European Journal of Public Health, Vol. 35, No. 2, 263–269

https://doi.org/10.1093/eurpub/ckaf015 Advance Access published on 10 March 2025

Modelling annual prevalence of tobacco consumption in Spain, 1991–2020

Carla Guerra-Tort [1], Esther López-Vizcaíno [2], María Isolina Santiago-Pérez [3], Julia Rey-Brandariz [1,4,5,*], Cristina Candal-Pedreira [1,4,5], Leonor Varela-Lema [1,4,5], Iñaki Galán [6,7], Alberto Ruano-Ravina [1,4,5], Mónica Pérez-Ríos [1,4,5]

[1]Department of Preventive Medicine and Public Health, Universidade de Santiago de Compostela, Santiago de Compostela, Spain
[2]Diffusion and Information Service, Galician Institute of Statistics, Santiago de Compostela, Spain
[3]Health Information Service, Directorate-General of Public Health, Galician Regional Health Authority, Santiago de Compostela, Spain
[4]Health Research Institute of Santiago de Compostela (IDIS), Santiago de Compostela, Spain
[5]Consortium for Biomedical Research in Epidemiology & Public Health (CIBER Epidemiología y Salud Pública—CIBERESP), Madrid, Spain
[6]National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain
[7]Department of Preventive Medicine and Public Health, Autonomous University of Madrid/IdiPAZ, Madrid, Spain

*Corresponding author. Department of Preventive Medicine and Public Health, Universidade de Santiago de Compostela. Rúa San Francisco s/n, 15782 Santiago de Compostela, Spain. E-mail: juliarey.brandariz@usc.es.

Abstract

The aim of this study was to estimate the series of tobacco smoking prevalence year-by-year in Spain, by sex and age group, for the period 1991–2020. Based on smoking prevalence obtained from national surveys and smoking-related auxiliary information from public statistics, we fitted a multinomial logistic mixed model with random area and time effects. Joinpoint regression was used to identify changes in the prevalence series across the period. To analyse the precision of the model-based estimates, we calculated the coefficients of variation. Between 1991 and 2020, the prevalence of smoking in Spain decreased in both sexes. In the 15–24 age group, the prevalence of smokers showed no differences by sex until 2007, after which prevalence in men exceeded that of women. However, in women aged 55 and over prevalence of smoking has been rising since 1991. After applying the model, the precision of smoking prevalence estimates improved. The reconstruction of a detailed series of tobacco smoking prevalence provides insight into the evolution of the tobacco epidemic in Spain. A detailed analysis by sex and age shows different trends in the prevalence of smoking among women that should be considered when control measures are formulated.

Introduction

Currently, over a billion people smoke worldwide, which means ~15% of the world population are smokers [1]. According to Eurobarometer 2021, 24 of every 100 Spanish over the age of 14 smoke, placing Spain above the European prevalence set at 23% [2].

Since 1987, the Spanish Health Survey (NHS) (*Encuesta Nacional de Salud de España/ENSE*) and the European Health Survey in Spain (EHS) (*Encuesta Europea de Salud en España/EESE*) have monitored the prevalence of several health determinants, among them tobacco smoking. Surveys are available for only 12 of the 34 years that comprise the period 1987–2020. Despite the fact that every edition provides information on tobacco smoking, in Spain, we have an incomplete picture of the evolution of the smoking epidemic. According to data furnished by these surveys, the prevalence of tobacco use has been falling in Spain since the 1990s, due essentially to the 30 percentage-point drop in the prevalence of smoking among men. Smoking in any of the years for which data are available is more prevalent among men than women across all age groups. Also the male/female prevalence ratio has fallen appreciably, from 2.4 in 1987 to 1.4 in 2020. However, a detailed analysis year-by-year of prevalence is crucial for understanding the history of the smoking epidemic, and for predicting its future evolution [3]. Nevertheless, the lack of regular surveys makes it necessary to fill the gaps in knowledge.

Previously, this was achieved by estimating the prevalence retrospectively by aggregating the data of several editions of the NHS [4–6]. This approach fails to take advantage of all the available NHS and EHS data. Collaterally, the further one moves away from the first year of the aggregated surveys, the less the sample represents the population of earliest years, leading to biased estimates. Until now, there has been no historical year-by-year picture of smoking prevalence in Spain broken down by sex and age.

The aim of this study was to estimate the annual series of prevalence of smokers, ex-smokers, and never smokers from 1991 to 2020, by sex and 5-year age group (from 15–19 through 80–84 years).

Methods

A multinomial logistic mixed model with random area and time effects that uses aggregate smoking-related data and auxiliary information [7, 8] sourced from available records was applied.

Data sources

Smoking-related data: Anonymized data from the NHS and EHS, available on the National Statistics Institute (NSI) (*Instituto Nacional de Estadística/INE*) and Ministry of Health websites [3]: 1987, 1993, 1995, 1997, 2001, 2003, 2006, 2009, 2011, 2014, 2017,

and 2020. These surveys are targeted at the population aged 15 years and older residing in family dwellings nationwide. In the NHS and EHS, the questionnaire includes a set of questions related to tobacco use, which served to create the categorical variable "smoker status." We defined a "smoker" (S) as anyone who smokes at the date of the survey, an "ex-smoker" (ExS) as anyone who had smoked but no longer did so, and a "never smoker" (NS) as anyone who had never smoked.

