Approximate Bayesian Inference for Smoking Habit Dynamics in Tuscany

Alessio Lachi, Cecilia Viscardi, and Michela Baccini

Abstract Smoking is a major risk factor for lung cancer, as well as for many other chronic diseases, and understanding smoking habits is essential to evaluate and compare tobacco control policies. We developed a compartmental model to describe the evolution of smoking habits in Tuscany, a region of central Italy. Our model relies on flexible modelling of age and sex-dependent probabilities of starting, quitting, and relapsing from smoking. Furthermore, we considered smoking intensity as a risk factor affecting mortality. The resulting model has an intractable likelihood function, so we used Approximate Bayesian Computation, a powerful simulation-based inference method, to provide posterior estimates of the model's parameters. Using these approximate posterior distributions, we predicted the prevalence of current, former, and never smokers in Tuscany up to 2043. The model results suggest that the prevalence of smokers will decrease over time.

Keywords Compartmental model · Smoking prevalence · Approximate Bayesian Computation

1 Introduction

Smoking is a major risk factor for many common chronic diseases. It reduces length and quality of life, and over 85% of lung cancers are attributable to smoking [19].

The World Health Organization Framework Convention on Tobacco Control (WHO FCTC) emphasizes the importance of tobacco control policies (TCPs) to

A. Lachi (✉) · C. Viscardi
DiSIA, University of Florence, Viale Morgagni 65, Firenze, Italy
e-mail: alessio.lachi@unifi.it

C. Viscardi
e-mail: cecilia.viscardi@unifi.it

M. Baccini
DiSIA, University of Florence, Viale Morgagni 59, Firenze, Italy
e-mail: michela.baccini@unifi.it

A. Avalos-Pacheco et al. (eds.), *Bayesian Statistics, New Generations New Approaches*, Springer Proceedings in Mathematics & Statistics 435,
https://doi.org/10.1007/978-3-031-42413-7_6

reduce the prevalence of smokers in the population and the burden of mortality and morbidity associated with smoking [20].

Compartmental models have been shown to be a powerful tool to describe the dynamics of smoking habits in the population and to assess the effectiveness of alternative TCPs. These models start with an initial population, divided into non-overlapping compartments according to observed smoking prevalence, and simulate the evolution of their sizes through a system of equations defined in discrete time. The mechanistic nature of these models allows for easy simulation of the system's evolution under different scenarios. In the literature, several compartmental models have been developed to forecast future smoking prevalence [3, 4, 14–16]. In particular, the SimSmoke model developed by Levy [15] has been widely used and applied to several countries, including Italy [14]. Carreras et al. [3, 4] extended the SimSmoke model to take into account that former smokers can relapse smoking and that the probability of this event depends on the time since cessation.

In a previous study [12], we developed a compartmental model aimed at describing the evolution of smoking habits from 1993 to 2043 in Tuscany, a region of central Italy. We estimated the probabilities of starting, quitting, and relapsing from smoking and we forecast the prevalence of current, former, and never smokers. We introduced several novel elements by modelling the dependence of the probabilities of starting and quitting smoking on age in a flexible way. Furthermore, we included smoking intensity as a risk factor affecting mortality. The resulting model provided a more realistic description of smoking habits, but it also introduced several complexities in the system of difference equations that governs the dynamics, making the inference more challenging.

The aim of this work is to provide a fully Bayesian estimate of a model largely inspired by the one presented in [12].

To overcome the intractability of the likelihood associated with the model, we exploited the easiness of simulating dynamics based on compartmental models and resorted to an Approximate Bayesian Computation approach.

2 Smoking Habit Compartmental Model

We consider a smoking habit compartmental (SHC) model in which people, classified by 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 sub-groups denoted by C_i and F_i for $i \in \{l, m, h\}$. Subscripts l, m and h stand for smoke intensity levels: low (cig/day$<$10), medium (10$\leq$cig/day$\leq$19), and high (cig/day$>$19).

