UNIVERSITÀ DEGLI STUDI FIRENZE

UNIVERSITÀ DEGLI STUDI DI PERUGIA

University of Florence, University of Perugia, INdAM consortium in the CIAFM

Ph.D.
in Mathematics, Computer Science, Statistics

Curriculum in Statistics
Cycle XXXVI

Administrative office in University of Florence
Coordinator Prof. Matteo Focardi

Compartmental Models and Uncertainty Quantification in Epidemiology

Scientific disciplinary sector MED/01

Ph.D. candidate
Alessio Lachi

Supervisor
Professor Michela Baccini
Co-supervisor
Doctor Cecilia Viscardi

Coordinator
Professor Matteo Focardi

Academic years 2020/2023

University of Florence

Compartmental Models and Uncertainty Quantification in Epidemiology

by

Alessio Lachi

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

in the
Department of Statistics, Computer Science, Applications "Giuseppe Parenti"

May 14, 2024

"The strong principle for the real world is: never use a model if you don't know its limitations and side effects"

Nassim Nicholas Taleb

Abstract

Smoking is a major public health problem in the world. It is a leading cause of preventable diseases, including lung cancer, heart diseases, and respiratory problems. Efforts to combat smoking include awareness campaigns, stricter regulations, and support for cessation programs. The main goal of this dissertation is to produce a model for simulating the evolution of smoking habits in Tuscany, a region of central Italy, which can be easily adapted to other contexts. The developed model is based on a system of differential equations. These particular types of models in epidemiology are called compartmental models and are widely used for understanding dynamic population phenomena. Their mechanistic nature allows us to straightforwardly predict the evolution of a phenomenon by simulating its dynamic under different scenarios. However, the enormous complexity of these models makes the definition of the underlying likelihood function very difficult. For this reason in this thesis, we also present an overview of the suitable estimation methods for compartmental models, emphasizing the relevance of likelihood-free methods. Another important limitation due to the complexity of these models is given by the high uncertainty of the results and the difficulties in quantifying them. In this thesis, through the use of Global Sensitivity Analysis, we produce a robustification of the inference resulting from our model and conclude that the assumptions underlying our model are reasonable. Furthermore, we were able to assess the impact, in terms of the actual effectiveness, of implementing hypothetical tobacco control policies in Tuscany. As a conclusion to this thesis, we trace the evolution of sensitivity analysis over the years and assess its possible future progress.

Acknowledgements

I would like to extend my sincere thanks to Professor Michela Baccini. This thesis would not have been realized without her guidance. Professor Baccini, your influence and support in my academic journey will remain essential.

Family, you're the backbone of everything I do. Thanks for always having my back. And a special shout-out to my pets, Willy, my little cat who will always be in our hearts and Asia, the new addition to my family.

I want to give a big shout-out to all my buddies who've had my back throughout this academic adventure. You guys are awesome, and I'm lucky to call you friends. In particular, a huge thanks to my lifelong pals, Martina and Valentina. Your support has been like a rock, and I'm grateful for all the good times we've had together. To my new friends – Cecilia, Maria, Lorenzo, and Veronica – a heartfelt thank you for making this journey more interesting. Again, a special mention to my "co-tutor", Doctor, and friend Cecilia! Your support has been crucial.

A big thank you goes to Professor Andrea Saltelli for hosting me for four months in Barcelona. It has been a unique experience. Special thanks also go to Doctor Giulia Carreras and Doctor Giulia Cereda, who helped me in writing the articles that make up this thesis.

Finally, I want to thank an extraordinary person who took me into her heart, my girlfriend Diana.

Thanks to all of you!

Contents

List of Figures

Dedicated to my nephew, Mattia

Preface

Smoking is a significant public health concern worldwide. The habit of smoking tobacco not only harms individuals but also poses a substantial burden on society. Moreover, smoking is a leading cause of preventable diseases such as lung cancer, heart diseases, and respiratory problems. Additionally, exposure to secondhand smoke can also have detrimental effects. Efforts to combat this issue include public awareness campaigns, stricter regulations on tobacco sales, and support for smoking cessation programs. Individuals need to recognize the risks associated with smoking and make informed choices to protect their health and the well-being of those around them. Understanding the complex dynamics of tobacco use and its effects on the population thus remains paramount for crafting effective public health policies, and the development of models that simulate smoking dynamics under different Tobacco Control Policies (TCPs) is useful for comparing hypothetical future interventions. Since the 2000s, several models have been proposed [1–4] to simulates smoking dynamics and most of these are compartmental models. Compartmental models, starting from a baseline year, perform macro-simulations so that the population evolves through deaths, births and changes in smoking habits [1–3]. The SimSmoke model is the most used compartmental model in this field. Developed by [2] to capture smoking dynamics, it has been implemented in a wide number of countries including Italy [5–8]. Based on a simple procedure of calibration, that minimizes the difference between observed and predicted smoking dynamics, the SimSmoke model estimates the relevant parameters that govern the transitions between compartments and makes projections. Additionally, it explores short and long-term impacts on smoking dynamics and related health outcomes of different hypothetical TCPs, the effects of which are taken from literature [2, 3, 5, 6, 8–20].

This dissertation was motivated by the need for a deeper understating of smoking dynamics in Tuscany, a region in central Italy, and their impacts on population health. The thesis has three main objectives:

- *Formulation of the Smoking Habits Compartmental model.* We formulated an adequate compartmental model that, grounding on previous published works [3, 12, 21, 22], describes the evolution of smoking dynamics in Tuscany from 1993 to 2019 and forecasts

them until 2043. This model includes flexible terms for modelling the age-specific probability of starting and stopping smoking, and accounts for different levels of smoking intensity.

- *Systematisation of inference methods for compartmental models.* We provided a systematic overview of the inference methods for compartmental models, from both a frequentist and a Bayesian point of view, with a special focus on likelihood-free approaches.
- *Uncertainty assessment and impact evaluation.* We provided robustified inference from the model on Tuscany data, according to the methods proposed in [23], and we evaluated the expected impact of alternative TCPs considering all sources of uncertainty.

Formulation of the Smoking Habits Compartmental model

We developed a compartmental model that describes the evolution of smoking habits in Tuscany, a region of Central Italy, from 1993 to 2019, and forecasts them until 2043. The model assumes that at each point in time, the population is divided into non-overlapping groups called compartments, defined according to smoking status (never, current, and former smokers), age and sex [24]. Transitions between compartments are described by probabilistic rules and the annual evolution of the size of the compartments is governed by a system of difference equations.

In order to avoid identifiability problems, usual in complex compartmental models, we fixed some of the parameters to values from the literature, estimating the others via a two-step calibration. In particular we estimated the age-specific probabilities of starting and quitting smoking, the probability of relapsing smoking, and the mortality rates, performing a calibration on the observed prevalence of never, current, and former smokers for the years 1993-2019, arising from yearly local surveys. The objective function of the calibration was based on a simple discrepancy measure between observed and predicted prevalence.

Compared to previous models on smoking dynamics, we introduced a flexible modelling based on regression splines for the age-specific probabilities of starting and stopping smoking, usually assumed as constant over age. Similarly, we allowed the probability of relapsing into smoke to change with time since cessation. In addition, the model considered different levels of smoking intensity. The model was used not only to infer the model parameters but also to quantify the impact of smoking on health in terms of smoking-attributable deaths.

Notably, we addressed the problem of accounting for sampling variability, never considered in previously published works, through a parametric bootstrap procedure [25, 26], assuming a Dirichlet distribution on the smoking prevalence. In this way, we obtained confidence intervals for the parameters and compartment sizes. Moreover, we assessed the predictive performance of the model using cross-validation on a rolling basis, considering different projection horizons.

Ultimately, integrating a transparent formalization of the model equations with calibration, bootstrap, and cross-validation, we proposed a novel and reproducible pipeline for the statistical analyses in this field. However, it should be stressed that the estimation procedure used has some limitations. We estimated the parameters in a deterministic way, in the sense that we considered the distributional assumptions on the prevalence only in the bootstrap procedure but not in the calibration phase. While likelihood-based approaches are unfeasible in this framework, being the model extremely complex, likelihood-free inference methods such as Approximate Bayesian Computation algorithms would allow a full uncertainty quantification [22]. This consideration motivated the work described in the next section.

Systematisation of inference methods for compartmental models

Compartmental models often exhibit an intractable likelihood function. In some cases, it can be complex to specify an analytical form for the likelihood, due to the complexity of the model – e.g., due to the large number of compartments or to challenging definitions of the allowed transitions and related probabilistic rules. In other cases an analytical form of the likelihood is available but its point-wise evaluation is infeasible due to the presence of a large number of unobserved latent variables. In both cases, the estimation of the model parameters requires suitable methods.

In the literature, the existing reviews on the methods to make inferences on compartmental models focus only on very specific approaches McKinley et al. [27], Tang et al. [28] and there is still a lack of a critical and comprehensive overview.

In order to fill this gap, we critically presented different frequentist and Bayesian estimation strategies in compartmental models, providing a comparison among them. Particular attention is paid to the distinction between mathematical and statistical compartmental models and to how introducing stochastic components leads to the definition of a likelihood function associated with the model. We discussed some of the reasons that make the likelihood intractable and methods suitable for dealing with this intractability, such as likelihood-free methods. In particular, we focused on the frequentist approach that combines model calibration with parametric bootstrap, used for inference on Smoking Habits Compartmental (SHC) model (see previous section), and Approximate Bayesian Computation (ABC). Calibration and ABC algorithms are very similar in spirit. The main difference between them is that calibration uses, in the point estimation phase, the mathematical model, while ABC resorts only to the statistical model. The presented ABC strategy allows us to consider two sources of variability in a single procedure: the uncertainty over the parameter space described by prior distributions and the sampling variability reproduced by the simulator. Instead, the calibration must be combined with an adequate bootstrap procedure to quantify the sampling variability.

Using the SHC model for Tuscany as a real-world example, we showed the great potential of likelihood-free methods, that can retrieve estimates and forecasts even when dealing with very complex models that prevent the use of whatever likelihood-based method. From a comparison between calibration and ABC, we concluded that the results are coherent with each other, thus they give support to the reliability of both methods.

However, calibration and ABC do not take into account the propagation of uncertainty due to modeling choices and assumptions. For example, an important issue in compartmental models concerns the problem of parameter identifiability [26]. Complex models with many compartments obviously have many parameters governing the admitted transitions, but unfortunately observed data are often insufficient to estimate all of them. To overcome this problem, we need to keep some parameters fixed at values coming from the literature. This suggests that the estimation procedure may be heavily affected by the set of fixed parameters, thus indicating the importance of conducting sensitivity analyses on these model assumptions. Moreover, implementing a formal procedure to assess uncertainty propagation is crucial for enhancing our understanding of smoking dynamics. This practice is strongly recommended, especially during health policy evaluation phases [29]. These considerations, along with the evaluation of different TCPs, motivated the work described in the next section.

Uncertainty assessment and impact evaluation

Sensitivity analysis, is traditionally associated with scenario analysis, where individual model parameters are altered, and resultant variations in outcomes are examined. This approach, commonly known as “one at a time" (OAT) sensitivity analysis, relies on partial derivatives to isolate the impact of each parameter on the model output. While OAT sensitivity analysis may seem logical, as changes in output are unequivocally attributed to individual variables, it falls short in contexts where multiple parameter inputs interact to influence model outcomes. To address this limitation, a variance based approach called Global Sensitivity Analysis (GSA) was introduced [29]. GSA is the study of how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input.

In statistical modeling, uncertainty may arise from various sources, such as measurement errors, sampling variability, model misspecification, and unobserved variables. Uncertainty assessment can be integrated with GSA, through the use of sensitivity indexes [30]. Especially when the overall uncertainty around the output is high, performing GSA is paramount to understanding which factors mostly influence the answer to our research question. This can help to identify model assumptions that are most critical and quantify the relative impact of different sources of uncertainty on the research findings.

The assumptions addressed underlying the SHC model represents only one of the numerous modeling choices available [31]. While compartmental models serve as invaluable tools for representing real-world phenomena, they inherently rely on simplifying assumptions that may limit their fidelity to reality.

We proposed a procedure aimed at robustifying inference, forecasting, and TCPs assessment produced by the SHC model. We also illustrated the use of SHC model to predict the future impact of hypothetical interventions that act on the probabilities of starting and quitting smoking. This robustification procedure, which can also be reproduced in different models and contexts, takes into account uncertainties involved in model definition, inference, and forecasting, and consists of an uncertainty assessment and a variance-based GSA [29, 32].

The discussion of the main objectives of the thesis is organized as follows:

- In Chapter 1 we addressed the objective *Formulation of the Smoking Habits Compartmental model*. In this chapter, you will encounter a paper titled *A Compartmental Model for Smoking Dynamics in Italy: A Pipeline for Inference, Validation, and Forecasting Under Hypothetical Scenarios*. This paper, submitted to *BMC Medical Research Methodology*, successfully passed the first revision. Although a condensed version is published in Lachi et al. [21].

- In Chapter 2, we addressed the objective *Systematisation of inference methods for compartmental models*. In this chapter, you will find a paper entitled *Frequentist and Bayesian inference on compartmental models in epidemiology: A critical review with a focus on likelihood-free approaches*. This paper is under review at *Statistics in Medicine*. Although a shorter version of this paper is published in [22].

- In Chapter 3, we addressed the objective *Uncertainty assessment and impact evaluation*. In this chapter, you will find the draft of a paper, to be submitted to a *peer review journal*, entitled *Smoking dynamics in Tuscany (Italy) under alternative tobacco control policies: Robustifying inference and forecasting via uncertainty propagation and Global Sensitivity Analysis*.

- In Chapter 4, you will find a paper entitled *An Annotated Timeline of Sensitivity Analysis* that provide an history overview of the evolution of GSA. The paper is published by *Environmental Modelling and Software* [33].

- In Chapter 5, you will find the abstract of all the other published papers during the Ph.D. period and related to the dissertation topics [34–37].

Bibliography

[1] D. Mendez, K.E. Warner, and P.N. Courant. Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States. *American Journal of Epidemiology*, 148(3):249–258, 1998. ISSN 0002-9262, 1476-6256. doi: 10.1093/oxfordjournals.aje.a009632. URL https://academic.oup.com/aje/article-lookup/doi/10.1093/oxfordjournals.aje.a009632.

[2] D.T. Levy and K. Friend. A Simulation Model of Policies Directed at Treating Tobacco Use and Dependence. *Medical Decision Making*, 22(1):6–17, 2002. ISSN 00000000, 0272989X. doi: 10.1177/02729890222062874. URL http://mdm.sagepub.com/cgi/doi/10.1177/02729890222062874.

[3] G. Carreras, S. Gallus, L. Iannucci, and G. Gorini. Estimating the probabilities of making a smoking quit attempt in Italy: stall in smoking cessation levels, 1986-2009. *BMC Public Health*, 12(1):183, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-183. URL http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-183.

