Compartmental models in epidemiology: Application on Smoking Habits in Tuscany

Modelli compartimentali in epidemiologia: Applicazione sull'abitudine al fumo in Toscana

Alessio Lachi, Cecilia Viscardi, Maria Chiara Malevolti, Giulia Carreras, and Michela Baccini

Abstract We developed a compartmental model to describe the smoking habits evolution in Tuscany, relying on flexible modelling of age and sex-dependent smoking transition rates. Calibrating on observed data, we estimated the prevalence of current, former and never smokers in the population, and quantified the fraction of deaths attributable to smoking from 1994 to 2033. The model results indicate that smokers prevalence is decreasing over time. We speculate that this reduction could be related to health policies implemented up to now. Results highlight also that smoking habits are different among males and females.

Abstract *Abbiamo sviluppato un modello compartimentale che descrive l'evoluzione delle abitudini al fumo in Toscana, modellando in modo flessibile i tassi di fumo, età e sesso dipendenti. Calibrando su dati osservati, abbiamo stimato la prevalenza di fumatori, ex fumatori e non fumatori nella popolazione e abbiamo quantificato la frazione di decessi attribuibili al fumo dal 1994 al 2033. I risultati del modello indicano che la prevalenza dei fumatori sta diminuendo nel tempo. Tale riduzione potrebbe essere dovuta alle politiche sanitarie sinora attuate. I risultati mostrano inoltre che le dinamiche di abitudine al fumo sono diverse tra uomini e donne.*

Key words: Compartmental Model, Smoke, Tobacco control, Prevalence, Population attributable fraction.

Alessio Lachi
DiSIA (UNIFI), Viale Morgagni 65 - Firenze, e-mail: alessio.lachi@unifi.it

Cecilia Viscardi
DiSIA (UNIFI), Viale Morgagni 65 - Firenze, e-mail: cecilia.viscardi@unifi.it

Maria Chiara Malevolti
ISPRO, Via Cosimo il Vecchio 2 - Firenze, e-mail: m.malevolti@ispro.toscana.it

Giulia Carreras
ISPRO, Via Cosimo il Vecchio 2 - Firenze, e-mail: g.carreras@ispro.toscana.it

Michela Baccini
DiSIA (UNIFI), Viale Morgagni 59 - Firenze, e-mail: michela.baccini@unifi.it

1 Introduction

Smoking is a major risk factor for many common chronic diseases. In particular, over 85% of lung cancers are attributable to smoking. Also, smoking reduces length and quality of life [13].

The World Health Organization Framework Convention on Tobacco Control (WHO FCTC) considers the implementation of tobacco control policies (TCP) as the primary prevention strategy to reduce the prevalence of smokers in the population and the burden of mortality and morbidity attributable to smoking.

Aggregate or compartmental models allow to describe the dynamics of smoking habits in time and to compare the effectiveness of TCPs. They start from an initial population, divided into non overlapping compartments according to observed smoking prevalence, and reproduce the evolution of their sizes through a system of continuous-time dynamic equations [11, 9]. The mechanistic nature of such models results in the easiness of simulation of the system evolution.

This study aims to develop a compartmental model that describes the evolution of smoking habits from 1994 to 2033 in Tuscany, a region of central Italy, estimating the rates of starting, quitting, and relapsing smoking, and forecasts the prevalence of actual, former, and never smokers over time, allowing the evaluation of the impact of previous TCPs.

2 Methods

2.1 Model specification

We considered a compartmental smoking model (CSM) in which people, classified by gender (g) and age (a), are grouped into non overlapping compartments based on their smoking status: Current (C), Never (N), Former (F) smokers, and the related deaths compartments. The compartments C and F are further divided into subgroups denoted by C_i and $F_{i,c}$ for $i \in \{l,m,h\}$ and $c \in \{0,...,15\}$. Subscripts l, m and h stand for smoke intensity: low (cig/day$<$10), medium (10$\leq$cig/day$\leq$19), and high (cig/day$>$19). Subscript c denotes the time from smoking cessation in years.

Transitions of the individuals from a given compartment to another one determine flows that generate changes in the size of the compartment. Such transitions occur with probabilities governed by the annual rate of starting (γ), quitting (ε) and relapsing (η_c) smoking, and by the mortality rates (δ_{C_i}, δ_N, and $\delta_{F_{i,c}}$).