Changes in smoking status are modelled through transitions of the individual from a given compartment to another. These transitions determine flows that modify the size of the involved compartments. The transitions allowed by the model occur with probabilities regulated by the probabilities of starting (γ), quitting (ε), and relapsing (η) smoking, and by the probabilities of deaths (δ). The probabilities of starting

and quitting smoking are assumed to be dependent on the age (a) and are modelled through natural cubic splines with 2 equidistant internal knots:

$$\ln \frac{\gamma(a)}{1-\gamma(a)} = s(a;\boldsymbol{\psi}) \quad \ln \frac{\varepsilon(a)}{1-\varepsilon(a)} = s(a;\boldsymbol{\phi}), \tag{1}$$

where $\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3)$ and $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3)$ are vectors of unknown parameters governing the cubic splines $s(\cdot;\cdot)$.

The probability of smoking relapse is assumed to be a function of the time from cessation (c):

$$\eta(c) = 1 - \exp[-\omega_0\omega_1 \exp(-\omega_1 c)], \tag{2}$$

where $\boldsymbol{\omega} = (\omega_0, \omega_1)$ is a vector of unknown parameters, with ω_0 governing the lifetime probability of no relapse and ω_1 governing the decline of the relapse over time [10].

The probabilities of death ($\delta_{C_i}(a)$, $\delta_{F_i}(a,c)$, and $\delta_N(a)$) are assumed to be dependent from the smoking status, the age $a \in \{1, ..., 100\}$, and also from smoking intensity and $c \in \{1, ..., 15+\}$ for former smokers. Accordingly, the dynamic of the population is described by the following system of difference equations defined in discrete time $t \in \{1, ..., T\}$, with the year as time-unit:

$$\begin{cases}
N(t;0) = \nu(t-1)\Big(1-\delta_N(0)\Big) & \\
N(t;a) = N(t-1;a-1)\Big(1-\delta_N(a)\Big)\Big(1-\gamma(a)\Big) & \forall\, a \in \{1, ..., 100\} \\
C_i(t;0) = 0 & \\
C_i(t;a) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\Big(1-\varepsilon(a)\Big) + N(t-1;a-1)\Big(1-\delta_N(a)\Big)\gamma_i(a) + & \\
\qquad + \sum_{c\geq 1} F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\eta(c) & \forall\, i \in \{l,m,h\}\,,\, a \in \{1, ..., 100\} \\
F_i(t;0,c) = 0 & \\
F_i(t;a,0) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\varepsilon(a) & \forall\, i \in \{l,m,h\}\,,\, a \in \{1, ..., 100\} \\
F_i(t;a,c) = F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\Big(1-\eta(c)\Big) & \forall\, i \in \{l,m,h\}\,,\, a \in \{1, ..., 100\}\,, \\
 & \quad c \in \{1, ..., 15+\} \\
D_N(t;0) = D_N(t-1;0) + \nu(t-1)\delta_N(0) & \\
D_N(t;a) = D_N(t-1;a) + N(t-1;a-1)\delta_N(a) & \forall\, a \in \{1, ..., 100\} \\
D_{C_i}(t;0) = 0 & \\
D_{C_i}(t;a) = D_{C_i}(t-1;a) + C_i(t-1;a-1)\delta_{C_i}(a) & \forall\, i \in \{l,m,h\}\,,\, a \in \{1, ..., 100\} \\
D_{F_i}(t;0,c) = 0 & \\
D_{F_i}(t;a,c) = D_{F_i}(t-1;a,c) + F_i(t-1;a-1,c-1)\delta_{F_i}(a,c) & \forall\, i \in \{l,m,h\}\,,\, a \in \{1, ..., 100\}\,, \\
 & \quad c \in \{1, ..., 15+\}.
\end{cases} \tag{3}$$

In Eq. (3), $\gamma_i(a)$ is defined as $\pi_{C_i}\gamma(a)$ and π_{C_i} represents the percentage of new current smokers that have a smoking intensity i. $\nu(t)$ denotes the number of new births at time t, which along with the sizes of the compartments at $t = 0$, is taken from the observed data.