[4] National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Reports of the Surgeon General. Centers for Disease Control and Prevention (US), Atlanta (GA), 2014. URL http://www.ncbi.nlm.nih.gov/books/NBK179276/.

[5] D.T. Levy, S. Gallus, K. Blackman, G. Carreras, C. La Vecchia, and G. Gorini. Italy SimSmoke: the effect of tobacco control policies on smoking prevalence and smoking-attributable deaths in Italy. *BMC Public Health*, 12(1):709, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-709. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-709.

[6] A.M. Near, K. Blackman, L.M. Currie, and D.T. Levy. Sweden SimSmoke: the effect of tobacco control policies on smoking and snus prevalence and attributable deaths.

The European Journal of Public Health, 24(3):451–458, 2014. ISSN 1101-1262, 1464-360X. doi: 10.1093/eurpub/ckt178. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckt178.

[7] D.T. Levy, L.M. Sánchez-Romero, Y. Li, Z. Yuan, N. Travis, M.J. Jarvis, J. Brown, and A. McNeill. England SimSmoke: the impact of nicotine vaping on smoking prevalence and smoking-attributable deaths in England. *Addiction*, 116(5):1196–1211, 2021. ISSN 0965-2140, 1360-0443. doi: 10.1111/add.15269. URL https://onlinelibrary.wiley.com/doi/10.1111/add.15269.

[8] L.M. Sánchez-Romero, L. Zavala-Arciniega, L.M. Reynales-Shigematsu, B.S. De Miera-Juárez, Z. Yuan, Y. Li, Y.K. Lau, N.L. Fleischer, R. Meza, J.F. Thrasher, and D.T. Levy. The Mexico SimSmoke tobacco control policy model: Development of a simulation model of daily and nondaily cigarette smoking. *PLOS ONE*, 16(6):e0248215, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0248215. URL https://dx.plos.org/10.1371/journal.pone.0248215.

[9] D.T. Levy, L. Nikolayev, E. Mumford, and C. 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*, 16(4):359–371, 2005. ISSN 0957-5243, 1573-7225. doi: 10.1007/s10552-004-7841-4. URL http://link.springer.com/10.1007/s10552-004-7841-4.

[10] D.T. Levy, J.E. Bauer, and H.R. Lee. Simulation Modeling and Tobacco Control: Creating More Robust Public Health Policies. *American Journal of Public Health*, 96(3):494–498, 2006. ISSN 0090-0036, 1541-0048. doi: 10.2105/AJPH.2005.063974. URL https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2005.063974.

[11] D.T. Levy, A.L. Graham, P.L. Mabry, D.B. Abrams, and C.T. Orleans. Modeling the Impact of Smoking-Cessation Treatment Policies on Quit Rates. *American Journal of Preventive Medicine*, 38(3):S364–S372, 2010. ISSN 07493797. doi: 10.1016/j.amepre.2009.11.016. URL https://linkinghub.elsevier.com/retrieve/pii/S0749379709008575.

[12] G. Carreras, G. Gorini, and E. 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*, 5(6):874–882, 2012. ISSN 1940-6207, 1940-6215. doi: 10.1158/1940-6207.CAPR-12-0019. URL https://aacrjournals.org/cancerpreventionresearch/article/5/6/874/49965/Can-a-National-Lung-Cancer-Screening-Program-in.

[13] G. Carreras, G. Gorini, S. Gallus, L. Iannucci, and D.T. Levy. Predicting the future prevalence of cigarette smoking in Italy over the next three decades. *European Journal of*

Public Health, 22(5):699–704, 2012. ISSN 1464-360X, 1101-1262. doi: 10.1093/eurpub/ckr108. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckr108.

[14] E.J. Feuer, D.T. Levy, and W.J. McCarthy. Chapter 1: The Impact of the Reduction in Tobacco Smoking on U.S. Lung Cancer Mortality, 1975-2000: An Introduction to the Problem: Introduction: Impact of the Reduction in Tobacco Smoking on U.S. Lung Cancer Mortality. *Risk Analysis*, 32:S6–S13, 2012. ISSN 02724332. doi: 10.1111/j.1539-6924.2011.01745.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.2011.01745.x.

[15] D.T. Levy, H. Fouad, J. Levy, A.D. Dragomir, and F. El Awa. Application of the *Abridged SimSmoke* model to four Eastern Mediterranean countries. *Tobacco Control*, 25(4):413–421, 2016. ISSN 0964-4563, 1468-3318. doi: 10.1136/tobaccocontrol-2015-052334. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tobaccocontrol-2015-052334.

[16] D.T. Levy, R. Borland, A.C. Villanti., R. Niaura, Z. Yuan, Y. Zhang, R. Meza, T.R. Holford, G.T. Fong, K.M. Cummings, and D.B. Abrams. The Application of a Decision-Theoretic Model to Estimate the Public Health Impact of Vaporized Nicotine Product Initiation in the United States. *Nicotine and Tobacco Research*, 19(2):149–159, 2017. ISSN 1462-2203, 1469-994X. doi: 10.1093/ntr/ntw158. URL https://academic.oup.com/ntr/article-lookup/doi/10.1093/ntr/ntw158.

[17] J. Tam, D.T. Levy, J. Jeon, J. Clarke, S. Gilkeson, T. Hall, E.J. Feuer, T.R. Holford, and R. Meza. Projecting the effects of tobacco control policies in the USA through microsimulation: a study protocol. *BMJ Open*, 8(3):e019169, 2018. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2017-019169. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2017-019169.

[18] M. Sanna, W. Gao, Y.W. Chiu, H.Y. Chiou, Y.H. Chen, C.P. Wen, and D.T. Levy. Tobacco control within and beyond WHO MPOWER: outcomes from Taiwan SimSmoke. *Tobacco Control*, 29(1):36–42, 2020. ISSN 0964-4563, 1468-3318. doi: 10.1136/tobaccocontrol-2018-054544. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tobaccocontrol-2018-054544.

[19] L.M. Sánchez-Romero, Z. Yuan, Y. Li, and D.T. Levy. The Kentucky SimSmoke Tobacco Control Policy Model of Smokeless Tobacco and Cigarette Use. *International Journal of Health Policy and Management*, page 1, 2020. ISSN 2322-5939. doi: 10.34172/ijhpm.2020.187. URL https://www.ijhpm.com/article_3936.html.

[20] D.T. Levy, J. Tam, L.M. Sanchez-Romero, Y. Li, Z. Yuan, J. Jeon, and R. Meza. Public health implications of vaping in the USA: the smoking and vaping simulation

model. *Population Health Metrics*, 19(1):19, 2021. ISSN 1478-7954. doi: 10.1186/s12963-021-00250-7. URL https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-021-00250-7.

[21] A. Lachi, C. Viscardi, M.C. Malevolti, G. Carreras, and M. Baccini. Compartmental models in epidemiology: Application on smoking habits in tuscany. In *Book of short papers SIS*, pages 1437–1442. Pearson, 2022.

[22] A. Lachi, C. Viscardi, G. Cereda, G. Carreras, and M. Baccini. A compartmental models for smoking dynamics in Italy: A pipeline for inference, validation, and forecasting under hypothetical scenarios. preprint, In Review, 2023. URL https://www.researchsquare.com/article/rs-3303111/v1.

[23] S. Lo Piano, R. Sheikholeslami, A. Puy, and A. Saltelli. Unpacking the modelling process via sensitivity auditing. *Futures*, 144:103041, 2022. ISSN 00163287. doi: 10.1016/j.futures.2022.103041. URL https://linkinghub.elsevier.com/retrieve/pii/S0016328722001410.

[24] L.D. Broemeling. *Bayesian analysis of infectious diseases: COVID-19 and beyond*. Chapman and Hall/CRC biostatistics series. CRC Press, Taylor and Francis Group, Boca Raton London New York, 2021. ISBN 978-0-367-64724-7 978-0-367-63386-8.

[25] B. Efron and R. Tibshirani. *An introduction to the bootstrap*. Number 57 in Monographs on statistics and applied probability. Chapman and Hall, New York, 1993. ISBN 978-0-412-04231-7.

[26] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042717300234.

[27] T.J. McKinley, I. Vernon, I. Andrianakis, N. McCreesh, J.E. Oakley, R.N. Nsubuga, M. Goldstein, and R.G. White. Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models. *Statistical Science*, 33(1):4 – 18, 2018. doi: 10.1214/17-STS618.

[28] L. Tang, Y. Zhou, L. Wang, S. Purkayastha, L. Zhang, J. He, F. Wang, and P.X.K. Song. A review of multi-compartment infectious disease models. *International Statistical Review*, 88(2):462–513, 2020. doi: 10.1111/insr.12402.

[29] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-05997-5.

[30] I.M. Sobol. *A primer for the Monte Carlo method*. CRC Press, Boca Raton, 1994. ISBN 978-0-8493-8673-2.

[31] S. Steegen, F. Tuerlinckx, A. Gelman, and W. Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5):702–712, 2016.

[32] S. Lo Piano and L. Benini. A critical perspective on uncertainty appraisal and sensitivity analysis in life cycle assessment. *Journal of Industrial Ecology*, 26(3):763–781, June 2022. ISSN 1088-1980, 1530-9290. doi: 10.1111/jiec.13237. URL `https://onlinelibrary.wiley.com/doi/10.1111/jiec.13237`.

[33] S. Tarantola, F. Ferretti, S. Lo Piano, M. Kozlova, A. Lachi, R. Rosati, A. Puy, P. Roy, G. Vannucci, M. Kuc-Czarnecka, and A. Saltelli. An annotated timeline of sensitivity analysis. *Environmental Modelling & Software*, page 105977, 2024. doi: https://doi.org/10.1016/j.envsoft.2024.105977.

[34] G. Carreras, A. Lachi, B. Cortini, S. Gallus, M.J. Lopez, A.N. Lopez, J.B. Soriano, E. Fernandez, O. Tigova, G. Gorini, and TackSHS Project Investigators. Burden of disease from second-hand tobacco smoke exposure at home among adults from european union countries in 2017: an analysis using a review of recent meta-analyses. *Preventive medicine*, 145: 106412, 2021. doi: https://doi.org/10.1016/j.ypmed.2020.106412.

[35] G. Carreras, A. Lachi, B. Cortini, S. Gallus, M.J. Lopez, A.N. Lopez, A. Lugo, M.T. Pastor, J.B. Soriano, E. Fernandez, and TackSHS Project Investigators. Burden of disease from exposure to secondhand smoke in children in europe. *Pediatric research*, 90(1):216–222, 2021. doi: https://doi.org/10.1038/s41390-020-01223-6.

[36] M.C. Malevolti, A. Lugo, M. Scala, S. Gallus, G. Gorini, A. Lachi, and G. Carreras. Dose-risk relationships between cigarette smoking and cervical cancer: a systematic review and meta-analysis. *European Journal of Cancer Prevention*, 32(2):171–183, 2023. doi: https://doi.org/10.1097/cej.0000000000000773.

[37] M. Marra, M. Baccini, G. Cereda, M. Culasso, M. De Sario, I. Eboli, A. Lachi, Z. Mitrova, R. Saulle, and A. Bena. Strategie di screening per contenere la diffusione del covid-19 nella scuola: una revisione sistematica della letteratura. *Epidemiologia e Prevenzione*, 47(3): 152–171, 2023. doi: https://doi.org/10.19191/EP23.3.A576.054.

CHAPTER 1

A compartmental model for smoking dynamics in Italy: A pipeline for inference, validation, and forecasting under hypothetical scenarios

Paper submitted to *BMC Medical Research Methodology*, under first review

Alessio Lachi[1,2], Cecilia Viscardi[1,3], Giulia Cereda[1,3], Giulia Carreras[4], Michela Baccini[1,3]

[1] Department of Statistics, Computer Science, and Applications “Giuseppe Parenti" (DiSIA), University of Florence, Viale Giovan Battista Morgagni 59-65 50134 - Florence, Italy

[2] Epidemiology and Health Research, Institute of Clinical Physiology of the Italian National Institute Research Council (IFC-CNR), Via Giuseppe Moruzzi 1 56124 - Pisa, Italy

[3] Florence Center for Data Science, University of Florence, Viale Giovanni Battista Morgagni 59-65 50134 - Florence, Italy

[4] Oncologic Network, Prevention and Research Institute (ISPRO), Health Service of Tuscany, Via Cosimo il Vecchio 2 50139 - Florence, Italy

Abstract

We propose a compartmental model for investigating smoking dynamics in an Italian region. Calibrating the model on local data from 1993 to 2019, we estimate the probabilities of starting and quitting smoking and the probability of smoking relapse. Then, we forecast the evolution of smoking prevalence until 2043 and assess the impact on mortality in terms of attributable deaths. We introduce elements of novelty with respect to previous studies in this field, including a formal definition of the equations governing the model dynamics and a flexible modelling of smoking probabilities based on cubic regression splines. We estimate model parameters by defining a two-step procedure and quantify the sampling variability via a parametric bootstrap. We propose the implementation of cross-validation on a rolling basis and variance-based Global Sensitivity

Analysis to check the robustness of the results and support our findings. Our results suggest a decrease in smoking prevalence among males and stability among females, over the next two decades. We estimate that, in 2023, 18% of deaths among males and 8% among females are due to smoking. We test the use of the model in assessing the impact on smoking prevalence and mortality of different tobacco control policies, including the tobacco-free generation ban recently introduced in New Zealand.

Keywords - Compartmental Models, Smoking Dynamics, Tobacco Control Policies, Global Sensitivity Analysis, Parametric Bootstrap, Cross Validation, Smoking Attributable Deaths, Forecasting, Calibration, Regression Splines

1.1 Background

Smoking is a significant risk factor for many common chronic diseases, including cancer, cardiovascular, cerebrovascular and respiratory diseases, diabetes, and a leading preventable cause of premature death [1, 2]. Also, smoking reduces length and quality of life [3], and contributes to health inequities [4]. The Global Burden of Disease study [5] reports that in 2019 smoking was responsible for around 8,709,000 deaths in the World (15.4% of all deaths), 907,000 in Europe, and 96,000 in Italy.

The importance of Tobacco Control Policies (TCP) has been firmly established within the World Health Organization's (WHO) Framework Convention on Tobacco Control (FCTC), an international treaty that came into force in 2005 and has been ratified by 182 countries. Specifically, tobacco control has been included as one of the global development goals, recognized as crucial and necessary to achieve a one-third reduction in premature mortality by 2030 [6].

Focusing on Italy, data from the Italian surveillance system PASSI (Progressi delle Aziende Sanitarie per la Salute in Italia) highlighted that in 2021 23.7% of Italians (27.2% in men and 20.2% in women) described themselves as current smokers [7]. Among adolescents, smoking prevalence stalled in the last years, with a prevalence of current smokers between 27.3% and 32.4% in young people aged 13-16 years [8, 9].