Smoking-related auxiliary information: By way of auxiliary information, we included variables associated with tobacco, such as: (i) nationality: percentage of foreign population [9, 10]; (ii) educational level: percentage of population with basic, secondary, or higher education [11]; (iii) relationship with work activity: percentage of population gainfully employed, economically inactive, and the unemployment rate [11]; (iv) main occupation: percentage of directors, managers, technicians, and professionals, percentage of skilled workers, and percentage of unskilled workers [11]; (v) occupational sector: percentage of population gainfully employed in agriculture, industry, business, transport and hospitality, construction, health and education, information, communications, finance, insurance and real estate activities, and other services [11]; and (vi) morbidity rate: hospitalized patients due to lung cancer per 1000 persons [12]. The rate denominator refers to the reference populations at the 1 of July, available from the NSI, for each year of the period 1991–2020 [13].

The data sources of the variables are shown in Supplementary Table S1. The summary measures of the variables are shown in Supplementary Table S2.

Statistical analysis

Imputation of missing data

To reconstruct the prevalence series of tobacco use, it is necessary to ascertain interviewees' smoking history, along with the age of initiation and cessation, where applicable. In 1987 NHS does not record the age at initiation and cessation, so this survey could not be considered for analysis purposes. The 2017 NHS and the 2009, 2014, and 2020 EHS partially record this information. For these years, the age of smoking initiation and cessation were imputed in S and ExS by means of Multivariate Imputations by Chained Equations (MICE) [14], using data from the remaining surveys with available information.

Estimation of smoking prevalence

Based on eight NHS and three EHS, we reconstructed individual smoking history for each year from 1991 to 2020. For this purpose, we considered an individuals' relationship with tobacco at the year of the survey and, where appropriate, age at initiation and age at cessation. Current S were considered NS prior to the year of initiation and S from the year of initiation until the year of the survey. ExS were defined as NS prior to the year of initiation, S from the year of initiation until the year of cessation, and ExS from the year of cessation until the year of the survey. NS were considered NS throughout the period.

The prevalence of S, ExS, and NS were then estimated by sex and 5-year age group for each year of the period 1991–2020, applying a weighted ratio estimator (direct estimator).

$$\widehat{p} = \frac{\sum_i W_i X_i}{\sum_i W_i},$$

where i is the individual; X_i represents the characteristic estimated (S, ExS, or NS), which takes values 0–1; and W_i is the sampling weight. We calculated the variance of this estimator using a linear approach in Taylor's series, and based on this, we obtained the coefficients of variation (CVs).

Modelling smoking prevalence

Based on the smoking prevalence obtained by applying the direct estimator and the smoking-related auxiliary information, we fitted a multinomial logistic mixed model with random area and time effects, with the aim of improving the precision of the estimates [7]. The areas were the $D = 28$ groups defined on the basis of sex and 14 age groups. The time periods were the $T = 30$ years of the period 1991–2020. In the model, the response variable is a vector with the number of S and ExS in each area and time. The explanatory variables are the aggregate smoking-related auxiliary information. The model is:

$$p_{dkt} = \frac{\exp(\eta_{dkt})}{1 + \exp(\eta_{d1t}) + \exp(\eta_{d2t})},$$

$$\eta_{dkt} = \log\left(\frac{p_{dkt}}{p_{d3t}}\right) = \boldsymbol{x}_{dkt}\boldsymbol{\beta}_k + \boldsymbol{u}_{1,dk} + \boldsymbol{u}_{2,dkt}, \quad d = 1, \ \ldots, \ D, \quad k = 1, 2, \quad t = 1, \ \ldots, \ T,$$

where p_{dkt} is the prevalence of each category k corresponding to area d and time t; $\boldsymbol{x}_{dkt} = (x_{dkt1}, \ldots, x_{dktr_k})'$ is the set of explanatory variables corresponding to category k, area d and time t; and $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kr_k})$ is the vector of regression parameters. The subindex k refers to the category of S ($k = 1$) or ExS ($k = 2$). The third category, NS ($k = 3$), is taken as the reference category, so that $p_{d3t} = 1 - (p_{d1t} + p_{d2t})$. The model considers the random effects $\boldsymbol{u}_{1,dk}$ and $\boldsymbol{u}_{2,dkt}$ associated with area d and category k, and with area d, category k and time t, respectively. To fit the model, we combined the penalized quasi-likelihood method for the estimation and prediction of $\boldsymbol{\beta}_{kr_k}$, $\boldsymbol{u}_{1,dk}$, and $\boldsymbol{u}_{2,dkt}$, and the restricted maximum likelihood method to estimate the components of the variance.

To assess the precision of the model-based estimates we calculated the mean squared error (MSE) using a parametric bootstrap procedure [15]. We then obtained the 95% confidence intervals (95% CIs) for the prevalence (95% CI $= \widehat{p} \pm 1.96\widehat{p}\sqrt{\text{MSE}}$) and the CVs (CV $= \sqrt{\text{MSE}}$). CVs below 30% were deemed acceptable, considering the criteria applied by the National Center for Health Statistics [16].