Some of the more relevant assumptions underlying the model are:

(1) Each individual experiences at most one event per year (starting, quitting, relapsing, or death); (2) The population is closed to immigration and emigration but open to newborns and deaths; (3) The transition rates do not change over calendar time; (4) The distribution of smokers by smoke intensity, $(\pi_{C_l}, \pi_{C_m}, \pi_{C_h})$, is constant over calendar time during the study period; (5) Smokers do not change their smoke intensity during their entire life and if a former smoker of intensity i returns to smoking, he/she returns to being a current smoker of intensity i; (6) The probabilities of quitting and relapsing into smoking do not depend on the level of smoking intensity, but only on age; (7) People can start smoking between the ages of 14 and 34; (8) People can quit smoking only after 20 years of age; (9) The relapsing rate depends only on time since smoking cessation, but not on age; (10) The mortality rate of current smokers does depend on age but does not consider the time from starting smoking.

2.1 The Probabilistic Model

For identifiability reasons, we fix the parameters π_{C_i}, $\delta_{C_i}(a)$, $\delta_{F_i}(a, c)$, and $\delta_N(a)$ to values provided by the "National Institute of Statistics" (ISTAT) [7]. Let $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\omega})$ be the vector of parameters object of our inference. We chose to set on them non-informative prior distributions. In particular, we define the prior distributions for the spline parameters in Eq. (1) as $U[-10, 10]$, and the parameters in Eq. (2) as $U(0, 10]$. Note that, exploiting some knowledge about $\eta(c)$, we impose that the parameters governing the probability of smoking relapse assume positive values. In particular, $\omega_1 > 0$ guarantees that $\eta(c)$ is a decreasing function of the time from cessation and $\omega_0 > 0$ ensures that the rate of smoking relapse $(\omega_0\omega_1 \exp(-\omega_1 c))$ assumes positive values.

Observed data are the vectors of the prevalence of each smoking status, $p^{\text{obs}}(a, t)$=$(p_C^{\text{obs}}(a, t), p_F^{\text{obs}}(a, t), p_N^{\text{obs}}(a, t))$, for each age a and year t. Hereafter, for the sake of simplicity, the collection of all vectors $p^{\text{obs}}(a, t)$ will be denoted by y^{obs}.

Let us denote by X the number of individuals who transit from a generic compartment to another one. We assume $X \sim Binomial(n_x, q_x)$, where n_x is the number of individuals being allowed to transition and q_x is the probability of that transition. As an example, consider the number of smokers of age a with a low intensity that quit smoking at time t: n_x is the number of current smokers with low intensity and age a that do not die during the year t, and q_x is equal to $\varepsilon(a)$. The same reasoning applies to the number of individuals relapsing smoking and the number of deaths. While the number of individuals who start smoking at age a is distributed according to a Multinomial distribution with the vector of probabilities $(\gamma_l(a), \gamma_m(a), \gamma_h(a))$.