Dynamic simulation models are widely used to describe and project the evolution of smoking habits in the population over time and to estimate the impact of past and hypothetical future TCPs. Since the 2000s, several models have been proposed [10–13], some of which developed within the Cancer Intervention and Surveillance Modelling Network (CISNET), a consortium of investigators funded by the National Cancer Institute, that uses mathematical modelling to study the impact of cancer control interventions [14]. These models are mainly of two types: compartmental models and agent-based models. Compartmental models, starting from a baseline year, perform macro-simulations so that the population evolves through deaths, births and changes

in smoking habits [10–12]. The SimSmoke model [15] is the most used compartmental model [16], implemented for a wide number of countries including Italy [17–20], among others. Agent-based models, also called micro-simulation models, simulate individual life trajectories and interactions to assess their effects on the system as a whole [21, 22].

In this paper, grounding on previous works [12, 23–25], we developed a compartmental model that describes the evolution of smoking habits in Tuscany, a region of Central Italy, from 1993 to 2019, and forecasts them until 2043. The model assumes that at each point in time, the population is divided into non-overlapping groups called compartments, defined according to smoking status (never, current, and former smokers), age and sex [26]. Transitions between compartments are described by simple probabilistic rules and the evolution of the size of the compartments is governed by a system of differential equations.

While some of the transition parameters in the model were assumed as fixed, we estimated via a two-step calibration the age-specific probabilities of starting and quitting smoking, modelled in a flexible way through cubic regression splines [27], the probability of relapsing smoking, modelled as a nonlinear function of the time from quitting [28], and the mortality rate. We calibrated the model on the observed prevalence of never, current, and former smokers for the years from 1993 to 2019, arising from yearly local surveys.

Once we estimated the transition parameters, we predicted the prevalence of never, current, and former smokers in the regional population over time, and we quantified the impact of smoking in terms of the number of smoking-attributable deaths (SAD) and population attributable fraction (PAF). With simple examples, we also illustrated the use of the compartmental model to predict the future impact of hypothetical interventions that act on the probabilities of starting and quitting smoking.

Compared to previous studies that dealt with the same problem, we aimed to prove some methodological advances both in the modelling and estimation strategies. First of all, grounding on a formal definition of the model equations, we addressed the problem of accounting for the sampling variability and provided confidence intervals for the estimates of the parameters and compartment sizes. To this end, due to the unavailability of the likelihood function associated with the model, we relied on a parametric bootstrap procedure [29, 30]. Also, we introduced a flexible modelling of the probabilities of starting and stopping smoking, usually assumed as constant. We allowed them to change over time as a function of age. Moreover, we assessed the predictive performance of the model using cross-validation on a rolling basis. Finally, we assessed parameter identifiability through Global Sensitivity Analysis (GSA) [31].

1.2 Methods

1.2.1 Data

The analyses relied on data from heterogeneous sources. We used data from the National Institute of Statistics (ISTAT) Multipurpose Surveys “Aspect of Daily Life" (AVQ) (`www.istat.it/it/archivio/91926`), which every year collects fundamental information related to the daily life of individuals and families in Italy. Each yearly AVQ survey enrols about 25,000 families distributed in about 800 Italian municipalities of different population sizes. Specifically, we obtained from the ISTAT AVQ surveys an estimate of the distribution by smoking habit (never, current, and former smokers) of the population residing in Tuscany for each year from 1993 to 2019, separately for males and females and by age class (14-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75+). We obtained from the same surveys the smoking intensity distribution for current smokers, by sex and age class.

We used data from the ISTATsurvey Multipurpose Surveys European Health Survey (EHIS) (`www.istat.it/it/archivio/167485`), a survey on the main aspects of public health carried out every 5 years from 1980 to all member states of the European Union, to obtain an estimate of the smoking intensity distribution among former smokers, as well as information about time since smoking cessation, separately for males and females and by the same age classes reported above. In particular, we considered the surveys for 1994, 1999, 2004, and 2013.

We obtained the size of the Tuscany population on January 1st 1993 and January 1st 2005, by age and sex, from the ISTAT website (`www.istat.it`). From the same website, we got the mortality rates by age and sex and the number of new births in Tuscany for the period 1993-2019. The relative risks (RR) of death for smokers and ex-smokers versus never smokers are those reported in Thun et al. [32].

1.2.2 Model specification

We specified a compartmental model for smoking habit dynamics in the population, which we call the Smoking Habits Compartmental (SHC) model. In order to better present the SHC model adopted for the analysis, we first introduce a simpler version of it, and then proceed step by step, adding elements of complexity.

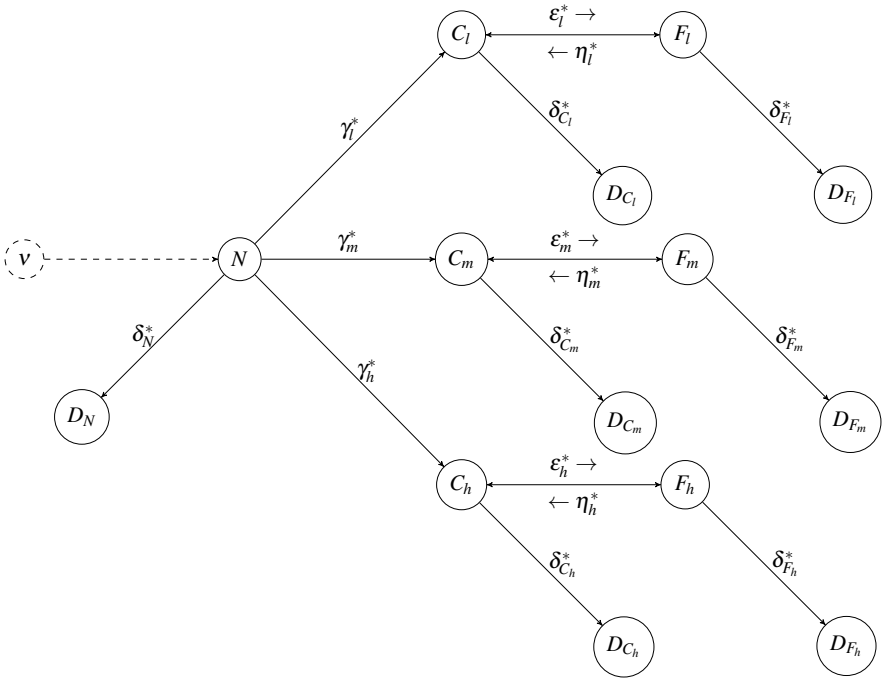


FIGURE 1.1: Smoking Habits Compartmental model in its simplest form.

The starting model assumes that at each time the alive population is divided into the following non-overlapping compartments: never (N), current (C), and former (F) smokers. We consider only cigarette smoking; current smokers are the individuals who have smoked at least 100 cigarettes in their life and smoked in the last 30 days, and former smokers are those who stopped smoking for at least 6 months [33]. Never smokers can become current smokers, current smokers can become former smokers, and former smokers may restart smoking (smoking relapse). The compartments C and F are further divided into sub-compartments denoted by C_i and F_i, where $i \in \{l,m,h\}$ indicates the level of smoking intensity, corresponding to low ($<$10 cigarettes/day), medium ($\geq$10 and $<$20 cigarettes/day), and high ($\geq$20 cigarettes/day) smoking intensity, respectively. During their life, individuals can change their smoking status, but, for the sake of simplicity, we assume that they cannot change their level of smoking intensity. The model admits deaths and new births. From each compartment, subjects can transit to a deceased compartment denoted by the letter D and a subscript corresponding to the compartment of origin. New births ($\nu(t)$ is the number of new births at time t) increase the size of the compartment N. Transitions of the individuals from a given compartment to another one determine flows regulated by the transition parameters, among which the rates of starting smoking (γ_i^*), stopping smoking (ε_i^*), and relapsing into smoking after having stopped (η_i^*). Note that these rates can depend on the level of smoking intensity i. Death happens with different rates for never (δ_N^*), current ($\delta_{C_i}^*$), and former ($\delta_{F_i}^*$) smokers. For current and former smokers, the mortality rates may depend also on smoking intensity. This compartmental model, graphically represented in Figure 1.1, is expressed by the following system of differential equations for each $i \in \{l,m,h\}$:

$$\begin{cases} \frac{dN(t)}{dt} = -N(t)\left(1-\delta_N^*\right)\gamma^* - N(t)\delta_N^* + \nu(t)\left(1-\delta_N^*\right) \\ \frac{dC_i(t)}{dt} = -C_i(t)\left(1-\delta_{C_i}^*\right)\varepsilon_i^* - C_i(t)\delta_{C_i}^* + N(t)\left(1-\delta_N^*\right)\gamma_i^* + F_i(t)\left(1-\delta_{F_i}^*\right)\eta_i^* \\ \frac{dF_i(t)}{dt} = -F_i(t)\left(1-\delta_{F_i}^*\right)\eta_i^* - F_i(t)\delta_{F_i}^* + C_i(t)\left(1-\delta_{C_i}^*\right)\varepsilon_i^* \\ \frac{dD_N(t)}{dt} = N(t)\delta_N^* + \nu(t)\delta_N^* \\ \frac{dD_{C_i}(t)}{dt} = C_i(t)\delta_{C_i}^* \\ \frac{dD_{F_i}(t)}{dt} = F_i(t)\delta_{F_i}^*, \end{cases} \tag{1.1}$$

where $\gamma^* = \gamma_l^* + \gamma_m^* + \gamma_h^*$ is the overall transition rate from the status of never smoker to the status of a current smoker. The initial conditions of the system, i.e. the sizes of the compartments at time 0, set to the 1^{st} of January 1993, are $N(0) = n_0$, $D_N(0) = 0$, $F_i(0) = f_0^i$, $C_i(0) = c_0^i$ and $D_{C_i}(0) = D_{F_i}(0) = 0 \; \forall i \in \{l,m,h\}$, where n_0 is the number of never smokers in the considered population on the first day of the study period, and f_0^i and c_0^i are the number of ex-smokers and current smokers with smoking intensity i.

For computational reasons, it is convenient to discretise the system of differential equations in Equation (1.1), assuming that the size of the compartments is constant during 1-year time steps. Hereafter, t will denote discrete time, with the year as a time-unit ($t \in \{1,...,T\}$), and we replace the system in Equation (1.1) with a system of difference equations, where the annual probability of stopping smoking (ε_i), and the annual probabilities of smoking relapse (η_i) are derived from the corresponding rates in Equation (1.2), as well as the annual probabilities of death for never (δ_N), current (δ_{C_i}), and former (δ_{F_i}) smokers. In particular, $\delta_N = 1 - \exp(-\delta_N^*)$, and $\varepsilon_i = 1 - \exp(-\varepsilon_i^*)$, $\eta_i = 1 - \exp(-\eta_i^*)$, $\delta_{C_i} = 1 - \exp(-\delta_{C_i}^*)$, $\delta_{F_i} = 1 - \exp(-\delta_{F_i}^*)$ with $i \in \{l,m,h\}$. Regarding the probabilities of starting smoking for never smokers, the overall annual probability γ comes from the corresponding rate, $\gamma = 1 - \exp(-\gamma^*)$, while $\gamma_i = \pi_{C_i}\gamma$, where $\boldsymbol{\pi} = (\pi_{C_l}, \pi_{C_m}, \pi_{C_h})$ is the distribution of the level of smoking intensity among the new current smokers. Notice that, if λ is the rate of occurrence of an event, the probability of experiencing at least one event in the time unit is $1 - \exp(-\lambda)$. The resulting system of discretised equations for each $i \in \{l,m,h\}$ and $t \in \{1,...,T\}$ is:

$$\begin{cases} N(t) = N(t-1)\Big(1-\delta_N\Big)\Big(1-\gamma\Big) + \nu(t-1)\Big(1-\delta_N\Big) \\ C_i(t) = C_i(t-1)\Big(1-\delta_{C_i}\Big)\Big(1-\varepsilon_i\Big) + N(t-1)\Big(1-\delta_N\Big)\gamma_i + F_i(t-1)\Big(1-\delta_{F_i}\Big)\eta_i \\ F_i(t) = F_i(t-1)\Big(1-\delta_{F_i}\Big)\Big(1-\eta_i\Big) + C_i(t-1)\Big(1-\delta_{C_i}\Big)\varepsilon_i \\ D_N(t) = D_N(t-1) + N(t-1)\delta_N + \nu(t-1)\delta_N \\ D_{C_i}(t) = D_{C_i}(t-1) + C_i(t-1)\delta_{C_i} \\ D_{F_i}(t) = D_{F_i}(t-1) + F_i(t-1)\delta_{F_i}, \end{cases} \tag{1.2}$$

where $\nu(t-1)$ denotes the newborns in the year $t-1$. Note that the initial conditions of the system of equations in (1.2) coincide with those of the previous model in (1.1).

The SHC model extends the system in Equation (1.2) to account for two additional discrete time axes: age and time since smoking cessation. The final model is a compartmental model with separate compartments for each discrete age (a), where also a stratification by years since smoking cessation (c) is introduced for former smokers. Two separate SHC models are defined by sex. The final SHC model is defined by the following system of equations for each $i \in \{l,m,h\}$ and $t \in \{1,...,T\}$:

$$
\begin{cases}
N(t;a) = \nu(t-1)\Big(1-\delta_N(a)\Big) & \text{if } a=0 \\
N(t;a) = N(t-1;a-1)\Big(1-\delta_N(a)\Big)\Big(1-\gamma(a)\Big) & \text{if } a>0 \\
C_i(t;a) = 0 & \text{if } a=0 \\
C_i(t;a) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\Big(1-\varepsilon(a)\Big)+ & \\
\qquad N(t-1;a-1)\Big(1-\delta_N(a)\Big)\gamma_i(a)+ & \\
\qquad \sum_{c>0} F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\eta(c) & \text{if } a>0 \\
F_i(t;a,c) = 0 & \text{if } a=0,\ c\geq 0 \\
F_i(t;a,c) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\varepsilon(a) & \text{if } a>0,\ c=0 \\
F_i(t;a,c) = F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\Big(1-\eta(c)\Big) & \text{if } a>0,\ c>0 \\
D_N(t;a) = \nu(t-1)\delta_N(a) & \text{if } a=0 \\
D_N(t;a) = D_N(t-1;a)+N(t-1;a-1)\delta_N(a) & \text{if } a>0 \\
D_{C_i}(t;a) = 0 & \text{if } a=0 \\
D_{C_i}(t;a) = D_{C_i}(t-1;a)+C_i(t-1;a-1)\delta_{C_i}(a) & \text{if } a>0 \\
D_{F_i}(t;a,c) = 0 & \text{if } a=0,\ c\geq 0 \\
D_{F_i}(t;a,c) = 0 & \text{if } a>0,\ c=0 \\
D_{F_i}(t;a,c) = D_{F_i}(t-1;a,c)+F_i(t-1;a-1,c-1)\delta_{F_i}(a,c) & \text{if } a>0,\ c>0.
\end{cases}
\tag{1.3}
$$

The initial conditions of the system are obtained by generalizing those of the model in Equation (1.2), to take into account the stratification by age for current smokers, and the stratification by age and time since cessation for former smokers.