Using the model-based estimates, we calculated the prevalence in 10-year age groups for men and women alike and analysed their trend by fitting joinpoint regression models. We considered a maximum of five change points. In each segment identified, we estimated the annual percentage change (APC), its 95% CI and significance level (P values). APCs with a P values <0.05 were deemed to be significant.

Data were processed using the Stata v17 computer software program [17], the estimation of the model was performed with the mme package for R [18], and trends were analysed with the Joinpoint Regression Program v4.9.1.0 [19].

Results

After pooling the eight NHS and three EHS, debugging and processing the resulting database, a final sample of 271 181 records was achieved.

The estimated coefficients of the models for S and ExS are shown in Table 1. The sets of auxiliary information yielding significant variables (P values < 0.05) to the models, of both S and ExS, were educational level, relationship with activity, occupational sector, and morbidity. The variables that most contributed to predicting the prevalence of S were the percentage of employees engaged in information, communications, finance, insurance, and real estate activities, and the percentage of population economically inactive. The variables that most contributed in the case of ExS were the percentage of employees in construction, and business. The value of the Akaike Information Criterion of the model was −9437.18.

Table 1. Estimated coefficients of the model (Beta), standard error (SE), and significance (*P* values)

	Beta	SE	*P* values
Smokers			
Intercept	0.526	0.395	0.182
% with secondary education	2.527	0.104	<0.001
% with higher education	−2.134	0.247	<0.001
% gainfully employed	−1.047	0.383	0.006
% economically inactive	−4.755	0.374	<0.001
% in construction	−1.793	0.311	<0.001
% in information and communications, finance, insurance, and real estate activities	5.594	0.738	<0.001
Hospitalizations due to lung cancer/1000 population	0.350	0.008	<0.001
Ex-smokers			
Intercept	−2.234	0.115	<0.001
% with secondary education	0.936	0.105	<0.001
% gainfully employed	3.365	0.076	<0.001
% construction employees	−4.765	0.299	<0.001
% business employees	−4.410	0.339	<0.001
Hospitalizations due to lung cancer/1000 population	0.466	0.007	<0.001

Regarding the precision of the estimates, Supplementary Fig. S1 shows the CVs by smoker status associated with the direct estimator and the model. The NS prevalence obtained with the model yielded a CV below 30%. In the case of S and ExS, 15 and 8 estimates, respectively, yielded a CV above the value taken as reference. Most of these CVs pertained to women S and ExS in the older age groups. The prevalence of smoking in these women, especially until 2009, was very low, which accounts for the lower degree of precision in these estimates and the higher CV. Despite this, the results indicate a general improvement in the precision of the prevalence estimated with the model. Thus, the median of the CV decreased by 2.8% for S (from 6.0% to 5.8%) and by 8% for ExS (from 8.9% to 8.2%). Furthermore, no CV associated with model-based estimates exceeded 40%.

For the years for which survey information is available, the median of the absolute value differences between the direct estimates and the model-based estimates were: 0.23 for S; 0.19 for ExS; and 0.31 for NS. Ninety-four percent of the differences between estimates from the two methods were <1.5 percentage points.

Figures 1–3 show the model-based estimated series of annual prevalence of S, ExS, and NS, respectively, along with their 95% CI. The series is shown in men and women. Supplementary Tables S3–S5 show the results of the joinpoint regression.

Since 1991, there has been a decrease in the overall prevalence of S in Spain, with some sex-related differences. In the case of women S, the APC for the period 1991–2020 was estimated at −1.3 (95% CI: −1.6, −1.0). That said, the prevalence of women S increased 3 percentage points between 1991 and 1998 (from 25.8% to 28.8%), after which it began to fall until reaching 18.2% in 2020. This change in trend was detected by joinpoint regression, but it was not significant (*P* values = 0.069). In terms of age, a general decrease in the prevalence of women S was seen among the youngest women. Between the ages of 35–54, there was a rise in the initial years of the period followed by a drop; whereas between the ages of 55–64 and 65–74 prevalence rose across the entire period. This increase was most pronounced from 1999 onwards for women aged 55–64, with an estimated APC of 7.4 (95% CI: 6.4, 8.3) from 1999 to 2014. For women aged 65–74, the main rise took place between 2001 and 2020, with an APC of 7.1 (95% CI: 6.4, 7.8). Among older women, the prevalence of S ranged from 1.4% to 3.1%. In the case of men, the prevalence of S fell in all age groups across the entire period, with an APC of −2.4 (95% CI: −2.5, −2.2). In 2020, the prevalence of men S was estimated at 25.7%. In all age groups, the prevalence of S was higher in men, except in the 15–24 age group, in which the prevalence series for men and women resembled each other until 2007 (Fig. 1).

The prevalence of ExS rose in both sexes across the entire period. From 1991 through 2020, the APC was estimated at 1.8 (95% CI: 1.3, 2.3) for men, and 4.3 (95% CI: 3.5, 5.0) for women. Between the ages of 15–44, the prevalence series for men and women ExS were very similar. In men, from the age of 45 years onwards, the prevalence of ExS increased steadily across the entire period. In women, we observed a decrease in the prevalence of ExS between the ages of 45–54 from 2012, and an increase between the ages of 55–74, which became more pronounced from 2011 onwards (Fig. 2).