2.2 Approximate Bayesian Computation for the SHC Model

In the Bayesian framework, the mathematical object of interest is the posterior distribution $p(\boldsymbol{\theta}|y^{\text{obs}}) \propto p(\boldsymbol{\theta})p(y^{\text{obs}}|\boldsymbol{\theta})$, derived through Bayes's formula starting from the prior distribution ($p(\boldsymbol{\theta})$) and the likelihood function ($p(y^{\text{obs}}|\boldsymbol{\theta})$). The analytical expression for the posterior distribution is often not available, but Monte Carlo or Markov Chain Monte Carlo methods [22] allow obtaining samples from the posterior distribution to produce an approximation of the posterior quantities of interest. However, these methods require multiple point-wise evaluations of the likelihood function $p(y^{\text{obs}}|\boldsymbol{\theta})$. When dealing with complex models these evaluations may be computationally prohibitive or impossible. In such cases, a possible solution is to resort to likelihood-free methods. In particular, Approximate Bayesian Computation (ABC) is a broad class of likelihood-free algorithms that allow Bayesian inference for complex models. They do not require exact likelihood computation since they only require the ease of simulation of the data generative process. Compartmental models are classical examples of models with intractable likelihood functions and easy-to-simulate generative processes [2, 11]. In particular, in the SHC model, the intractability of the likelihood comes from the high number of involved compartments, the complexity of the model that expresses transition probabilities as a function of the age and time from cessation, and the presence of high-dimensional latent variables. Indeed, in our model observed data, y^{obs}, are only the collection of vectors of prevalence $p^{\text{obs}}(a, t)$ for each a and t. The number of transitions, as well as the size of all compartments at each point in time, represent latent variables.

The key idea behind ABC is to use Bayes' Theorem, as interpreted by Rubin in [23], to convert samples from the prior distribution into samples from the posterior distribution through comparisons between observed data and pseudo-data generated from a computer program reproducing the data generative process—i.e. the simulator. The basic ABC algorithm [21, 26] proceeds as follows: (1) draw S parameter proposals from the prior distribution; (2) for each $s \in \{1, ..., S\}$ give the parameter proposal $\boldsymbol{\theta}^{(s)}$ as an input to the simulator to produce pseudo-data $y^{(s)}$; (3) retain only parameter proposals such that $\rho(y^{(s)}, y^{\text{obs}}) \leq e$, where $\rho(\cdot, \cdot)$ is a distance function and e is a positive tolerance threshold. The output of the algorithm is a sample from an approximate posterior distribution, the accuracy of which depends on the tolerance threshold. For a comprehensive description of the method, we refer the reader to [25]. The literature includes more advanced sampling schemes, such as the Population Monte Carlo ABC presented in [1] and some adaptive versions inspired by it (see Algorithm 1).

Algorithm 1 Adaptive Population Monte Carlo ABC

for j in $1 : M$ **do**
  Simulate $\boldsymbol{\theta}_j^{(1)} \sim p(\boldsymbol{\theta})$ and $y \sim p(y|\boldsymbol{\theta}_j^{(1)})$ until $\rho(y, y^{\text{obs}}) < e$
  Set $w_j^{(1)} = \frac{1}{M}$
end for
for s in $2 : S$ **do**
  for j in $1 : M$ **do**
    Set $\tau_s^2 = 2Var(\boldsymbol{\theta}_j^{(s-1)})$
    Pick $\boldsymbol{\theta}_j^*$ from $\boldsymbol{\theta}_j^{(s-1)}$ with probabilities $w_j^{(s-1)}$
    Gen $\boldsymbol{\theta}_j^{(s)}|\boldsymbol{\theta}_j^* \sim MVN(\boldsymbol{\theta}_j^*, \tau_s^2)$ and $y_j^{(s)} \sim p(y|\boldsymbol{\theta}_j^{(s)})$
    Set $w_j^{(s)} \propto \dfrac{p(\boldsymbol{\theta}_j^{(s)})}{\sum_{m=1}^{M} w_m^{(s-1)} \phi\{\tau_s^{-1}(\boldsymbol{\theta}_m^{(s)} - \boldsymbol{\theta}_m^{(s-1)})\}} \mathbb{1}\{\rho(y_j^{(s)}, y^{\text{obs}}) < e_s\}$
  end for
  Select e_{s+1} using an adaptive strategy.
end for

Here, we relied on the strategy proposed in [13, Algorithm 5], where the threshold is automatically selected during the execution of the algorithm in a way ensuring that the tolerance level decreases from one iteration to the next.