The age a takes values from 0 to 100. We set $\gamma(a)$ to 0 until 13 and from 35 years of age, and, in order to account for the possible non-linearity between 14 and 34, we modeled the *logit* transformation of $\gamma(a)$ through a natural cubic regression spline of age, with 2 equidistant internal knots. The choice of the number of nodes is motivated by the need to allow flexibility containing at the same time model complexity and computational times. Similarly, we set $\varepsilon(a)$ to 0 until 19 years of age; we introduced a natural cubic regression spline with 2 equidistant internal knots to model non-linearity for $a \geq 20$. The resulting functions are the following:

$$
\gamma(a) = \begin{cases} 0 & 0 \leq a \leq 13 \cup a \geq 35 \\ \frac{\exp(s(a;\boldsymbol{\psi}))}{1+\exp(s(a;\boldsymbol{\psi}))} & 14 \leq a \leq 34 \end{cases}
\qquad
\varepsilon(a) = \begin{cases} 0 & 0 \leq a \leq 19 \\ \frac{\exp(s(a;\boldsymbol{\phi}))}{1+\exp(s(a;\boldsymbol{\phi}))} & a \geq 20, \end{cases}
$$

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 probabilities of starting and quitting smoking, respectively. The relapsing rate, $\eta^*(c)$,

was modeled as a negative exponential function of the time since cessation, with parameters $\boldsymbol{\omega} = (\omega_0, \omega_1)$:

$$\eta^*(c) = \begin{cases} 0 & c = 0 \\ \omega_0 \omega_1 \exp(-\omega_1 c) & 1 \leq c \leq 15 \\ \omega_0 \omega_1 \exp(-\omega_1 15) & c \geq 16, \end{cases}$$

where ω_0 governs the lifetime probability of no relapse and ω_1 tunes how fast the rate of smoking relapse declines with the time from cessation [12, 23, 28, 34]. Both ω_0 and ω_1 are assumed to be positive so that $\eta^*(c)$ is a positive, decreasing function of c. The assumptions on which the SHC model is based are summarized in Section A.1, Supplemental Material.

1.2.3 Estimation strategy

An important issue in compartmental models concerns parameter identifiability [30]. Complex models with many compartments, such as the model in Equation (1.3), have many parameters governing the admitted transitions, but unfortunately observed data are often insufficient to estimate all of them. To overcome this problem we fixed some of the parameters to values from the literature or external data, leaving as unknown the mortality risks and the spline coefficients $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, and $\boldsymbol{\omega}$. Details on the values assigned to the fixed parameters are provided in Section A.2, Supplemental Material. The unknown parameters have been estimated following the two step-procedure described in the next section.

1.2.3.1 Two-step estimation

In order to estimate the unknown parameters, we adopted a two-step procedure. Both steps use as observed data the prevalence of never, current and former smokers from ISTAT AVQ, here denoted by $p^{obs}(t;a^*) = \left(p_C^{obs}(t;a^*), p_N^{obs}(t;a^*), p_F^{obs}(t;a^*)\right)$, where t denotes the year and a^* the age class. In particular, we considered years from 1993 to 2019 and age classes $a^* \in \{14-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75+\}$.

First step. We estimated the age-specific risks of mortality for never smokers $\delta_N(a)$ using both observed prevalence and relative risks coming from the literature. In particular, the age-specific risks of dying for current and former smokers in the population at time t are respectively $\delta_C(t;a) = RR_C \times \delta_N(t;a)$ and $\delta_F(t;a) = RR_F \times \delta_N(t;a)$, with RR_C and RR_F the relative risks of dying for current smokers and former smokers versus never smokers. Let $p(t;a) = \left(p_N(t;a), p_C(t;a), p_F(t;a)\right)$ be the distribution of never, current and former smokers in

the population. The overall mortality at age a in the year t, $\delta_{pop}(t;a)$, is a weighted average of $\delta_N(t;a)$, $\delta_C(t;a)$, and $\delta_F(t;a)$ with weights $p(t;a)$. Thus, $\delta_N(t;a)$ can be derived as the ratio:

$$\delta_N(t;a) = \frac{\delta_{pop}(t;a)}{p_N(t;a) + RR_C p_C(t;a) + RR_F p_F(t;a)}. \tag{1.4}$$

Therefore, separately for each year t in the period 1993-2019, we obtained an estimate of $\delta_N(t;a)$, plugging into equation (1.4) the mortality risk at age a reported for Tuscany, the relative risks for current and former smokers versus never smokers [32], and the observed age-specific prevalence of never, current and former smokers $p^{obs}(t;a^*)$. Finally, we averaged the year-specific $\hat{\delta}_N(t;a)$ over t, obtaining the overall estimate $\hat{\delta}_N(a)$. The risks of dying for current and former smokers by i and c were then derived as:

$$\hat{\delta}_{C_i}(a) = RR_{C_i} \times \hat{\delta}_N(a) \qquad \hat{\delta}_{F_i}(a,c) = RR_{F_i}(c) \times \hat{\delta}_N(a).$$

Second step. After fixing the mortality risks to the values computed at the first step, $\hat{\boldsymbol{\delta}}(a,c)$, we calibrated the model on the observed prevalence $p^{obs}(t;a^*)$ to estimate the vector of parameters which were still unknown, $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\omega})$. $p(t;a^*,\boldsymbol{\theta}) = \big(p_C(t;a^*,\boldsymbol{\theta}), p_N(t;a^*,\boldsymbol{\theta}), p_F(t;a^*,\boldsymbol{\theta})\big)$ be the vector of the prevalence of never, current and former smokers belonging to the class of age a^* at time t, calculated on the population predicted by the model (1.3), given a specific value of $\boldsymbol{\theta}$. With calibration, we searched for the value of $\boldsymbol{\theta}$ that leads to predicted prevalence as close as possible to the observed ones. To compare observed and simulated trajectories, we considered the following objective function, where $H(\cdot,\cdot)$ denotes the Hellinger distance [35] between two discrete probability distributions:

$$\begin{aligned} Obj(\boldsymbol{\theta}) =& \frac{1}{T \times A^*} \sum_{t,a^*} H\left(p(t;a^*,\boldsymbol{\theta}), p^{obs}(t;a^*) \right) = \\ =& \frac{1}{T \times A^* \times \sqrt{2}} \sum_{t,a^*} \sqrt{\sum_{k \in \{C,N,F\}} \left(\sqrt{p_k(t;a^*,\boldsymbol{\theta})} - \sqrt{p_k^{obs}(t;a^*)} \right)^2}, \end{aligned} \tag{1.5}$$

where A^* is the number of age classes a^*. We minimized the objective function in Equation (1.5) over $\boldsymbol{\theta}$ via a global optimization procedure, resorting to the JULIA package `Optim.jl` [36]. It is well-known that, in the context of compartmental models, optimization results often depend on the chosen starting points of the algorithm [37, 38]. To avoid the problem of getting stuck in local minima, we performed several optimizations using different starting points, then we selected the solution that brought to the minimum Hellinger distance [30, 37]. The two-step procedure was

performed separately by sex, obtaining different estimates for males and females and sex-specific evolution of the compartment sizes.

We estimated the compartment sizes up to 2043 by projecting the model dynamics assuming that parameters and model structure do not change after 2019.

1.2.3.2 Parametric bootstrap procedure

We quantified the sampling variability around point estimates and projections by using a parametric bootstrap procedure [29, 30]. Let $\hat{\boldsymbol{\theta}}$ be the vector of parameters minimizing the objective function in Equation (1.5) and $p(t;a^*,\hat{\boldsymbol{\theta}})$ the corresponding estimated vector of prevalence for never, current and former smokers of age a^* in the population at time t. Let $n(t;a^*)$ be the number of subjects belonging to the age class a^*, enrolled in the ISTAT AVQ in the year t in Tuscany (i.e. the denominator of the observed prevalence $p^{obs}(t;a^*)$). The bootstrap procedure consisted of the following steps:

1. for each a^* and t, we sampled a vector of prevalence from a Dirichlet distribution: $p^b(t;a^*) \sim Dirichlet\Big(p_C(t;a^*,\hat{\boldsymbol{\theta}})n(t;a^*), p_N(t;a^*,\hat{\boldsymbol{\theta}})n(t;a^*), p_F(t;a^*,\hat{\boldsymbol{\theta}})n(t;a^*)\Big)$;
2. we considered the collection of these sampled vectors as the observed values and performed the two-step estimation, computing the vector $\boldsymbol{\delta}^b(a,c)$ and finding $\boldsymbol{\theta}^b$ that minimized the objective function;
3. we repeated the previous two steps $B = 1000$ times, collecting a sample of B bootstrap estimates of $\boldsymbol{\delta}(a,c)$ and $\boldsymbol{\theta}$ to be used to estimate as many curves describing the transition parameters and compartment size trajectories;
4. we calculated the 90% confidence intervals for the quantities of interest as the 5^{th} and 95^{th} percentiles of the bootstrap estimates; pointwise confidence intervals were calculated for the curves.

1.2.4 Model validation

In order to evaluate the predictive performance of the estimation procedure described in Section Estimation strategy, we applied cross-validation (CV) on a rolling basis. We started defining the first 3 years of the period 1993-2019 as the training set, and the subsequent q years as the test set. Then, we calibrated the compartmental model in Equation (1.3) on the training set and used the estimated model to forecast the prevalence of never, current, and former smokers in the years belonging to the test time window. The discrepancy between observed and projected prevalence was evaluated in terms of absolute percentage error. Then, we progressively extended

the length of the training set by adding one year at a time, and we obtained the projections for the q subsequent years every time. We stopped when the last training set considered the years between 1993 to 2016. We finally computed the Mean Absolute Percentage Error (MAPE), by averaging the absolute percentage errors across different types of smokers over time, age classes, and training sets. Note that in general, for a set of n observations, MAPE is defined as $\frac{100}{n}\sum_{i=1}^{n}\frac{|O_i-E_i|}{O_i}$, where O_i is the observed value and E_i is the expected one for unit i. We calculated the MAPE for different forecasting horizons by setting $q = 3,6,9,12$ years.

1.2.5 Sensitivity analysis

A key assumption of our model is that the dynamics of the studied phenomenon, particularly the transition probabilities between compartments, remain constant from 1993 to 2019 (and continue to do so until 2043). To verify its appropriateness, we conducted two separate analyses, first calibrating the model on the period 1993-2004 and then on the period 2005-2019, and compared the results. Notice that in the analysis 2005-2019 the initial sizes of the compartments were set to values obtained from 2005 surveys (see Section A.3, Supplemental Material).

Another crucial point concerns the fact that the inference results could be affected by the model parameters assumed as fixed. To address this issue, we utilized a variance-based approach to Global Sensitivity Analysis (GSA) [31]. Given K_X mutually independent inputs $(X_1, X_2, ..., X_{K_X})$ and a model which, given the inputs, returns K_Y outputs $(Y_1, Y_2, ..., Y_{K_Y})$, this approach quantifies the relative importance of each input to the model's outcomes by propagating uncertainty from the inputs to the outputs and computing variance indices. In our application, given the model in Equation 1.3, we considered as inputs all the parameters, both fixed and unknown, and the Hellinger distance in Equation (1.5) as the output Y. Note that, considering the Hellinger distance as the output, we directly measure the influence of the inputs on the discrepancy between observed and predicted data, thus, ultimately, on the inference results. Then we calculated, for each input X_i, the so-called total variance index (S_i^{tot}), which S_i^{tot} measures the overall effect of the i-th input on the output Y, including all the interactions of X_i with the other inputs. This index corresponds to the expected variance of Y that would be left on average when all the parameters but X_i, $X_{\sim i}$, are fixed:

$$S_i^{tot} = \frac{E_{X\sim i}(Var_{X_i}(Y|X_{\sim i}))}{Var(Y)}.$$

A total variance index close to zero indicates that the parameter X_i does not influence Y, and therefore, the inference results. Conversely, a large total variance index indicates that the parameter does have an impact on them. In the former case, the parameter can be fixed without affecting our estimates, or in other words, our model and data do not provide information on

this parameter. The computation of S_i^{tot} relies on Monte Carlo simulations [39]. We simulated $K = 10,000$ different combinations of the model inputs, then, for each of them, we predicted the prevalence values via the model in Equation (1.3) and calculated the corresponding Hellinger distance. Specifically, we draw the model parameters from the distributions reported in Table A.3, Supplemental Material, adopting a quasi-random numbers sampling which provides a more efficient exploration of the sample space [40, 41]. On the basis of the simulated Hellinger distance and the combination of the parameters, we computed the total variance indices as described in [39]. It is worth noting that in the GSA we did not include the age-specific mortality rates, $\delta_{pop}(t;a)$, among the model inputs. It is reasonable, as done elsewhere [23], to treat these parameters as not affected by uncertainty, given that they were estimated based on the entire population.

1.2.6 Health impact assessment

The impact of smoking was quantified in terms of attributable deaths. We calculated the Smoking-Attributable Deaths (SADs) in the year t as the difference between the number of deaths occurring in that year under the actual scenario, i.e. the number of deaths predicted by the model in Equation (1.3) given $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\delta}}(a)$, and the deaths we would observe under a specific counterfactual condition. We considered the counterfactual condition where current smokers and former smokers in the year t were never smokers. Therefore, for each age a, we applied to the size of the compartments of smokers or ex-smokers the excess risk relative to never-smokers. The excess risk is defined as the difference between risks. For example, the excess risk of current smokers of age a and smoking intensity i relative to never-smokers is $\delta_{C_i}(a,c) - \delta_N(a)$. Thus, for the year t, the number of SADs among people of age a was calculated as:

$$\text{SAD}(t;a) = \sum_i C_i(t,a;\hat{\boldsymbol{\theta}})(\hat{\delta}_{C_i}(a) - \hat{\delta}_N(a)) + \sum_i \sum_c F_i(t,a,c;\hat{\boldsymbol{\theta}})(\hat{\delta}_{F_i}(a,c) - \hat{\delta}_N(a)).$$

The age-specific $\text{SAD}(t;a)$ can be summed over a to obtain the total number of attributable deaths in population or in a certain class of age: $\text{SAD}(t) = \sum_a \text{SAD}(t;a)$. The impact of smoking on population health can be expressed also in terms of Population Attributable Fraction (PAF), defined as the proportion of deaths that would be avoided if all current and former smokers in the population or in a subset of it were never smokers [42]. For details, see Section A.4, Supplemental Material. We calculated SADs and PAFs over the period 1993-2043, separately by sex and for the ages 35+ and 65+.

1.2.6.1 Impact of future hypothetical policies

In order to illustrate the use of the compartmental model to assess the impact of hypothetical TCPs on SAD, we focused on three policies acting on the rates of starting and stopping smoking, $\gamma^*(a)$ and $\varepsilon^*(a)$. We assumed that all the defined policies are implemented in 2023 and that, in the absence of policies, the smoking habit dynamics would not change.