Regarding NS, the overall prevalence increased in men and decreased in women from 1991 through 2020. Between the ages of 15–34, the prevalence of men and women NS were similar, with a rising trend. The main differences between sexes were seen from the age of 35 years onwards (Fig. 3).

Discussion

The estimates of annual prevalence of smokers, ex-smokers, and never smokers in Spain from 1991 to 2020 show that tobacco use has steadily fallen in men since 1991. In the 15–24 age group, the prevalence of S shows no differences by sex until 2007, after which prevalence in men exceeds that of women. In the 55–64 and 65–74 age groups, the prevalence of women S has risen since 1991. This rise is an indicator of a possible future morbidity and mortality epidemic among women in Spain.

In this study, a multinomial logistic mixed model with random area and time effects was applied to estimate the annual prevalence series of S, ExS, and NS in Spain with a breakdown by sex and 5-year age groups, improving the precision of the estimates obtained directly from surveys. In the absence of stratification, direct estimates can provide a good result. However, when stratifying by sex and age, as the number of individuals in the sample decreases the variability in the estimates increases. Our target variable, the annual prevalence of S and ExS by sex and age group, is not a population value, but a sample-based estimate with an associated error. In addition to improving precision, the proposed model gives access to the standard error of the estimator.

When comparing the model-based estimates with the direct estimates provided by the NHS and EHS, no relevant changes are observed over the period. Ninety-four percent of the differences between the model-based and direct estimates were <1.5 percentage points. However, it should be noted that the model-based estimates place the highest peak of S among women in 1998, while the direct estimates place the highest peak in 2001. This fact highlights the importance of having an annual prevalence series of health determinants in order to accurately assess evolution over time. Moreover, differentiation by sex and age is crucial.

Many studies which reconstruct the prevalence of tobacco use from surveys over an extended period have used smoking history and age at initiation and cessation to fill in the gaps between surveys [4–6, 20–23]. With the reconstructed data, the prevalence of tobacco use is estimated for each year of the period by sex, age, or educational level. The prevalence is estimated as the proportion of S among the total population. In these papers, the reconstructed period varies from 14 to 83 years. An 83-year period may be too long, compromising both the precision of the estimates and the representativeness of age groups for the initial years of the reconstructed series. This limitation may be especially relevant when the sample size is too small or stratified analyses are performed. Applying the multinomial logistic mixed model mitigates this limitation.

Holford *et al.* [24] proposed an age–period–cohort model with constrained natural splines from individual survey data to estimate