In the case at hand, the comparison between observed and simulated trajectories of the vector of prevalence is based on the following average of Hellinger distances [9] denoted by $H(\cdot, \cdot)$:

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

where $p^{\boldsymbol{\theta}}(t, a)$ is the vector of prevalence computed from the sizes of compartments at time t for the age a, outputted by the simulator when the vector of parameter $\boldsymbol{\theta}$ is given as an input.

3 Results

We used data collected by ISTAT in the Multipurpose Surveys "Aspect of Daily Life" (AVQ) [6]. They include fundamental information related to the daily life of individuals and families in Italy. Each yearly survey enrols about 24,000 families and 54,000 persons distributed in about 850 Italian municipalities of different population sizes. We estimated the model described in Sect. 2.1 using data collected in Tuscany, a region in central Italy. In particular, our observed data are the vectors of the prevalence of never, current, and former smokers for each year from 1993 to 2019 and in each of the following age classes: 14–17, 18–19, 20–24, 25–34, 35–44, 45–54, 55–59, 60–64, 65–74, and 75+.

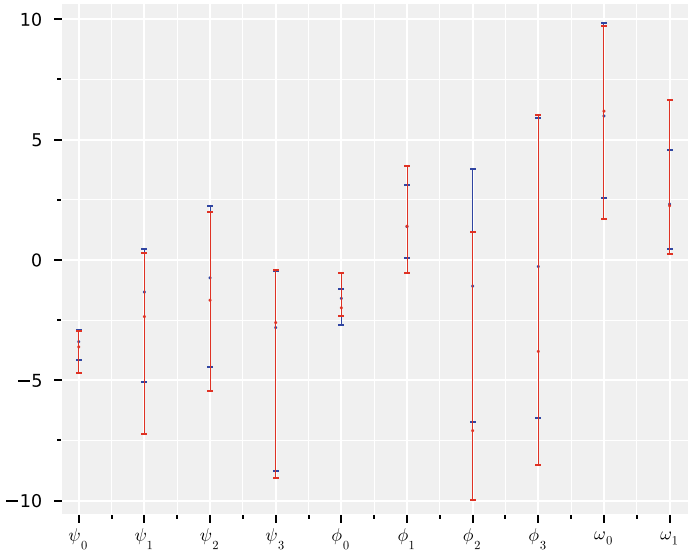

Fig. 1 Maximum a posteriori estimates along with 90% credible intervals for the model parameters for males (in blue) and females (in red)

We estimated separate models for each gender and simulated the evolution of the system up to 2043. We ran the ABC algorithms in parallel on 50 cores getting final effective sample sizes of 534 and 515 with ABC tolerance thresholds of 0.073 and 0.078, respectively for males and females.

Figure 1 reports the approximate posterior estimates of $\boldsymbol{\theta}$.

Using the approximate posterior distribution derived for $\boldsymbol{\theta}$, we estimated the prevalence of each compartment from 1993 to 2043, both for males and females, see Fig. 2 (Panel (a)). Looking at the observed data (blue and pink dots), we observe an adequate model fitting. The forecasts suggest a decrease in smoking prevalence over the next 24 years.

Panel (b) shows the posterior estimates of the probabilities of starting, quitting, and relapsing from smoking. They reveal that males are more likely to start and stop smoking, but also more likely to relapse, compared to females.

4 Discussion

Previous literature has used compartmental models to forecast future smoking rates and evaluate the impact of tobacco control policies (TCPs) [3, 4, 14, 15]. However, these models have relied on strong assumptions and have only provided point estimates without any uncertainty quantification. In [12], we introduced several novel elements to obtain more realistic trajectories, but also encountered a more complex