Taking inspiration from [43], and from a recent policy introduced in New Zealand (`www.bbc.com/news/world-asia-63954862`) we defined the following hypothetical TCPs starting from 2023:

- TCP1, a policy able to reduce the rate of starting smoking by 25% in 10 years for subjects between 14 and 34 years of age; for simplicity, we assumed a linear decrease, starting with a decrease of 2.5% the first year, a decrease of 5% the second one and so on, up to a final decrease of 25% after 10 years;
- TCP2, a policy able to increase the rate of stopping smoking by 25% in 10 years for subjects between 25 and 100 years of age; for simplicity, we assumed a linear growth, starting with an increase of 2.5% the first year, an increase of 5% the second one and so on, up to a final increase of 25% after 10 years;
- TCP3, a policy that imposes a complete smoking ban on cohorts born since 2009.

For each policy, we calculated the evolution of smoking prevalence and the number of avoided deaths expected from its implementation, taking the scenario without policies as a reference (TCP0). To better appreciate the impact of policies in terms of SAD, limited to this analysis, we extended the projections up to 2063.

1.3 Results

The Tuscany population in 1993 counted 1,697,495 million males and 1,824,090 million females, and the proportions of never, current, and former smokers estimated from the ISTAT AVQ survey were respectively 35%, 34%, 31% for males and 67%, 20%, 13% for females.

Figure (1.2), Panel (a) and (b) show, separately for males and females, the estimates of the parameters left unknown in the SHC model in Equation (1.3), with their 90% confidence intervals (CI), as obtained from the two-step estimation procedure and bootstrap. In particular, Panel (b) compares the estimated risk of death for never smokers with the one in the general population.

It is worth noting that while the two risks are similar for females (the mortality among never-smokers is 8% lower than among the general population), a not negligible difference is observed for males (25% lower) as noted also in [44].

Figure (1.2), Panel (c) shows the estimates of the probabilities of starting and quitting smoking and the probability of smoking relapse, derived from the estimated coefficients in Panel (a). Table 1.1 reports some summaries of the curves. Males are more likely to start and quit smoking than females. In particular, the probability of starting smoking has a peak around 20.0 years of age for males and 19.2 for females, with a maximum of just over 9% for males and just over 6% for females. The mean age of initiation is 20.8 for males and 20.8 for females. The probability of stopping smoking increases after 50 years of age, reaching a maximum of 27.0% for males and a maximum of 28.0% for females. The probability of smoking relapse is affected by large sampling variability. However, our results seem to indicate that it is about 80% in the first year, then declines to about 40% after two years and progressively becomes negligible after 3 years (Figure (1.2)). On average, former smokers relapse into smoking after 0.6 and 0.5 years, for males and females respectively (Table 1.1).

Panel (d) shows the estimated prevalence of never, current, and former smokers among those over 14 years old from 1993 to 2043, predicted through the SHC model, together with the observed data used to calibrate the model (blue and red dots respectively for males and females with their 90% CI). The model fit appears to be adequate, with the predicted values close to the observed ones. Our forecasts, starting from 2020, suggest that the smoking prevalence will decrease in the coming years. Panels (e) and (f) show the predicted SAD and PAF over the period 1993-2043, separately for males and females, calculated for the population over 35 years of age and for the population over 65 years of age. The impact on males is higher than on females both in absolute and relative terms. However, while a clear reduction of the attributable deaths is expected in the coming years for males, for females they slightly decline only after having reached a maximum around 2030 [45]. Note that the majority of attributable deaths in the population over 35 are due to deaths in individuals over 65, as shown by the similarity of the curves.

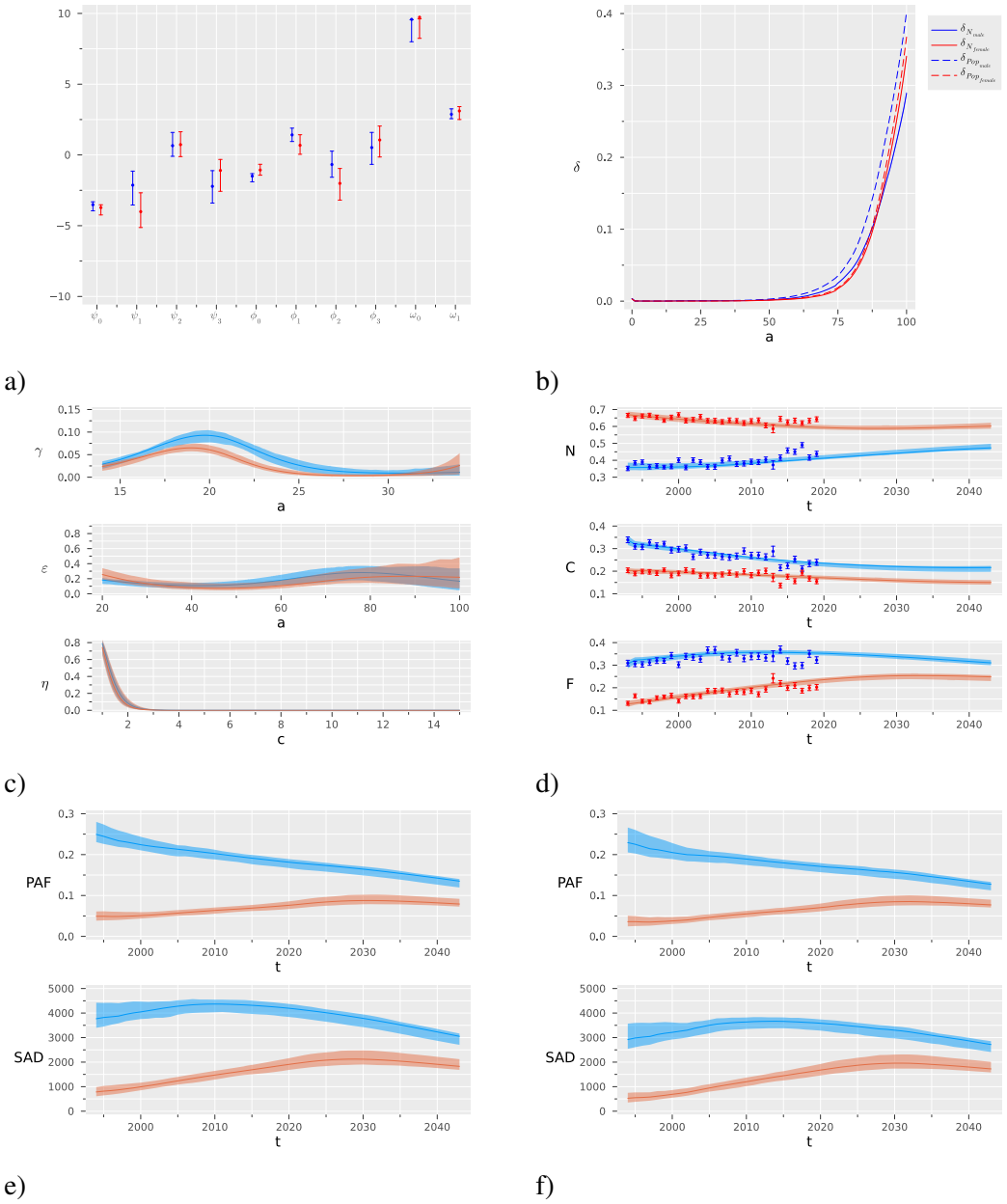


FIGURE 1.2: Results of the two-step estimation procedure for males in blue and females in red, with their bootstrap 90% confidence intervals: parameters tuning the probabilities of starting ($\boldsymbol{\psi}$) and stopping smoking ($\boldsymbol{\phi}$), and the probability of smoking relapse ($\boldsymbol{\omega}$) (a), age-specific mortality for never smokers and for the general population (b), probabilities of starting ($\gamma(a)$), and stopping smoking ($\varepsilon(a)$) and probability of smoking relapse ($\eta(c)$) (c), observed and predicted prevalence for never (N), current (C) and former (F) smokers (d), Population Attributable Fraction (PAF) and Smoking Attributable Deaths (SAD) for people over 35 years old (e) and over 65 years old (f).

TABLE 1.1: Summaries with 90% confidence intervals of the probabilities of starting and stopping smoking, and of the probability of smoking relapse, for males and females.

Smoking event	Sex	Maximum probability	Age at maximum	Mean probability	Mean time at the event
Starting	**Male**	0.09 (0.08 - 0.10)	20.0 (19.8 - 20.2)	0.04 (0.03 - 0.05)	20.8[1] (20.6 - 21.1)
	Female	0.06 (0.05 - 0.07)	19.2 (19.0 - 19.4)	0.03 (0.02 - 0.04)	20.8[1] (20.3 - 21.4)
Stopping	**Male**	0.27 (0.20 - 0.34)	75.4 (73.8 - 76.2)	0.17 (0.13 - 0.20)	63.9[1] (62.3 - 64.9)
	Female	0.28 (0.23 - 0.34)	20.1 (20.0 - 20.2)	0.12 (0.09 - 0.15)	54.7[1] (52.2 - 74.9)
Relapsing	**Male**	0.78 (0.73 - 0.81)	-	0.03 (0.02 - 0.04)	0.6[2] (0.5 - 0.7)
	Female	0.75 (0.68 - 0.80)	-	0.03 (0.02 - 0.04)	0.5[2] (0.4 - 0.6)

[1] Mean age of starting and stopping smoking
[2] Mean number of years from smoking cessation to relapse

Tables 1.2 and 1.3 report the percentages of never, current, and former smokers, the SAD and PAF, estimated every 10 years from 1993 to 2043, with their 90% confidence intervals. As an example, we estimated that in Tuscany in 2023 smoking was responsible for 4,070 (90% CI:3,795-4,247) deaths among men over 35 years old and 1,976 (90% CI:1,741-2,407) deaths among women in the same age class, corresponding to a PAF of 18% and 8%, respectively. Most of the attributable burden, however, was on people older than 65 (3,497 SAD for men and 1,765 for women).

TABLE 1.2: Estimated prevalence (%) of never, current, and former smokers in the population with 90% confidence intervals, evaluated every 10 years from 1993 to 2043, for males and females.

	Never		Current		Former	
Year	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
1993	35.7 (33.8 - 37.7)	66.9 (64.9 - 68.7)	33.7 (31.6 - 35.9)	20.3 (18.8 - 21.8)	30.6 (28.6 - 32.5)	12.8 (11.5 - 14.3)
2003	36.5 (35.0 - 38.3)	63.3 (61.7 - 64.9)	28.7 (27.6 - 29.4)	19.4 (18.5 - 20.2)	34.8 (33.7 - 35.9)	17.3 (16.5 - 18.3)
2013	39.2 (38.0 - 41.0)	60.6 (59.2 - 62.1)	25.1 (23.9 - 25.6)	18.2 (17.2 - 18.7)	35.7 (34.7 - 37.0)	21.2 (20.4 - 22.4)
2023	42.2 (41.2 - 44.0)	58.9 (57.7 - 60.3)	22.9 (21.4 - 23.3)	17.0 (15.8 - 17.5)	34.9 (33.7 - 36.2)	24.1 (23.1 - 25.4)
2033	45.2 (44.3 - 47.0)	59.0 (58.0 - 60.6)	21.7 (20.0 - 22.1)	15.9 (14.7 - 16.5)	33.2 (31.9 - 34.5)	25.1 (24.0 - 26.3)
2043	47.5 (46.8 - 49.7)	60.1 (59.0 - 61.8)	21.5 (19.7 - 22.0)	15.4 (14.1 - 16.1)	31.0 (29.7 - 32.2)	24.6 (23.4 - 25.8)

TABLE 1.3: Estimated number of Smoking Attributable Deaths (SAD), Population Attributable Fraction (PAF) (%), and the ratio between population size and SAD in the years 1993, 2003, 2013, 2023, 2033, 2043, 2053 and 2063, with 90% confidence intervals, among males and females aged over 35 and over 65.

		SAD (90% CI)		PAF (90% CI)		Pop/SAD (90% CI)	
Age	**Year**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
35+	**1993**	3,770 (3,371 - 4,219)	785 (640 - 970)	24.9 (22.9 - 27.1)	4.9 (4.1 - 6.0)	788 (704 - 882)	4,375 (3,540 - 5,366)
	2003	4,187 (3,911 - 4,446)	1,157 (999 - 1,317)	21.7 (20.3 - 22.9)	5.5 (4.8 - 6.2)	768 (722 - 824)	3,126 (2,745 - 3,625)
	2013	4,319 (4,084 - 4,498)	1,669 (1,501 - 1,800)	19.5 (18.5 - 20.2)	6.9 (6.3 - 7.5)	761 (730 - 806)	2,159 (1,999 - 2,404)
	2023	4,058 (3,804 - 4,213)	2,127 (1,906 - 2,268)	17.5 (16.5 - 18.2)	8.5 (7.7 - 9.1)	748 (719 - 799)	1,536 (1,438 - 1,716)
	2033	3,597 (3,335 - 3,743)	2,219 (1,983 - 2,370)	15.7 (14.6 - 16.3)	9.3 (8.3 - 9.9)	763 (732 - 826)	1,313 (1,227 - 1,472)
	2043	3,014 (2,763 - 3,155)	1,952 (1,745 - 2,084)	13.3 (12.2 - 14.0)	8.4 (7.6 - 9.0)	826 (787 - 904)	1,338 (1,251 - 1,500)
65+	**1993**	2,919 (2,512 - 3,358)	527 (381 - 709)	22.9 (20.4 - 25.5)	3.6 (2.6 - 4.8)	309 (268 - 359)	2,413 (1,790 - 3,339)
	2003	3,369 (3,096 - 3,627)	849 (694 - 999)	19.9 (18.4 - 21.3)	4.3 (3.5 - 5.1)	316 (291 - 347)	1,674 (1,418 - 2,054)
	2013	3,640 (3,425 - 3,810)	1,418 (1,250 - 1,547)	18.2 (17.2 - 19.0)	6.2 (5.5 - 6.8)	317 (301 - 339)	1,044 (954 - 1,188)
	2023	3,481 (3,249 - 3,635)	1,906 (1,691 - 2,044)	16.6 (15.5 - 17.3)	8.0 (7.2 - 8.6)	335 (319 - 361)	755 (702 - 854)
	2033	3,155 (2,911 - 3,304)	2,077 (1,844 - 2,229)	14.9 (13.8 - 15.6)	9.1 (8.1 - 9.7)	393 (374 - 429)	701 (651 - 793)
	2043	2,681 (2,448 - 2,821)	1,847 (1,642 - 1,919)	12.6 (11.5 - 13.2)	8.2 (7.3 - 8.8)	463 (438 - 511)	763 (710 - 862)

Regarding the CV procedure, the average values of MAPE for different prediction horizons are lower than 30% (Table 1.4), indicating that the predictive performance of the model is adequate, even if not optimal [46]. The MAPE is lower for the model on the male population than for the model on the female one.

TABLE 1.4: Cross-validation results: MAPE (%) calculated on four-time horizons for the model on males and the model on females.