The assumptions underlying the model are: 1) the population is closed to immigration and emigration but opened to newborns, (birth rates from the Italian Institute of Statistics (ISTAT) database of 2011); 2) people start smoking in age $\in \{14,...,24\}$ ($\gamma(a) = 0$ for $a \in \{0,...,13\} \cup \{25,...,100\}$; 3) people are allowed becoming former smokers in age $\in \{25,...,100\}$ ($\varepsilon(a) = 0$ for $a \in \{0,...,24\}$); 4) the probability of starting, quitting, and relapsing smoking are independent from mortality.

Accordingly, the dynamic of the system is described by the following differential equations defined for each cohort [1] (y), and gender (g):

$$\begin{cases} \frac{dN^y(t)}{dt} = & N^y(t)\Big(1-\delta_N(t-y)\Big)\Big(1-\gamma(t-y)\Big) \\ \frac{dC_i^y(t)}{dt} = & C_i^y(t)\Big(1-\delta_{C_i}(t-y)\Big)\Big(1-\varepsilon(t-y)\Big)+N^y(t)\Big(1-\delta_N(t-y)\Big)\gamma(t-y)\pi_{C_i}(t-y)\mathbb{1}_{14\leq(t-y)\leq 24}+ \\ & \sum_{c=1}^{15+}F_{i,c-1}^y(t)\Big(1-\delta_{F_c}(t-y)\Big)\eta_c\mathbb{1}_{(t-y)>24} \\ \frac{dF_i^y(t)}{dt} = & \sum_{c=1}^{15+}F_{i,c-1}^y(t)\Big(1-\delta_{F_c}(t-y)\Big)\Big(1-\eta_c\Big)+C_i^y(t)\Big(1-\delta_{C_i}(t-y)\Big)\varepsilon(t-y) \\ \frac{dDN^y(t)}{dt} = & N^y(t)\delta_N(t-y) \\ \frac{dDC_i^y(t)}{dt} = & C_i^y(t)\delta_{C_i}(t-y) \\ \frac{dDF_i^y(t)}{dt} = & \sum_{i\in\{l,m,h\}}\sum_{c=1}^{15+}F_{i,c-1}^y(t)\delta_{F_c}(t-y). \end{cases} \tag{1}$$

Note that in Eq. (1) a can be viewed as $t-y$ and π_{C_i} represents the percentage of current smokers for each level of intensity $i\in\{l,m,h\}$.

We considered γ, ε, and η_c as unknown parameters to be estimated. In particular, we define γ and ε as natural cubic regression splines of a with 4 and 3 degrees of freedom:

$$\gamma(a)=s(a;\psi_0,\psi_1,\psi_2,\psi_3);\quad \varepsilon(a)=s(a;\phi_0,\phi_1,\phi_2).$$

The relapsing smoke rate, η_c is modelled as a negative exponential function of time from smoking cessation c:

$$\eta(c)=\alpha\beta e^{-12c\beta}$$

where β governs how fast the relapse rate declines over time from smoking cessation, and α governs the lifetime probability of no relapse [7].

The parameters π_{C_i} are fixed to values taken from ISTAT surveys [8].

The mortality rates δ_N, δ_{C_i} and $\delta_{F_{i,c}}$ are assumed gender-specific and constant over time. For each age we computed the mortality rates as follows:

$$\delta_N(a)=\delta_{Pop}(a);\ \ \delta_{C_i}(a)\approx RR_{C_i}\times\delta_{Pop}(a);\ \ \delta_{F_{i,c}}(a)\approx RR_{F_{i,c}}\times\delta_{Pop}(a).$$

We relied on the Relative Risks (RR) estimated in [12] and the mortality $\delta_{Pop}(a)$ rate taken from ISTAT database of 2011.

[1] People born in the same year

2.2 Model calibration and uncertainty quantification

Let us denote by $\theta = (\psi_0, \psi_1, \psi_2, \psi_3, \phi_0, \phi_1, \phi_2, \alpha, \beta)$ the vector of the parameters to be estimated and by $p^{\theta}(t,a) = \left(p_C^{\theta}(t,a), p_N^{\theta}(t,a), p_F^{\theta}(t,a)\right)$ the vector of prevalence computed from the sizes of compartments at time t for the age a [2], given a specific value of θ. A calibration procedure consists of searching the vector θ leading to $p^{\theta}(t,a)$ as closed as possible to $p^{obs}(t,a) = \left(p_C^{obs}(t,a), p_N^{obs}(t,a), p_F^{obs}(t,a)\right)$, the vector of the observed prevalence.