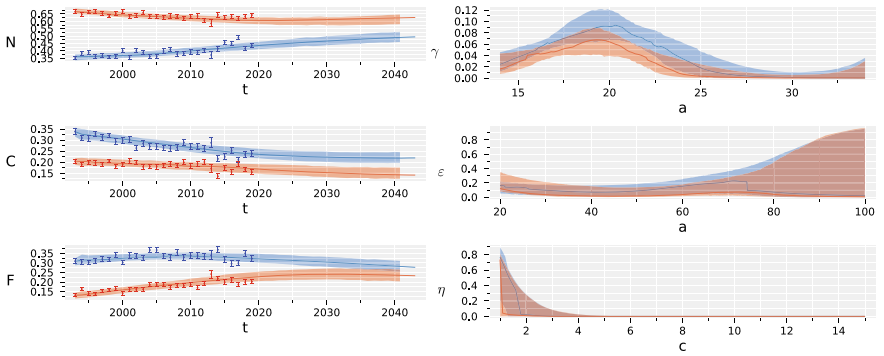

Fig. 2 Maximum a posteriori estimates and 90% credible intervals of the prevalence for never, current, and former smokers (Panel (**a**)). Maximum a posteriori estimates and 90% credible intervals of $\gamma(a)$, $\varepsilon(a)$, and $\eta(c)$, for males in blue and females in red (Panel (**b**))

compartmental model with an intractable likelihood. In [12], we provided both point estimates and frequentist confidence intervals using an optimization and bootstrap procedure [5, 8], respectively.

In this study, we employed a fully Bayesian approach using ABC to overcome the limitations of calibration methods based on optimization, such as the risk of getting stuck in local minima, as discussed in [5]. ABC allowed us to obtain point estimates as well as a quantification of uncertainty around parameter estimates and prevalence forecasts.

The presented model may be a valuable tool for evaluating the impact of tobacco control policies (TCPs) on future smoking habits and determining the population attributable fraction, which measures the proportion of deaths that could be avoided if smoking were completely eliminated or reduced to specific counterfactual levels.

However, it has several limitations. To address identifiability issues, we fixed certain parameters to values obtained from the literature or from local surveys. Additionally, the model relies on strong assumptions (closure of the population to immigration and emigration, constant mortality and new births, and constant transition probabilities over time). Some of them are strictly related to the Italian context. In particular, the assumptions regarding the age at which people can start or quit smoking are based on several research studies which show that very few people start smoking after 34 and quit before 20 years of age [17, 18]. This evidence is also supported by the data regarding our study period that are made available by ISTAT [7].

Global sensitivity analysis approaches [24] should be conducted to evaluate the robustness of the model results to different specifications of the fixed parameters and the impact of the structural assumptions.

Future developments, should also include prior sensitivity analyses. Finally, we plan to conduct a simulation study to evaluate the performance of the employed ABC algorithm.