Horizon	**Male**	**Female**
3 years	24.3	28.6
6 years	24.4	28.8
9 years	24.7	29.3
12 years	24.9	29.9

Figure 1.3 reports the results of the two separate calibrations of the SHC model, one on the prevalence data from 1993 to 2004 and one on the prevalence data from 2005 to 2019. The confidence bands are wider in the second period of calibration than in the first one. For males, there is evidence of a downward shift of age corresponding to the maximum probability of starting smoking. For females, calibrating the model in the first years brought a lower projection of the prevalence of never smokers, which likely reflects a change over time in the smoking habits among women. Apart from these differences, the two calibrations provided qualitatively similar results. For numerical details see Tables and Figures in Section A.5, Supplemental Material.

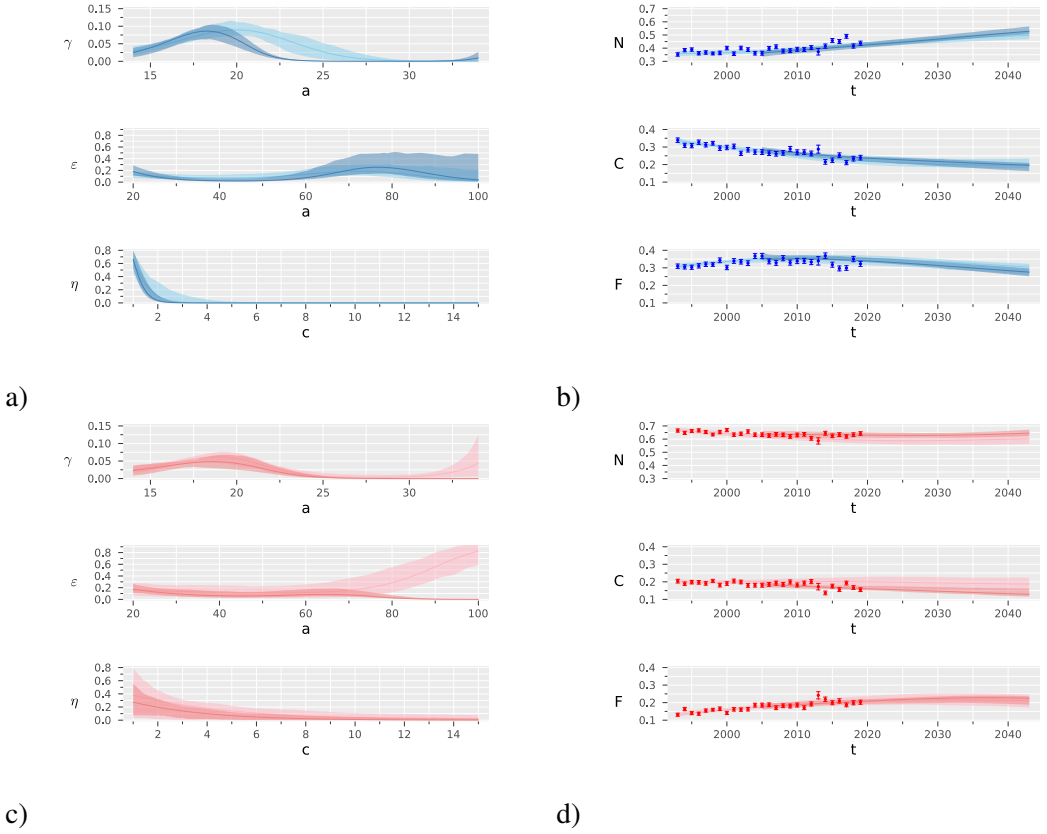


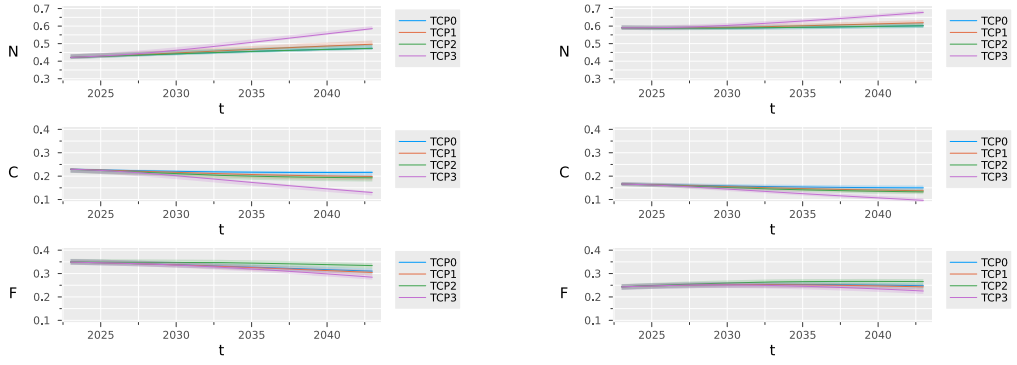
FIGURE 1.3: Results of the two-step estimation procedure for males in blue and females in red, by the period of calibration (from 1993 to 2004 in a light color and from 2005 to 2019 in a dark color): probabilities of starting ($\gamma(a)$) and stopping smoking ($\varepsilon(a)$), and probability of smoking relapse ($\eta(c)$), with 90% confidence bands, (a) and (c); the prevalence of never (N), current (C) and former (F) smokers, with 90% confidence bands, (b) and (d).

The total variance indices derived from the GSA (Table 1.5) reveal that the primary factor contributing to the variability of the Hellinger distance is the probability of starting smoking and its interaction with the other model inputs, resulting in S_i^{tot} values of 0.58 for males and 0.76 for females. This is followed by the probability of quitting smoking, with values of 0.36 for males and 0.21 for females, and by the probability of smoking relapse, with values of 0.15 for males and 0.09 for females. Conversely, the parameters assumed to be fixed have a negligible impact on the Hellinger distance, with total variance indices very close to 0. On the one hand, this result suggests that calibration does not provide meaningful information about $\boldsymbol{\pi}$, ν, and the *RR*s, thereby supporting our decision to treat these parameters as fixed, while focusing on inferring $\boldsymbol{\psi}$, $\boldsymbol{\phi}$, and $\boldsymbol{\omega}$. On the other hand, it indicates that fixing the aforementioned parameters to specific values does not significantly affect the calibration results and prevalence estimates, demonstrating their robustness against variations in $\boldsymbol{\pi}$, ν, and the *RR*s specifications.

TABLE 1.5: Total variance indices quantifying the contribution of each input on the Hellinger distance, calculated for the model on males and the model on females.

Input	**Male**	**Female**
$\boldsymbol{\psi}$	0.58	0.76
$\boldsymbol{\phi}$	0.36	0.21
$\boldsymbol{\omega}$	0.15	0.09
$\boldsymbol{\pi}$	< 0.01	< 0.01
$\boldsymbol{\nu}$	< 0.01	< 0.01
RRs	< 0.01	< 0.01

Figure 1.4 compares the evolution of smoking habits in the male and female populations under three alternative scenarios that simulate hypothetical tobacco control policies. These scenarios are compared with the status quo, corresponding to the absence of actions to reduce tobacco consumption (TCP0). We assumed that the TCPs have been applied since 2023. They have no substantial effect on the prevalence of never, former, and current smokers during the 10 years following their implementation. TCP3 has the largest impact: in 2043 it is expected to increase by 12 percentage points the prevalence of never-smokers among males and by 8 among females, compared with TCP0 (see Table 1.6).



a) b)

FIGURE 1.4: Estimated prevalence of never (N), current (C) and former (F) smokers under different tobacco control policies (TCP) with 90% confidence intervals, for males (a) and females (b).

TABLE 1.6: Estimated prevalence (%) of never, current, and former smokers in the population under different tobacco control policies (TCP) evaluated in 2023, 2033, and 2043, with 90% confidence intervals, for males and females.

		Male				Female			
Status	**Year**	**TCP0**	**TCP1**	**TCP2**	**TCP3**	**TCP0**	**TCP1**	**TCP2**	**TCP3**
Never	**2023**	42.2 (41.2 - 44.0)	42.2 (41.2 - 44.0)	42.2 (41.2 - 44.0)	42.2 (41.3 - 44.0)	58.9 (57.7 - 60.3)	58.9 (57.7 - 60.3)	58.9 (57.7 - 60.3)	58.9 (57.7 - 60.3)
	2033	45.2 (44.3 - 47.0)	46.0 (45.2 - 47.9)	45.1 (44.0 - 47.0)	49.0 (48.1 - 50.6)	59.0 (58.0 - 60.6)	59.7 (58.6 - 61.2)	59.0 (58.0 - 60.5)	61.7 (60.7 - 63.0)
	2043	47.5 (46.8 - 49.7)	49.8 (49.1 - 51.9)	47.5 (46.8 - 49.7)	58.7 (58.2 - 60.2)	60.1 (59.0 - 61.8)	61.8 (60.8 - 63.4)	60.0 (59.0 - 61.8)	67.7 (66.9 - 68.9)
Current	**2023**	22.9 (21.4 - 23.3)	22.9 (21.4 - 23.3)	22.9 (21.4 - 23.3)	22.9 (21.4 - 23.3)	17.0 (15.8 - 17.5)	17.0 (15.8 - 17.5)	17.0 (15.8 - 17.5)	17.0 (15.8 - 17.5)
	2033	21.7 (20.0 - 22.1)	20.9 (19.3 - 21.4)	20.2 (18.6 - 20.7)	18.4 (16.9 - 18.8)	15.9 (14.7 - 16.5)	15.3 (14.1 - 16.0)	14.8 (13.6 - 15.5)	13.7 (12.6 - 14.3)
	2043	21.5 (19.7 - 22.0)	19.7 (18.0 - 20.2)	19.2 (17.5 - 19.7)	12.9 (11.6 - 13.3)	15.4 (14.1 - 16.1)	14.1 (12.9 - 14.8)	13.6 (12.4 - 14.3)	10.1 (9.1 - 10.7)
Former	**2023**	34.9 (33.7 - 36.2)	34.9 (33.7 - 36.2)	35.0 (33.7 - 36.3)	34.9 (33.7 - 36.2)	24.1 (23.1 - 25.4)	24.1 (23.1 - 25.4)	24.1 (23.1 - 25.4)	24.1 (23.1 - 25.4)
	2033	33.2 (31.9 - 34.5)	33.0 (31.8 - 34.4)	34.7 (33.3 - 36.0)	32.7 (31.5 - 34.0)	25.1 (24.0 - 26.3)	25.0 (23.9 - 26.2)	26.1 (25.0 - 27.4)	24.6 (23.5 - 25.8)
	2043	31.0 (29.7 - 32.2)	30.5 (29.2 - 31.6)	33.4 (31.9 - 34.5)	28.4 (27.3 - 29.6)	24.6 (23.4 - 25.8)	24.1 (22.9 - 25.2)	26.3 (25.1 - 27.5)	22.3 (21.3 - 23.3)

In order to better appreciate the impact of the TCPs on mortality, we extended the forecasting horizon up to 2063. Table 1.7 reports the predicted number of attributable deaths every 10 years, from 2023 to 2063, for both males and females under different TCPs, for the classes of age 35+ and 65+. TPC2, which increases the probability of stopping smoking, is the policy that most impacts mortality in both classes of age. TPC3, which bans access to smoking to the new generations, despite its effectiveness in reducing current smokers, does not reduce SADs within the time window considered. Indeed, this policy is expected to have a longer-term impact, which is not visible before 2063. Additional Tables and Figures are reported in Section A.5, Supplemental Material.

TABLE 1.7: Expected number of Smoking Attributable Deaths (SAD) under different tobacco control policies (TCP), in the years 2023, 2033, 2043, 2053, and 2063, with 90% confidence intervals, among males and females aged over 35 and over 65.

		Male				Female			
Age	**Year**	**TCP0**	**TCP1**	**TCP2**	**TCP3**	**TCP0**	**TCP1**	**TCP2**	**TCP3**
35+	**2023**	4,058 (3,804 - 4,212)	4,058 (3,804 - 4,213)	4,058 (3,804 - 4,212)	4,058 (3,804 - 4,213)	2,127 (1,906 - 2,268)	2,127 (1,906 - 2,268)	2,127 (1,906 - 2,268)	2,127 (1,906 - 2,268)
	2033	3,597 (3,335 - 3,743)	3,596 (3,334 - 3,742)	3,547 (3,288 - 3,695)	3,597 (3,335 - 3,743)	2,219 (1,983 - 2,370)	2,218 (1,983 - 2,370)	2,188 (1,953 - 2,339)	2,219 (1,983 - 2,370)
	2043	3,014 (2,763 - 3,155)	3,011 (2,760 - 3,152)	2,848 (2,605 - 2,989)	3,014 (2,763 - 3,155)	1,952 (1,745 - 2,084)	1,950 (1,743 - 2,083)	1,849 (1,645 - 1,981)	1,952 (1,745 - 2,084)
	2053	2,409 (2,194 - 2,514)	2,397 (2,182 - 2,501)	2,188 (1,984 - 2,294)	2,360 (2,157 - 2,463)	1,590 (1,418 - 1,688)	1,584 (1,412 - 1,683)	1,452 (1,285 - 1,549)	1,575 (1,402 - 1,673)
	2063	1,892 (1,711 - 1,968)	1,858 (1,678 - 1,934)	1,667 (1,500 - 1,739)	1,742 (1,570 - 1,815)	1,101 (976 - 1,165)	1,087 (963 - 1,150)	969 (853 - 1,019)	1,053 (930 - 1,116)
65+	**2023**	3,481 (3,249 - 3,635)	3,481 (3,249 - 3,635)	3,481 (3,249 - 3,635)	3,481 (3,249 - 3,635)	1,906 (1,691 - 2,044)	1,906 (1,691 - 2,044)	1,906 (1,691 - 2,044)	1,906 (1,691 - 2,044)
	2033	3,155 (2,911 - 3,304)	3,155 (2,911 - 3,304)	3,113 (2,869 - 3,262)	3,155 (2,911 - 3,304)	2,077 (1,844 - 2,229)	2,077 (1,844 - 2,229)	2,049 (1,816 - 2,199)	2,077 (1,844 - 2,229)
	2043	2,681 (2,448 - 2,821)	2,681 (2,448 - 2,821)	2,534 (2,307 - 2,675)	2,681 (2,448 - 2,821)	1,847 (1,642 - 1,979)	1,847 (1,642 - 1,919)	1,750 (1,548 - 1,881)	1,847 (1,642 - 1,979)
	2053	2,081 (1,887 - 2,183)	2,081 (1,887 - 2,183)	1,892 (1,710 - 1,997)	2,081 (1,887 - 2,183)	1,486 (1,316 - 1,585)	1,486 (1,316 - 1,585)	1,359 (1,198 - 1,457)	1,486 (1,316 - 1,585)
	2063	1,553 (1,397 - 1,621)	1,552 (1,396 - 1,619)	1,372 (1,230 - 1,434)	1,553 (1,397 - 1,621)	994 (877 - 1,053)	993 (876 - 1,052)	877 (768 - 934)	994 (877 - 1,053)

1.4 Discussion

Interesting findings emerged from our analysis. We found that the probability of starting smoking reaches its maximum of just over 9% for males and just over 6% for females between 19 and 20 years of age. Considering that younger people have a large probability to become stable smokers [47], these probabilities are quite worrying. The difference in the mean age of initiation between males and females is lower than one year, confirming what is reported for high-income countries [48]. Regarding the probability of stopping smoking, we found that it increases after 50 years

of age and has a maximum of 29.5% for males and 24.0% for females, even if the confidence bands around these curves are quite wide. The 80% of ex-smokers relapse into smoking after 1 year, in line with the results of the Italian surveillance system PASSI for the years 2020-2021 (www.epicentro.iss.it/passi/dati/SmettereFumo). On average, former smokers relapse into smoking during the second year from cessation (after 1.7 and 1.5 years for males and females, respectively).