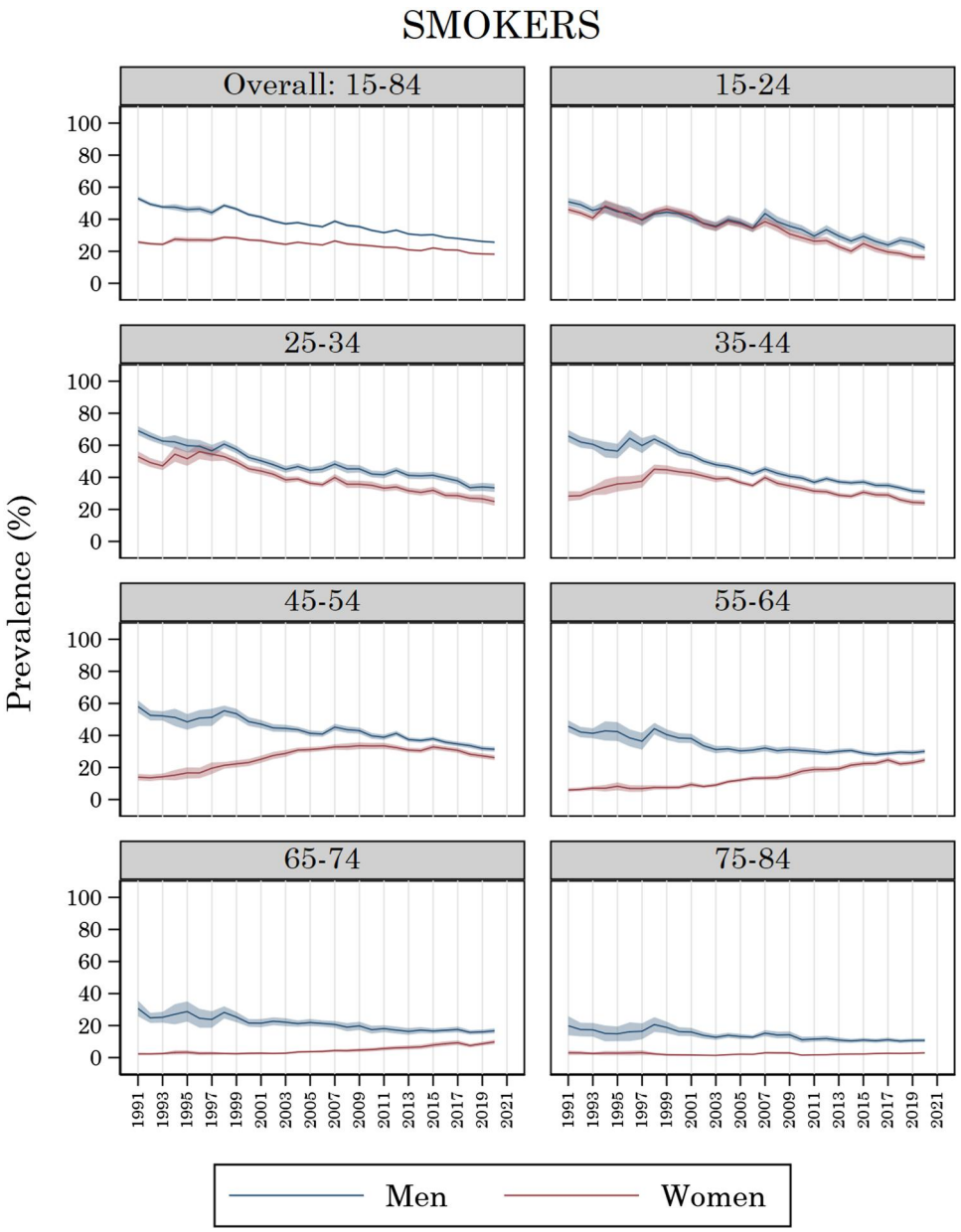

Figure 1. Annual prevalence of smokers, by sex and age group, from 1991 to 2020, with 95% confidence intervals.

smoking initiation and cessation probabilities and prevalence of S, ExS, and NS, as well as smoking intensity. As in our case, the set of S, ExS, and NS in each unit of analysis is exhaustive. Several papers applying the proposal of these authors show good precision of the estimates with narrow confidence intervals [25, 26]. However, some of them report better estimates for men than for women [27]. In addition, in the case of the youngest individuals under study, it is noted that prevalence estimates derived from the age–period–cohort model might be less reliable, as they are based on a smaller amount of data [28, 29]. These studies seek to explain the evolution of the prevalence series of S and ExS considering the birth cohort. However, in our study, the aim is to improve the precision of the prevalence estimates in small areas obtained directly from surveys. Since we work with combinations of sex and 5-year age groups we cannot apply age–period–cohort models.

Other studies have reconstructed the smoking prevalence series in Spain [4–6] based on the aggregation of 2 or 3 years of the NHS. Fernandez *et al.* [4] reconstructed the series of prevalence of S from the period 1910–1990. Given that this period does not coincide with that considered in our study, the results cannot be compared. Both Bilal *et al.* [5] and Martínez-Sánchez *et al.* [6] reconstructed the series of prevalence of S, the former for the period 1940–2007 using NHS data from 2003 and 2006, and the latter for the period 1989–2011 using NHS data from 2003, 2006, and 2011. The results of these two studies show a common, sharp downward trend in the prevalence of men S, and a stabilization and slight decrease in the prevalence of women S in the 2000s. This agrees with what is observed in our study, though our point estimates of prevalence are slightly higher. This could be due to differences in the methodology applied.

Our study has a series of limitations. The 2017 NHS and the 2009, 2014, and 2020 EHS only partially record the age of smoking initiation and cessation. To solve this inconvenience, the MICE method was applied. This is one of the most widely used methods for imputing missing data and provides good results [30–32]. In addition, the NHS and EHS do not allow the identification of relapses in ExS. This could underestimate the prevalence of S in our analysis. In our