References

1. Beaumont, M., Cornuet, J.-M., Marin, J.-M., Robert, C.: Adaptive approximate Bayesian computation. Biometrika (2009)
2. Brown, G., Porter, A., Oleson, J., Hinman, J.: Approximate Bayesian computation for spatial seir(s) epidemic models. Spatial and Spatio-temporal Epidemiology (2018)
3. Carreras, G., Gallus, S., Iannucci, L., Gorini, G.: Estimating the probabilities of making a smoking quit attempt in Italy: stall in smoking cessation levels, 1986–2009. BMC Public Health **183** (2012). http://www.biomedcentral.com/1471-2458/12/183
4. Carreras, G., Gorini, G., Paci, E.: Can a national lung cancer screening program in combination with smoking cessation policies cause an early decrease in tobacco deaths in Italy? Cancer Prev. Res. **5**(6), 874–882 (2012). https://doi.org/10.1158/1940-6207.capr-12-0019
5. Chowell, G.: Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. KeAi **2**(3), 379–398 (2017). https://doi.org/10.1016/j.idm.2017.08.001
6. Istituto Nazionale di Statistica. Indagine multiscopo della famiglia: Aspetti della vita quotidiana. www.istat.it/it/archivio/91926
7. Istituto Nazionale di Statistica. Istat. https://www.istat.it
8. Efron, B., Tibshirani, R.: An Introduction to the Bootstrap. Chapman and Hall/CRC (1993)
9. Hellinger, E.: Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. Journal für die reine und angewandte Mathematik **136**, 210–271 (1909). https://doi.org/10.1515/crll.1909.136.210
10. Hoogenveen, R., Van Baal, P., Boshuizen, H., Feenstra, T.: Dynamic effects of smoking cessation on disease incidence, mortality, and quality of life. BioMed Central **6**(1) (2008). https://doi.org/10.1186/1478-7547-6-1
11. Kypraios, T., Neal, P., Prangle, D.: A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. Math. Biosci. **287**, 42–53 (2017). https://doi.org/10.1016/j.mbs.2016.07.001
12. Lachi, A., Viscardi, C., Malevolti, M.C., Carreras, G., Baccini, M.: Compartmental models in epidemiology: application on smoking habits in Tuscany. In: Book of Short Papers SIS. Pearson (2022)
13. Lenormand, M., Jabot, F., Deffuant, G.: Adaptive approximate Bayesian computation for complex models. Comput. Stat. (2013)
14. Levy, D., Gallus, S., Blackman, K., Carreras, G., La Vecchia, C., Gorini, G.: Italy simsmoke: the effect of tobacco control policies on smoking prevalence and smoking-attributable deaths in Italy. BMC Public Health 12(709) (2012). https://doi.org/10.1186/1471-2458-12-709
15. Levy, D., Nikolayev, L., Mumford, E., Compton, C.: The healthy people 2010 smoking prevalence and tobacco control objectives: results from the simsmoke tobacco control policy simulation model (united states). Cancer Causes Control Vol. **16**(4), 359–371 (2005). https://cisnet.flexkb.net/mp/pub/CISNET_ModelProfile_LUNG_PIRE_001_01132012_83636.pdf, https://doi.org/10.1007/s10552-004-7841-4
16. Levy, D., Ross, H., Zaloshnja, E., Shuperka, R., Rusta, M.: 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. J. Public Health **16**(4), 189–98 (2008). https://doi.org/10.21101/cejph.a3486
17. Lugo, A., Asciutto, R., Pacifici, R., Colombo, P., la Vecchia, C., Gallus, S.: Smoking in Italy 2013–2014, with a focus on the young. Tumori J. (2013)
18. Lugo, A., Zuccaro, P., Pacifici, R., Gorini, G., Colombo, P., La Vecchia, C., Gallus, S.: Smoking in italy in 2015–2016: prevalence, trends, roll-your-own cigarettes, and attitudes towards incoming regulations. Tumori J. **103**(4) (2017). https://doi.org/10.5301/tj.5000644
19. World Health Organization. Personal Habits and Indoor Combustions: IARC monographs on the evaluation of carcinogenic risks to humans. International Agency for Research of Cancer (2012)

20. World Health Organization. Roadmap of actions to strengthen implementation of the who framework convention on tobacco control in the European region 2015–2025: making tobacco a thing of the past. International Agency for Research of Cancer (2015)
21. Pritchard, J., Seielstad, M., Perez-Lezaun, A., Feldman, M.: Population growth of human y chromosomes: a study of y chromosome microsatellites. Mol. Biol. Evol. (1999)
22. Robert, C., Casella, G.: Monte Carlo Statistical Methods. Springer (1999)
23. Rubin, D.: Bayesianly justifiable and relevant frequency calculations for the applies statistician. Ann. Stat. (1984)
24. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S.: Global Sensitivity Analysis: The Primer. Wiley (2007)
25. Sisson, S., Fan, Y., Beaumont, M.: Handbook of Approximate Bayesian Computation. Chapman & Hall (2019)
26. Tavare, S., Balding, D., Griffiths, R., Donnelly, P.: Inferring coalescence times from DNA sequence data. Genetics