According to our model, in 2023 in Tuscany, 23% of men smoke, while 35% are ex-smokers. These percentages are lower among women: 16% smoke and 24% are ex-smokers. The prevalence of smokers estimated by our model is lower than the one reported in the PASSI survey for the period 2020-2021 (26.1% and 20.5% in the age class 18-69 for males and females, respectively), but consistent if we consider that our estimates are calculated on all population, while PASSI focuses on the age class 18-69 (www.epicentro.iss.it/passi/pdf2020/Scheda-fumo-PASSI-regione-2016-2019.pdf).

We estimated that, in 2023, 18% of deaths among males and 8% among females are due to smoking, corresponding to 4,070 and 1,976 deaths, respectively. These PAFs are in line with those estimated by the Global Burden of Disease Study for Italy in 2019 (https://vizhub.healthdata.org/gbd-results/): 20.5% (CI: 19.5-21.7) in males and 8.17% (CI: 7.51-9.02) in females, slightly lower than those reported for Italy by the Tobacco Atlas initiative (https://tobaccoatlas.org/challenges/deaths/) and overall coherent with previous results for Italy and Tuscany ([49]; www.deathsfromsmoking.net).

As shown by the cross-validation results, the model produces quite reliable predictions of prevalence. Thus, subject to the assumption that all mechanisms underlying smoking dynamics and demographic evolution do not change in the future, we projected the dynamics. For the next two decades, we estimated an evident decrease in the prevalence of current smokers for males, due to an increase in the percentage of never-smokers. For females, substantial stability is expected. Similar considerations apply to PAFs: a decrease is observed for males and stability for females. These results confirm that Italy is in the fourth stage of the tobacco epidemic model, characterized by a continuing slow decline of smoking prevalence in both men and women with converging rates between sex [50, 51].

The proposed model can be used for assessing the impact of alternative TPCs. For illustrative purposes, we considered the impact of three policies aimed at reducing smoking in the population. The first two policies are completely hypothetical and defined in terms of their effect on the probability of starting (TCP1) and stopping (TPC2) smoking. They are not real policies but represent the intentions of the legislator to change the rates of smoking initiation and cessation. The third one (TCP3), which bans smoking in new cohorts since 2009, is inspired by the tobacco-free generation real intervention implemented in New Zealand as part of a plan for the tobacco endgame, including also additional strategies aimed at decreasing the affordability and availability

of smoking, reducing the levels of nicotine in tobacco products, and restricting sales to designated tobacco outlets. We evaluated the expected marginal impact that this tobacco-free generation intervention would have in Tuscany, assuming complete compliance of new generations to the smoking ban. The results indicate that under TPC1 and TPC2 the prevalence of current smokers is reduced by a few percentage points either for women or men. On the contrary, TCP3 produces a clear increase in never-smokers, thus a reduction in smoking prevalence, which is expected to decrease in ten years by 9 and 6 percentage points among males and females, respectively. The impact on mortality of the three policies, in particular TCP1 and TPC3, that act by increasing the number of never smokers, can be appreciated only by extending the time horizon of forecasting. Interventions able to increase the probability of stopping smoking, like TPC2, are expected to produce the largest reduction of SADs in the medium term, especially among the over-65s. However, this kind of policy does not contribute to reducing smoking among the youngest, thus effectively stopping the tobacco epidemic.

From a methodological point of view, we introduced several elements of novelty. First of all, we provided a formal definition of the equations that describe the system dynamics and made explicit assumptions on the distribution of the involved random variables. We also introduced cubic regression splines for modelling in a flexible way the probabilities of starting and quitting smoking as functions of age, thus obtaining more realistic trajectories. Furthermore, we included in the model dependencies from the smoking intensity, which may allow assessing the impact of personalized TCPs specific for heavy or moderate smokers, such as lung cancer screening, use of pharmacological treatment, or smoking cessation campaigns.

Regarding the inference on the unknown parameters, we proposed a two-step estimation strategy to estimate the curves describing the probability of starting and stopping smoking and the probability of smoking relapse, as well as the mortality risk among never, current and former smokers. At the second step of the estimation procedure, we defined the calibration objective function in terms of a Hellinger distance between observed and predicted prevalence, instead of the widely used sum of squares function. The use of this discrepancy measure is relatively new in this framework and allowed handling a bounded loss function, defined in $[0, 1]$, that is simple to minimise and to be interpreted. Finally, we provided confidence intervals/bands for the parameters/curves of interest. To the best of our knowledge, this is the first time that quantification of sampling variability is performed in this field. To this aim, we resorted to a parametric bootstrap procedure defined by adapting to our framework a method proposed for compartmental models describing infectious dynamics [30]. Note that this quantification of the sampling variability accounts for the sample size of the surveys from which we derived the observed prevalence used in the estimation. The estimation procedure has also limitations. We estimated the parameters in a deterministic way, in the sense that we considered the distributional assumptions on the prevalence only in the bootstrap procedure but not in the calibration phase. While likelihood-based approaches are unfeasible in this framework, likelihood-free inference

methods such as Approximate Bayesian Computation algorithms would allow a full uncertainty quantification [25].

The reliability of the model's results depends on several factors. First of all, it depends on the quality of the data used for calibration. In our case, we used data from yearly surveys conducted according to well-established methodology on reasonably large sample sizes. Secondly, it depends on the appropriateness of the values assigned to the parameters treated as fixed. We addressed this point by proposing the use of GSA in this framework. The GSA results revealed that the prevalence estimates were robust to variations of the fixed parameters within plausible ranges of values. Lastly, the reliability of the results depends on the structural assumptions on which the model is based, not assessed via GSA. Underneath, we qualitatively review the main assumptions of the model and discuss the limitations that may arise from them. With respect to demographic dynamics, we assumed that the population was close to immigration and emigration and that the number of new births did not vary during the study period, effectively feeding the model with identical cohorts of subjects each year. For more realistic modelling, we could use the observed yearly number of births to create the new cohorts up to 2019. However, in light of the GSA, we expect that the impact of this choice on the results has not been significant. We also assumed that the age-specific mortality rates did not vary over the study period.

Regarding smoking dynamics, we assumed that the probabilities of starting and stopping smoking were functions of age and that the probability of smoking relapse was a function of time since cessation, but not of age. We did not allow any of these probabilities to vary over time. By defining the transition probabilities in this way, we have made a clear choice about which time axes were most important in our opinion to capture appropriately the smoking dynamics in the population. This choice is not without problems because in some cases there is evidence suggesting otherwise. For example, a decreasing trend in the probability of starting smoking has been reported for both males and females in Europe [52], while evidence of a dependence between age and risk of smoking relapse has been found in the US population [53]. However, if introducing multiple time-axes dependence in the transition probabilities could lead to more realistic results, this would be at the price of further complicating the model by introducing new unknown parameters to be estimated. We partially explored the goodness of the assumption of no calendar time dependence through a simple sensitivity analysis, which confirmed that the probabilities of starting and stopping smoking, and the probability of smoking relapse were quite similar when two separate calibrations were performed on the periods 1993-2004 and 2005-2019. It is worth stressing that, even if these two periods correspond to before and after the introduction of the so-called Sirchia law that banned smoking in all indoor public places in Italy, it was not our goal to speculate about the causal effect of this intervention on smoking dynamics. We also assumed that people could not change their smoke intensity during their entire life, that the probability of stopping and relapsing did not depend on smoking intensity, and, again, that the distribution of smokers by smoking intensity did not change over the study period.

Regarding the assumption that individual cannot change level of smoking, it has been reported in the literature that the probability that smokers that do not stop smoking change their smoking habits is negligible [54]. Moreover, the assumption of no transitions between levels of smoking is necessary in order not to further complicate the already complex differential equations system governing the model. Regarding the assumption of defined the bounded transitions probabilities, relying on various research studies conducted by the Italian Superior Institute of Health and carried out in Lugo et al. [55], revealed that over 95% of current smokers began smoking regularly before the age of 25, and that in the age group 15-24 only 1.4% are former smokers. Furthermore, national data provided by ISTAT also confirm that there are few former smokers before the age of 20. Accordingly, we speculated that the number of individuals who quit smoking before 20 years of age is very small and that their impact on the whole dynamic may be considered as negligible.

In general, it is important to note that, underlying all the simplifications introduced in model specification, is the fact that adding details to a compartmental model goes along with the definition of new compartments and new transitions, and without available and reliable data, the model could become non-identifiable producing more uncertain and unstable results [56]. Moreover, microsimulation models or social network models, that explore smoking dynamics from an individual point of view, could be more suitable solutions to introduce detail and complexity, including those related, for example, to the course of disease [57, 58], or to explore the exposure to second-hand smoke that, being related to the social network of the individuals, was not considered in our analysis.

1.5 Conclusions

We developed an approach for modelling smoking dynamics in the population, that overcomes many of the limitations of previously proposed models. It includes validation tools like cross-validation on a rolling basis and GSA, aimed at checking the robustness of our results and supporting our findings. The model can be easily generalized and applied to other regions and countries, after carefully checking the validity of the model assumptions in different contexts. It can be also used to assess the impact of other tobacco control policies on smoking prevalence and mortality, beyond those considered in this paper.

Authors' contributions: A.L., C.V. and M.B. conceptualized the model and wrote the first version of the paper; A.L. wrote the code and conducted the statistical analysis; G.Ce. and G.Ca. provided critical feedback; M.B. supervised the project. All the authors read and approved the final version of the paper.

Bibliography

[1] IARC, editor. *Tobacco smoke and involuntary smoking: this publication represents the views and expert opinions of an IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, which met in Lyon, 11 - 18 June 2002*. Number 83 in IARC monographs on the evaluation of carcinogenic risks to humans. IARC, Lyon, 2004. ISBN 978-92-832-1283-6.

[2] Institute of Medicine (U.S.), R.J. Bonnie, K.R. Stratton, and R.B. Wallace, editors. *Ending the tobacco problem: a blueprint for the nation*. National Academies Press, Washington, DC, 2007. ISBN 978-0-309-10382-4.

[3] IARC, editor. *A review of human carcinogens*. Number 100 in IARC monographs on the evaluation of carcinogenic risks to humans. IARC, Lyon, 2012. ISBN 978-92-832-1329-1 978-92-832-1318-5 978-92-832-1319-2 978-92-832-1320-8 978-92-832-1321-5 978-92-832-1322-2 978-92-832-1323-9.

[4] B. Loring. *Tobacco and inequities: guidance for addressing inequities in tobacco-related harm*. World Health Organization, Regional Office for Europe, Copenhagen, Denmark, 2014. ISBN 978-92-890-5049-4.

[5] GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258):1223–1249, 2020. ISSN 01406736. doi: 10.1016/S0140-6736(20)30752-2. URL `https://linkinghub.elsevier.com/retrieve/pii/S0140673620307522`.

[6] World Health Organization. *Tobacco control for sustainable development*. Regional Office for South-East Asia, 2017. ISBN 978-92-9022-578-2.

[7] G. Gorini, G. Carreras, A. Lugo, S. Gallus, M. Masocco, L. Spizzichino, and V. Minardi. Electronic cigarette use as an aid to quit smoking: Evidence from PASSI survey, 2014–2021. *Preventive Medicine*, 166:107391, 2023. ISSN 00917435. doi: 10.1016/j.ypmed.2022.107391. URL `https://linkinghub.elsevier.com/retrieve/pii/S009174352200456X`.

[8] G. Gorini, S. Gallus, G. Carreras, B. De Mei, M. Masocco, F. Faggiano, L. Charrier, F. Cavallo, L. Spizzichino, D. Galeone, V. Minardi, S. Lana, A. Lachi, R. Pacifici, B. Cortini, L. Mastrobattista, C. Mortali, R. Di Pirchio, G. Ferrante, and F. Barone-Adesi. Prevalence of tobacco smoking and electronic cigarette use among adolescents in Italy: Global Youth Tobacco Surveys (GYTS), 2010, 2014, 2018. *Preventive Medicine*, 131:105903, 2020. ISSN 00917435. doi: 10.1016/j.ypmed.2019.105903. URL https://linkinghub.elsevier.com/retrieve/pii/S0091743519303834.

[9] S. Cerrai, E. Benedetti, E. Colasante, M. Scalese, G. Gorini, S. Gallus, and S. Molinaro. E-cigarette use and conventional cigarette smoking among European students: findings from the 2019 ESPAD survey. *Addiction*, 117(11):2918–2932, 2022. ISSN 0965-2140, 1360-0443. doi: 10.1111/add.15982. URL https://onlinelibrary.wiley.com/doi/10.1111/add.15982.

[10] D. Mendez, K.E. Warner, and P.N. Courant. Has Smoking Cessation Ceased? Expected Trends in the Prevalence of Smoking in the United States. *American Journal of Epidemiology*, 148(3):249–258, 1998. ISSN 0002-9262, 1476-6256. doi: 10.1093/oxfordjournals.aje.a009632. URL https://academic.oup.com/aje/article-lookup/doi/10.1093/oxfordjournals.aje.a009632.

[11] D.T. Levy and K. Friend. A Simulation Model of Policies Directed at Treating Tobacco Use and Dependence. *Medical Decision Making*, 22(1):6–17, 2002. ISSN 00000000, 0272989X. doi: 10.1177/02729890222062874. URL http://mdm.sagepub.com/cgi/doi/10.1177/02729890222062874.

[12] G. Carreras, S. Gallus, L. Iannucci, and G. Gorini. Estimating the probabilities of making a smoking quit attempt in Italy: stall in smoking cessation levels, 1986-2009. *BMC Public Health*, 12(1):183, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-183. URL http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-183.

[13] National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Reports of the Surgeon General. Centers for Disease Control and Prevention (US), Atlanta (GA), 2014. URL http://www.ncbi.nlm.nih.gov/books/NBK179276/.

[14] E.J. Feuer, D.T. Levy, and W.J. McCarthy. Chapter 1: The Impact of the Reduction in Tobacco Smoking on U.S. Lung Cancer Mortality, 1975-2000: An Introduction to the Problem: Introduction: Impact of the Reduction in Tobacco Smoking on U.S. Lung Cancer Mortality. *Risk Analysis*, 32:S6–S13, 2012. ISSN 02724332. doi: 10.1111/j.1539-6924.

2011.01745.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.2011.01745.x.