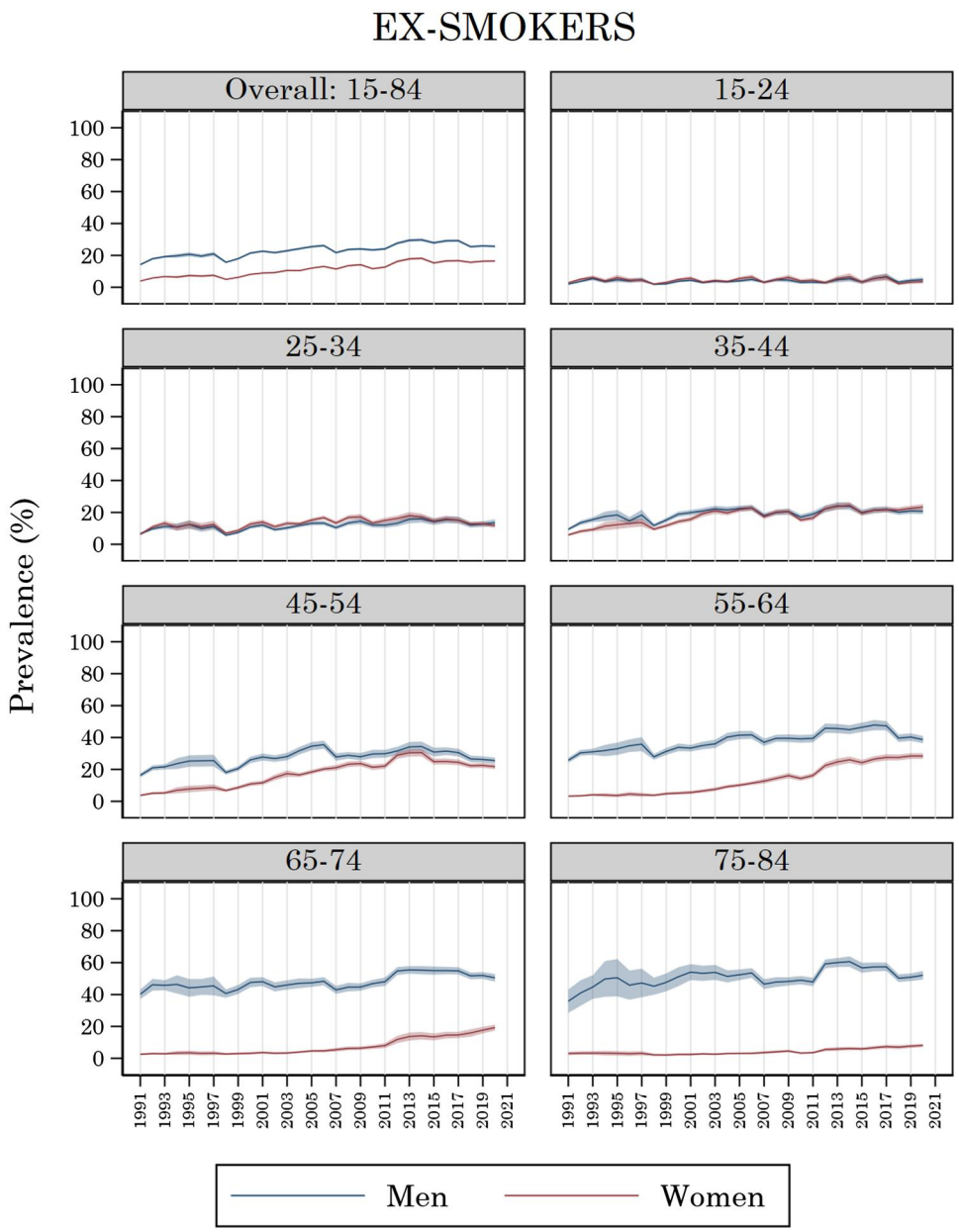

Figure 2. Annual prevalence of ex-smokers, by sex and age group, from 1991 to 2020, with 95% confidence intervals.

study, ExS are those who had been S but were not smoking at the time of the survey, independent of the time since cessation. In Holford *et al.*, ExS is defined as a smoking cessation period of at least two years before the date of the survey. But this definition also has its limitations, since it could underestimate the prevalence of ExS. Another limitation of our study is the use of hospitalizations due to lung cancer as auxiliary information, instead of lung cancer incidence, a well-known indicator of smoking. In our case, incidence data were not available, so hospitalizations were used as a proxy.

A strength of this study is the use of all NHS and EHS editions available from 1993 until the present. Therefore, a period of three decades was completed without having to go back more than 2 or 3 years, avoiding the problem of loss of representativeness due to age. In addition, the prevalence series of S, ExS, and NS were simultaneously estimated, resulting in a sum total of 100% in each area and year. Furthermore, the series was estimated in areas defined by sex and 5-year age group, providing a very detailed level of disaggregation. Another strength is the application of a model with random area and time effects which has been previously validated [33], making it possible to obtain more precise and smoother estimates. Our results show a general improvement in the precision of estimates of S, and particularly of ExS.

For the first time a detailed, year-by-year analysis of the historical series of smoking in Spain since 1991, with a breakdown by sex and age, has been made. In addition to surveillance, having precise, detailed annual estimates of prevalence is important as a benchmark to establish ecological relations between smoking and disease, the impact of tobacco control laws or to attribute mortality.

The results of this study, as well as its methodology, could be applied to other health determinants, making it possible to give a detailed account of evolution and impact on a population by sex and age.