We calibrated our model on the prevalence estimates derived from the multipurpose household surveys data "Aspect of daily life" (AVQ) [8] carried out by ISTAT. Since 1993 the multipurpose surveys AVQ collect fundamental information related to the daily life of individuals and families. In particular, we considered $t \in \{1994, ..., 2019\}$ and we simulated the evolution of the system up to 2033. The model was calibrated separately by gender.

To compare observed and simulated trajectories we considered the following objective function where $H(\cdot,\cdot)$ denotes the Hellinger distance [6]:

$$min\left[\frac{1}{t \times a}\sum_{t,a} H\left(p^{\theta}(t,a), p^{obs}(t,a)\right)\right] = \frac{1}{t \times a \times \sqrt{2}}\sum_{t,a}\sqrt{\sum_{i \in \{C,F,N\}}\left(\sqrt{p_i^{\theta}(t,a)} - \sqrt{p_i^{obs}(t,a)}\right)^2}. \quad (2)$$

We performed a constrained optimization procedure resorting to the R package `nloptr` [14]. To take into account the sampling variability and quantify the uncertainty around point estimates we used a parametric bootstrap procedure [4]. Following an approach similar to [3], we estimated percentile bootstrap confidence intervals assuming that each prevalence followed a Dirichlet distribution, $p^{\theta}(t,a) \sim Dirichlet\left(C^{\hat{\theta}}(t,a), N^{\hat{\theta}}(t,a), F^{\hat{\theta}}(t,a)\right)$ where $C^{\hat{\theta}}$, $F^{\hat{\theta}}$, and $N^{\hat{\theta}}$ represent the size of the compartments corresponding to the best estimate $\hat{\theta}$ derived by minimizing Eq. 2. The sampling procedure was repeated n=1000 times to obtain bootstrap replicates of the unknown parameters used to compute confidence intervals.

We also computed population attributable fraction (PAF) for all cause of death as defined in [5] with never smokers as counterfactual level. These PAFs measure the proportion of deaths that would be avoided if the smoking risk factor in the population were eliminated.

3 Results and discussion

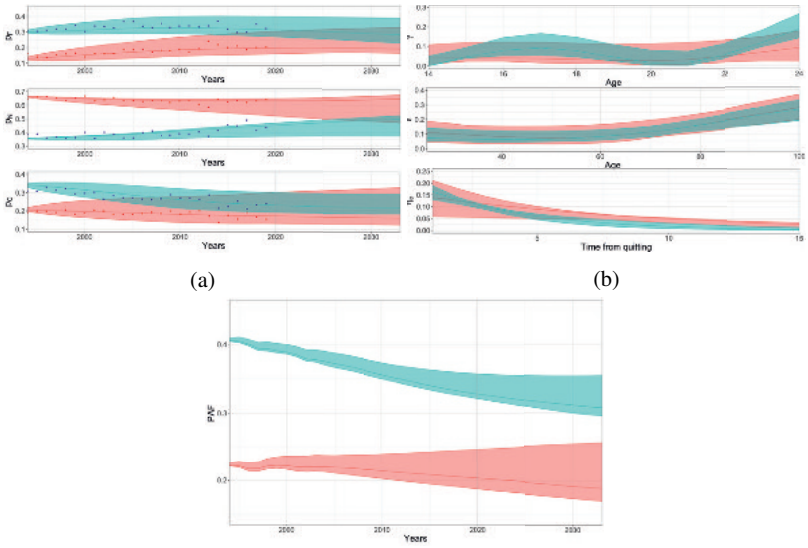
Fig. 1 panel a) shows the estimated prevalence from 1994 to 2033 for each gender. Looking at the observed data (blue and red dots), we observe an adequate model

[2] We considered the age classes 14-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75+

fit. Our forecasts suggest that the smoking prevalence will decrease in the next 15 years. This reduction may be due to the health policies so far implemented.

Panel b) shows the estimates of the rate of starting, quitting and relapsing smoking. From the comparison between rates of starting among males and females, we can see that males are more likely to start smoking than females. Moreover, the female rate exhibits an almost linear behaviour over time while the male rate shows a peak at seventeen. As regards the two rates of quitting, the two groups have the same behaviour, while the relapse rate among females becomes greater than among males after 5 years from quitting.

