Peto R, Boreham J, Lopez AD *et al.* Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *Lancet* 1992;**339**:1268–78.

19 Ma J, Siegel RL, Jacobs EJ *et al.* Smoking-attributable mortality by state in 2014. U.S. *Am J Prev Med* 2018;**54**:661–70.

20 Tu AW, Buxton JA, Stockwell T. Estimates of smoking-attributable mortality and hospitalization in BC, 2002-2007. *Can J Public Health* 2012;**103**:137–41.

21 Putrik P, Otavova M, Faes C *et al.* Variation in smoking attributable all-cause mortality across municipalities in Belgium, 2018: application of a Bayesian approach for small area estimations. *BMC Public Health* 2022;**22**:1699.

22 Song L, Mercer L, Wakefield J *et al.* Using small-area estimation to calculate the prevalence of smoking by subcounty geographic areas in King County, Washington, Behavioral Risk Factor Surveillance System, 2009–2013. *Prev Chronic Dis* 2019;**13**: E59. https://doi.org/10.5888/PCD13.150536.

23 Szatkowski L, Fahy SJ, Coleman T *et al.* Small area synthetic estimates of smoking prevalence during pregnancy in England. *Popul Health Metr* 2015;**13**:34. https://doi.org/10.1186/S12963-015-0067-8.

24 Twigg L, Moon G, Jones K. Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators. *Soc Sci Med* 2000;**50**:1109–20.

25 Srebotnjak T, Mokdad AH, Murray CJL. A novel framework for validating and applying standardized small area measurement strategies. *Popul Health Metr* 2010; **8**:26–13.

26 Pryce R, Angus C, Holmes J *et al.* Reweighting national survey data for small area behaviour estimates: modelling alcohol consumption in local authorities in England. *Popul Health Metr* 2020;**18**:1–12.

27 Zhang Z, Zhang L, Penman A *et al.* Using small-area estimation method to calculate county-level prevalence of obesity in Mississippi, 2007-2009. *Prev Chronic Dis* 2011; **8**:A85.

28 Xie D, Raghunathan TE, Lepkowski JM. Estimation of the proportion of overweight individuals in small areas—a robust extension of the Fay–Herriot model. *Stat Med* 2007;**26**:2699–715.

29 Rey-Brandariz J, Ruano-Ravina A, Santiago-Pérez MI *et al.* Evolution of smoking prevalence in Spain and its 17 autonomous regions (1987-2020). *Med Clin (Barc)* 2024;**162**:417–24.

30 Rey-Brandariz J, Blanco-Ferreiro A, Varela-Lema L *et al.* Estimations of smoking-attributable mortality in Spain at a regional level: comparison of two methods. *Ann Epidemiol* 2023;**82**:77–83.e3.

31 Agaku IT, King BA. Validation of self-reported smokeless tobacco use by measurement of serum cotinine concentration among US adults. *Am J Epidemiol* 2014; **180**:749–54.

European Journal of Public Health, 2024, 00, 1–7
https://doi.org/10.1093/eurpub/ckae104
Original Manuscript