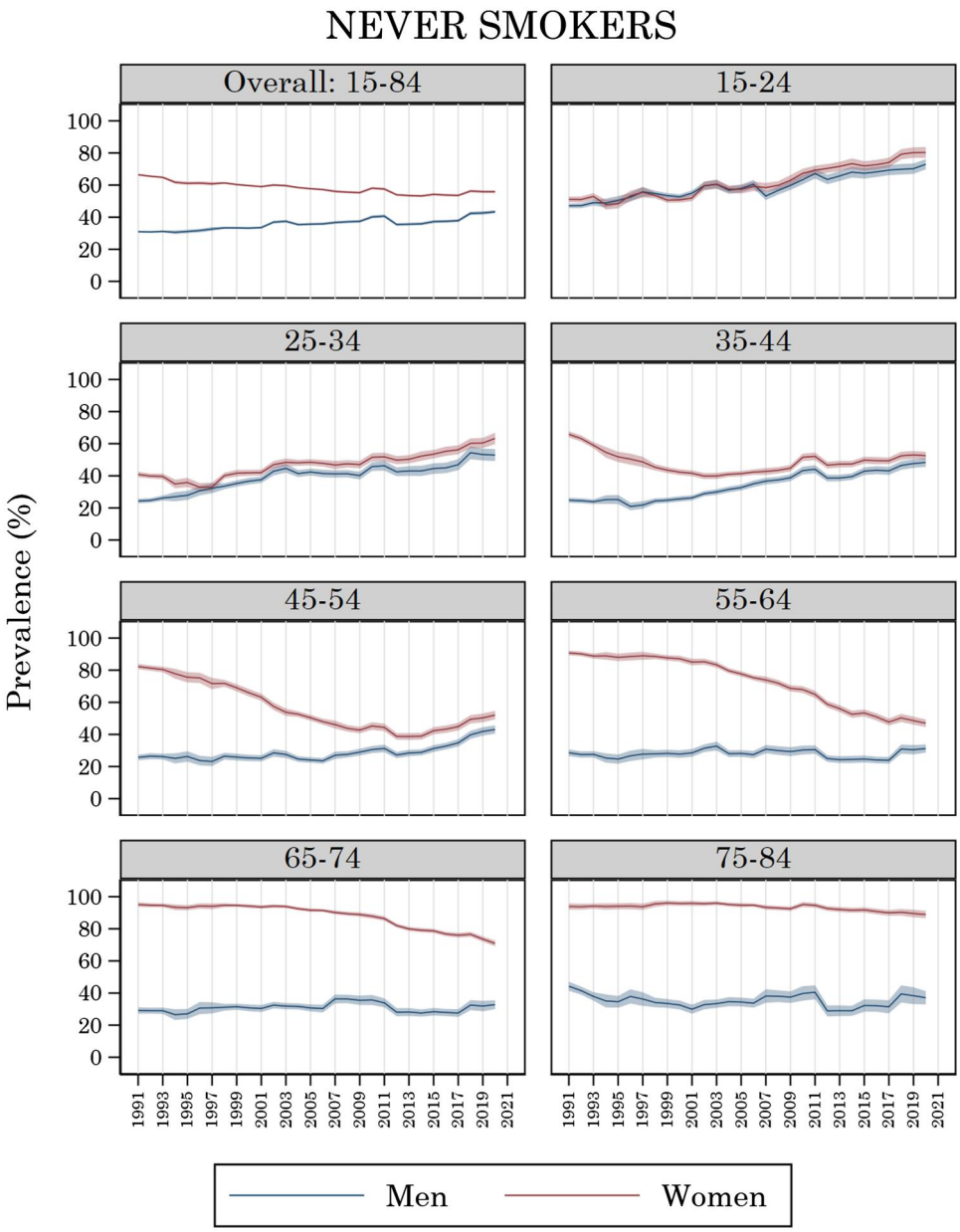

Figure 3. Annual prevalence of never smokers, by sex and age group, from 1991 to 2020, with 95% confidence intervals.

Author contributions

Conceptualization: Mónica Pérez-Ríos, Esther López-Vizcaíno, and María Isolina Santiago-Pérez; Technical-guidance: Esther López-Vizcaíno and María Isolina Santiago-Pérez; Data curation: Carla Guerra-Tort; Supervision: Alberto Ruano-Ravina and Mónica Pérez-Ríos; Writing-original draft: Carla Guerra-Tort; Visualization: Julia Rey-Brandariz, Cristina-Candal-Pedreira, Leonor Varela-Lema, and Iñaki Galán; Writing-review; Julia Rey-Brandariz, Cristina-Candal-Pedreira, and Leonor Varela-Lema; Editing: Esther López-Vizcaíno, María Isolina Santiago-Pérez, and Iñaki Gálan; Funding acquisition: Mónica Pérez-Ríos. All authors read and approved the final manuscript.

Supplementary data

Supplementary data are available at *EURPUB* online.

Conflict of interest: None declared.

Funding

This study was supported by the Instituto de Salud Carlos III (ISCIII) [grant number "PI22/00727"] and co-financed by the European Union.

Data availability

All data used in this study are publicly available. Sources of information and access links can be found in the Supplementary Material.

Ethics

Not applicable.

Key points

- A year-by-year series of smoking prevalence in Spain has been completed.
- Smoking prevalence in Spain has decreased overall between 1991 and 2020.
- Detailed analysis by age shows different trends in prevalence among women.
- The proposed methodology may be used to characterize the evolution of other health determinants.
- The year-by-year prevalence series estimated will facilitate future analysis.

References

1 Reitsma MB, Kendrick PJ, Ababneh E *et al.* Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet* 2021;**397**:2337–60.

2 European Commission, Directorate-General for Communication, Directorate-General for Health Food Safety. Special Eurobarometer 506: Attitudes of Europeans towards tobacco and electronic cigarettes. 2021. https://europa.eu/eurobarometer/surveys/detail/2240 (12 February 2024, date last accessed).

3 Instituto Nacional de Estadística. Encuesta Nacional de Salud en España. 2018. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176783&menu=ultiDatos&idp=1254735573175 (31 January 2024, date last accessed).

4 Fernandez E, Schiaffino A, Borras JM *et al.* Prevalence of cigarette smoking by birth cohort among males and females in Spain, 1910-1990. *Eur J Cancer Prev* 2003;**12**:57–62.

5 Bilal U, Fernandez E, Beltran P *et al.* Validation of a method for reconstructing historical rates of smoking prevalence. *Am J Epidemiol* 2014;**179**:15–9.

6 Martínez-Sánchez JC, Martinez-Sanchez JM, Bilal U *et al.* Sex and age specific projections of smoking prevalence in Spain: a Bayesian approach. *Nicotine Tob Res* 2018;**20**:725–30.