Panel c) represents PAFs. Our model predicts a reduction of PAF in the next years, due to the reduction of smoking prevalence. Males have a higher PAF than females, due to the higher prevalence of smoking among males.



(a) (b)

(c)

Fig. 1: Figure shows the prevalence estimates (a), the estimates of $\gamma - \varepsilon$ and η_c (b), and PAF (c) along with their 90% confidence bands by gender (males in blue and female in pink).

Several compartmental models have been developed to forecast future smoking rates and assess the impact of TCP [10, 9, 1, 2]. The Simsmoke model developed by Levy [10] has been largely used and applied to several countries, including Italy [9]. The model developed by Carreras et al [1, 2] extended the Simsmoke model by considering time from cessation in former smokers and by taking into account relapsing smoking. Our model adds further elements of novelty to this last model by considering also smoking intensity as a factor affecting mortality. Moreover, we introduced in the compartmental model cubic splines for estimating starting and

quitting smoking rates, thus obtaining more realistic trajectories. Finally, we evaluated parameter variability by estimating confidence intervals based on a parametric bootstrap procedure.

The model has several limitations. In particular, we assume the closeness of the population to immigration and emigration as well as constant mortality, birth rates, and, more in general, constant transition rates. Future work should include multivariate splines to model the dependence of the rates from calendar years. We should also consider second-hand smoking and other risk factors such as environmental pollution.

References

1. Giulia Carreras, Silvano Gallus, Laura Iannucci, and Giuseppe Gorini. Estimating the probabilities of making a smoking quit attempt in italy: stall in smoking cessation levels, 1986-2009. *BMC Public Health*.
2. Giulia Carreras, Giuseppe Gorini, and Eugenio Paci. Can a national lung cancer screening program in combination with smoking cessation policies cause an early decrease in tobacco deaths in italy? *Cancer prevention research*, 2012.
3. Gerardo Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *KeAi*, 2017.
4. Bradley Efron and Ryan Tibshirani. An introdcution to the bootstrap. 1993.
5. GBD. Gbd 2019 tobacco collaborators. 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. *The Lancet*, 2021.
6. Ernst Hellinger. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *De Gruyter*, 1909.
7. Rudolf Hoogenveen, Pieter Van Baal, Hendriek Boshuizen, and Talitha Feenstra. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life. *BioMed Central*, 2008.
8. ISTAT. Indagine multiscopo della famiglia: Aspetti della vita quotidiana. 1993-2019. URL: `www.istat.it/it/archivio/91926`.
9. David Levy, Silvano Gallus, Kenneth Blackman, Giulia Carreras, Carlo La Vecchia, and Giuseppe Gorini. Italy simsmoke: the effect of tobacco control policies on smoking prevalence and smoking attributable deaths in italy. *BMC Public Health*, 2012.
10. David Levy, Leonid Nikolayev, Elizabeth Mumford, and Christine Compton. The healthy people 2010 smoking prevalence and tobacco control objectives: results from the simsmoke tobacco control policy simulation model (united states). *Cancer Causes and Control volume*, 2005.
11. David Levy, Hana Ross, Eduard Zaloshnja, Roland Shuperka, and Meriglena Rusta. The role of tobacco control policies in reducing smoking and deaths caused by smoking in an eastern european nation: results from the albania simsmoke simulation model. *Journal of Public Health*, 2008.
12. Michael Thun, Brian Carter, Diane Feskanich, Neal Freedman, Ross Prentice, Alan Lopez, Patricia Hartge, and Susan Gapstur. 50-year trends in smoking-related mortality in the united states. the new england journal of medicine. *The new england journal of medicine*, 2013.
13. WHO. Tobacco smoke and involuntary smoking, iarc monographs on the evaluation of carcinogenic risks to humans. *IARC*, 2004.
14. Jelmer Ypma, Steven Johnson, Hans Borchers, Dirk Eddelbuettel, Brian Ripley, Kurt Hornik, Julien Chiquet, Avraham Adler, Xiongtao Dai, Aymeric Stamm, and Jeroen Ooms. R interface to nlopt. URL: `cran.r-project.org/web/packages/nloptr/nloptr.pdf`.