7 López-Vizcaíno E, Lombardía MJ, Morales D. Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. *J R Stat Soc Ser A Stat Soc*2015;**178**:535–65.

8 Santiago-Perez MI, Lopez-Vizcaino E, Perez-Rios 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**:1.

9 Instituto Nacional de Estadística. Estadística del Padrón continuo. 2023. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177012&menu=ultiDatos&idp=1254734710990 (11 January 2024, date last accessed).

10 Instituto Nacional de Estadística. Censos de población y viviendas. 2023. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176992&menu=ultiDatos&idp=1254735576757 (11 January 2024, date last accessed).

11 Instituto Nacional de Estadística. Encuesta de población activa. 2021. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595 (11 January 2024, date last accessed).

12 Ministerio de Sanidad, Gobierno de España. Registro de Actividad de Atención Especializada—RAE-CMBD. 2022. https://pestadistico.inteligenciadegestion.sanidad.gob.es/publicoSNS/S/rae-cmbd (12 January 2024, date last accessed).

13 Instituto Nacional de Estadística. Cifras de población. 2022. https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736176951&menu=resultados&idp=1254735572981 (11 January 2024, date last accessed).

14 van Buuren S, Oudshoorn K. *Flexible Multivariate Imputation by MICE*. Leiden, The Netherlands: Netherlands Organization for Applied Scientific Research (TNO), 1999, 1–20.

15 González-Manteiga W, Lombardía MJ, Molina I *et al.* Bootstrap mean squared error of small-area EBLUP. *J Stat Comput Simul* 2008;**78**:443–62.

16 Cohen R, Bloom B. Trends in health insurance and access to medical care for children under age 19 years: United States, 1998-2003. In: *Advance Data from Vital Health Statistics; No 355*. Hyattsville, MD: National Center for Health Statistics, **2005**, 1–12.

17 StataCorp. *Stata Statistical Software*. College Station. TX: StataCorp LLC, 2021.

18 López-Vizcaíno E, Lombardía M, Morales D. mme: Multinomial Mixed Effects Models. R package version 0.1-6. 2019. https://cran.r-project.org/web/packages/mme/index.html (10 April 2024, date last accessed).

19 Joinpoint Regression Program V4.9.1. *A Statistical Methodology and Applications Branch, Surveillance Research Program*. National Cancer Institute, 2022.

20 Fernández E, Schiaffino A, García M *et al.* Prevalencia del consumo de tabaco en España entre 1945 y 1995. Reconstrucción a partir de las Encuestas Nacionales de Salud. *Med Clín* 2003;**120**:14–6.

21 Borras JM, Fernandez E, Schiaffino A *et al.* Pattern of smoking initiation in Catalonia, Spain, from 1948 to 1992. *Am J Public Health* 2000;**90**:1459–62.

22 Birkett NJ. Trends in smoking by birth cohort for births between 1940 and 1975: a reconstructed cohort analysis of the 1990 Ontario Health Survey. *Prev Med* 1997;**26**:534–41.

23 Brenner H. A birth cohort analysis of the smoking epidemic in West Germany. *J Epidemiol Community Health* 1993;**47**:54–8.

24 Holford TR, Levy DT, McKay LA *et al.* Patterns of birth cohort-specific smoking histories, 1965-2009. *Am J Prev Med* 2014;**46**:e31–7.

25 Holford TR, Levy DT, Meza R. Comparison of smoking history patterns among African American and White cohorts in the United States born 1890 to 1990. *Nicotine Tob Res* 2016;**18** Suppl 1:S16–29.

26 Salgado MV, Mok Y, Jeon J *et al.* Smoking patterns by birth cohort in Argentina: an age-period-cohort population-based modeling study. *Lancet Reg Health Am* 2024;**36**:100823.

27 Manuel DG, Wilton AS, Bennett C *et al.* Smoking patterns based on birth-cohort-specific histories from 1965 to 2013, with projections to 2041. *Health Rep* 2020;**31**:16–31.

28 Tam J, Jaffri MA, Mok Y *et al.* Patterns of birth cohort–specific smoking histories in Brazil. *Am J Prev Med* 2023;**64**:S63–71.

29 Levy DT, Tam J, Jeon J *et al.* Summary and concluding remarks: patterns of birth cohort–specific smoking histories. *Am J Prev Med* 2023;**64**:S72–9.

30 Azur MJ, Stuart EA, Frangakis C *et al.* Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;**20**:40–9.

31 Mera-Gaona M, Neumann U, Vargas-Canas R *et al.* Evaluating the impact of multivariate imputation by MICE in feature selection. *PLoS One* 2021;**16**:e0254720.

32 Baneshi MR, Talei AR. Does the missing data imputation method affect the composition and performance of prognostic models? *Iran Red Crescent Med J* 2012;**14**:31–6.

33 Guerra-Tort C, López-Vizcaíno E, Santiago-Pérez MI *et al.* Validation of a small-area model for estimation of smoking prevalence at a subnational level. *Tob Induc Dis* 2023;**21**:112.

European Journal of Public Health, 2025, 35, 263–269
https://doi.org/10.1093/eurpub/ckaf015
Original Manuscript