[15] D.T. Levy, L. Nikolayev, E. Mumford, and C. 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*, 16(4):359–371, 2005. ISSN 0957-5243, 1573-7225. doi: 10.1007/s10552-004-7841-4. URL http://link.springer.com/10.1007/s10552-004-7841-4.

[16] A. Singh, N. Wilson, and T. Blakely. Simulating future public health benefits of tobacco control interventions: a systematic review of models. *Tobacco Control*, 30(4):460–470, 2021. ISSN 0964-4563, 1468-3318. doi: 10.1136/tobaccocontrol-2019-055425. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tobaccocontrol-2019-055425.

[17] D.T. Levy, S. Gallus, K. Blackman, G. Carreras, C. La Vecchia, and G. Gorini. Italy SimSmoke: the effect of tobacco control policies on smoking prevalence and smoking-attributable deaths in Italy. *BMC Public Health*, 12(1):709, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-709. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-709.

[18] D.T. Levy, L.M. Sánchez-Romero, Y. Li, Z. Yuan, N. Travis, M.J. Jarvis, J. Brown, and A. McNeill. England SimSmoke: the impact of nicotine vaping on smoking prevalence and smoking-attributable deaths in England. *Addiction*, 116(5):1196–1211, 2021. ISSN 0965-2140, 1360-0443. doi: 10.1111/add.15269. URL https://onlinelibrary.wiley.com/doi/10.1111/add.15269.

[19] A.M. Near, K. Blackman, L.M. Currie, and D.T. Levy. Sweden SimSmoke: the effect of tobacco control policies on smoking and snus prevalence and attributable deaths. *The European Journal of Public Health*, 24(3):451–458, 2014. ISSN 1101-1262, 1464-360X. doi: 10.1093/eurpub/ckt178. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckt178.

[20] L.M. Sánchez-Romero, L. Zavala-Arciniega, L.M. Reynales-Shigematsu, B.S. De Miera-Juárez, Z. Yuan, Y. Li, Y.K. Lau, N.L. Fleischer, R. Meza, J.F. Thrasher, and D.T. Levy. The Mexico SimSmoke tobacco control policy model: Development of a simulation model of daily and nondaily cigarette smoking. *PLOS ONE*, 16(6):e0248215, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0248215. URL https://dx.plos.org/10.1371/journal.pone.0248215.

[21] Institute of Medicine (U.S.), R.B. Wallace, A. Geller, and V.A. Ogawa, editors. *Assessing the use of agent-based models for tobacco regulation*. National Academies Press, Washington, D.C, 2015. ISBN 978-0-309-31722-1.

[22] J. Tam, D.T. Levy, J. Jeon, J. Clarke, S. Gilkeson, T. Hall, E.J. Feuer, T.R. Holford, and R. Meza. Projecting the effects of tobacco control policies in the USA through microsimulation: a study protocol. *BMJ Open*, 8(3):e019169, 2018. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2017-019169. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2017-019169.

[23] G. Carreras, G. Gorini, and E. 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*, 5(6):874–882, 2012. ISSN 1940-6207, 1940-6215. doi: 10.1158/1940-6207.CAPR-12-0019. URL https://aacrjournals.org/cancerpreventionresearch/article/5/6/874/49965/Can-a-National-Lung-Cancer-Screening-Program-in.

[24] A. Lachi, C. Viscardi, M.C. Malevolti, G. Carreras, and M. Baccini. Compartmental models in epidemiology: Application on smoking habits in tuscany. In *Book of short papers SIS*, pages 1437–1442. Pearson, 2022.

[25] A. Lachi, C. Viscardi, G. Cereda, G. Carreras, and M. Baccini. A compartmental models for smoking dynamics in Italy: A pipeline for inference, validation, and forecasting under hypothetical scenarios. preprint, In Review, 2023. URL https://www.researchsquare.com/article/rs-3303111/v1.

[26] L.D. Broemeling. *Bayesian analysis of infectious diseases: COVID-19 and beyond*. Chapman and Hall/CRC biostatistics series. CRC Press, Taylor and Francis Group, Boca Raton London New York, 2021. ISBN 978-0-367-64724-7 978-0-367-63386-8.

[27] M. Baccini, G. Cereda, and C. Viscardi. The first wave of the SARS-CoV-2 epidemic in Tuscany (Italy): A SI2R2D compartmental model with uncertainty evaluation. *PLOS ONE*, 16(4):e0250029, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0250029. URL https://dx.plos.org/10.1371/journal.pone.0250029.

[28] R.T. Hoogenveen, P.Hm. Van Baal, H.C. Boshuizen, and T.L. Feenstra. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Effectiveness and Resource Allocation*, 6(1):1, 2008. ISSN 1478-7547. doi: 10.1186/1478-7547-6-1. URL http://resource-allocation.biomedcentral.com/articles/10.1186/1478-7547-6-1.

[29] B. Efron and R. Tibshirani. *An introduction to the bootstrap*. Number 57 in Monographs on statistics and applied probability. Chapman and Hall, New York, 1993. ISBN 978-0-412-04231-7.

[30] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*,

2(3):379–398, 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042717300234.

[31] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-05997-5.

[32] M.J. Thun, B.D. Carter, D. Feskanich, N.D. Freedman, R. Prentice, A.D. Lopez, P. Hartge, and S.M. Gapstur. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine*, 368(4):351–364, 2013. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMsa1211127. URL http://www.nejm.org/doi/10.1056/NEJMsa1211127.

[33] H. Ryan, A. Trosclair, and J. Gfroerer. Adult Current Smoking: Differences in Definitions and Prevalence Estimates—NHIS and NSDUH, 2008. *Journal of Environmental and Public Health*, 2012:1–11, 2012. ISSN 1687-9805, 1687-9813. doi: 10.1155/2012/918368. URL http://www.hindawi.com/journals/jeph/2012/918368/.

[34] G. Carreras, G. Gorini, S. Gallus, L. Iannucci, and D.T. Levy. Predicting the future prevalence of cigarette smoking in Italy over the next three decades. *European Journal of Public Health*, 22(5):699–704, 2012. ISSN 1464-360X, 1101-1262. doi: 10.1093/eurpub/ckr108. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckr108.

[35] E. Hellinger. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136):210–271, 1909. ISSN 1435-5345, 0075-4102. doi: 10.1515/crll.1909.136.210. URL https://www.degruyter.com/document/doi/10.1515/crll.1909.136.210/html.

[36] P.K. Mogensen and A.N. Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615, 2018. ISSN 2475-9066. doi: 10.21105/joss.00615. URL http://joss.theoj.org/papers/10.21105/joss.00615.

[37] W. Zucchini, I.L. MacDonald, and R. Langrock. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC, New York, second edition edition, 2016. ISBN 978-1-03-217949-0 978-1-4822-5383-2. doi: 10.1201/b20790.

[38] K. Roosa and G. Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. *Theoretical Biology and Medical Modelling*, 16(1):1, 2019. ISSN 1742-4682. doi: 10.1186/s12976-018-0097-6. URL https://tbiomed.biomedcentral.com/articles/10.1186/s12976-018-0097-6.

[39] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance-based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018. URL https://linkinghub.elsevier.com/retrieve/pii/S0010465509003087.

[40] I.M. Sobol. *A primer for the Monte Carlo method*. CRC Press, Boca Raton, 1994. ISBN 978-0-8493-8673-2.

[41] S. Kucherenko, D. Albrecht, and A. Saltelli. Exploring multi-dimensional spaces: a comparison of latin hypercube and quasi monte carlo sampling techniques. *arXiv - University of Cornell (USA). JRC98050*, 2015. doi: https://doi.org/10.48550/arXiv.1505.02350.

[42] GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in the 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*, 397(10292):2337–2360, 2021. ISSN 01406736. doi: 10.1016/S0140-6736(21)01169-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673621011697.

[43] M.C. Kulik, W.J. Nusselder, H.C. Boshuizen, S.K. Lhachimi, E. Fernández, P. Baili, K. Bennett, J.P. Mackenbach, and H.A. Smit. Comparison of Tobacco Control Scenarios: Quantifying Estimates of Long-Term Health Impact Using the DYNAMO-HIA Modeling Tool. *PLoS ONE*, 7(2):e32363, 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0032363. URL https://dx.plos.org/10.1371/journal.pone.0032363.

[44] M. Hara, T. Sobue, S. Sasaki, S. Tsugane, and the JPHC Study Group. Smoking and Risk of Premature Death among Middle-aged Japanese: Ten-year Follow-up of the Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Diseases (JPHC Study) Cohort I. *Japanese Journal of Cancer Research*, 93(1):6–14, 2002. ISSN 09105050. doi: 10.1111/j.1349-7006.2002.tb01194.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1349-7006.2002.tb01194.x.

[45] M. Wensink, J.A. Alvarez, S. Rizzi, F. Janssen, and R. Lindahl-Jacobsen. Progression of the smoking epidemic in high-income regions and its effects on male-female survival differences: a cohort-by-age analysis of 17 countries. *BMC Public Health*, 20(1):39, 2020. ISSN 1471-2458. doi: 10.1186/s12889-020-8148-4. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-8148-4.

[46] N. Meade. Industrial and business forecasting methods, Lewis, C.D., Borough Green, Sevenoaks, Kent: Butterworth, 1982. Price: £9.25. Pages: 144. *Journal of Forecasting*, 2(2):194–196, 1983. ISSN 02776693, 1099131X. doi: 10.1002/for.3980020210. URL https://onlinelibrary.wiley.com/doi/10.1002/for.3980020210.

[47] S.D. Mahajan, G.G. Homish, and A. Quisenberry. Multifactorial Etiology of Adolescent Nicotine Addiction: A Review of the Neurobiology of Nicotine Addiction and Its Implications for Smoking Cessation Pharmacotherapy. *Frontiers in Public Health*, 9: 664748, 2021. ISSN 2296-2565. doi: 10.3389/fpubh.2021.664748. URL https://www.frontiersin.org/articles/10.3389/fpubh.2021.664748/full.

[48] M.B. Reitsma, L.S. Flor, E.C. Mullany, V. Gupta, S.I. Hay, and E. Gakidou. Spatial, temporal, and demographic patterns in the prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. *The Lancet Public Health*, 6(7): e472–e481, 2021. ISSN 24682667. doi: 10.1016/S2468-2667(21)00102-X. URL https://linkinghub.elsevier.com/retrieve/pii/S246826672100102X.

[49] G. Gorini, A. Costantini, G. Franchi, and R. Terrone. [Environmental tobacco smoke (ETS) at the workplace: considerations about a survey carried out in a pharmaceutical industry]. *Epidemiologia & Prevenzione*, 26(1):35–39, 2002. ISSN 1120-9763.

[50] A.D. Lopez, N.E. Collishaw, and T. Piha. A descriptive model of the cigarette epidemic in developed countries. *Tobacco Control*, 3(3):242–247, 1994. ISSN 0964-4563. doi: 10.1136/tc.3.3.242. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tc.3.3.242.

[51] G. Gorini, G. Carreras, E. Allara, and F. Faggiano. Decennial trends of social differences in smoking habits in Italy: a 30-year update. *Cancer Causes and Control*, 24(7):1385–1391, 2013. ISSN 0957-5243, 1573-7225. doi: 10.1007/s10552-013-0218-9. URL http://link.springer.com/10.1007/s10552-013-0218-9.

[52] A. Marcon, G. Pesce, L. Calciano, V. Bellisario, S.C. Dharmage, J. Garcia-Aymerich, T. Gislasson, J. Heinrich, M. Holm, C. Janson, D. Jarvis, B. Leynaert, M.C. Matheson, P. Pirina, C. Svanes, S. Villani, T. Zuberbier, C. Minelli, S. Accordini, and the Ageing Lungs In European Cohorts study. Trends in smoking initiation in Europe over 40 years: A retrospective cohort study. *PLOS ONE*, 13(8):e0201881, 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201881. URL https://dx.plos.org/10.1371/journal.pone.0201881.

[53] A. Alboksmaty, I.T. Agaku, S. Odani, and F.T. Filippidis. Prevalence and determinants of cigarette smoking relapse among US adult smokers: a longitudinal study. *BMJ Open*, 9(11):e031676, 2019. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2019-031676. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2019-031676.

[54] United States Public Health Service Office et al. Patterns of smoking cessation among us adults, young adults, and youth. In *Smoking Cessation: A Report of the Surgeon General [Internet]*. US Department of Health and Human Services, 2020.

[55] Alessandra Lugo, Rosario Asciutto, Roberta Pacifici, Paolo Colombo, Carlo la Vecchia, and Silvano gallus. Smoking in italy 2013-2014, with a focus on the young. *Tumori journal*, 2013.

[56] A. Puy, P. Beneventano, S.A. Levin, S. Lo Piano, T. Portaluri, and A. Saltelli. Models with higher effective dimensions tend to produce more uncertain estimates. *Science Advances*, 8(42):eabn9450, 2022. ISSN 2375-2548. doi: 10.1126/sciadv.abn9450. URL https://www.science.org/doi/10.1126/sciadv.abn9450.

[57] M.L Bongers, D. De Ruysscher, C. Oberije, P. Lambin, C.A. Uyl–de Groot, and V.M.H. Coupé. Multistate Statistical Modeling: A Tool to Build a Lung Cancer Microsimulation Model That Includes Parameter Uncertainty and Patient Heterogeneity. *Medical Decision Making*, 36(1):86–100, 2016. ISSN 0272-989X, 1552-681X. doi: 10.1177/0272989X15574500. URL http://journals.sagepub.com/doi/10.1177/0272989X15574500.

[58] S.A. Chrysanthopoulou. MILC: A Microsimulation Model of the Natural History of Lung Cancer. *International Journal of Microsimulation*, 10(3):5–26, 2016. doi: 10.34196/ijm.00164. URL http://www.microsimulation.org/IJM/V10_3/IJM_2017_10_3_1.pdf.

CHAPTER 2

Frequentist and Bayesian inference on compartmental models in epidemiology: A critical review with a focus on likelihood-free approaches

Paper submitted to *Statistics in Medicine*, under review

Cecilia Viscardi[1,2], Alessio Lachi[1,3], Michela Baccini[1,2,]

[1] Department of Statistics, Computer Science, and Applications “Giuseppe Parenti" (DiSIA), University of Florence, Viale Giovan Battista Morgagni 59-65 50134 - Florence, Italy

[2] Florence Center for Data Science, University of Florence, Viale Giovanni Battista Morgagni 59-65 50134 - Florence, Italy

[3] Epidemiology and Health Research, Institute of Clinical Physiology of the Italian National Institute Research Council (IFC-CNR), Via Giuseppe Moruzzi 1 56124 - Pisa

Abstract

Compartmental models have emerged as useful tools in various scientific domains, from epidemiology to pharmacokinetics and engineering. Due to their mechanistic nature, they provide insights into complex dynamic systems and allow predictions under different scenarios. In the last few years, they experimented with a vast spreading due to the increasing interest in modeling epidemic dynamics. However, despite their widespread use, there is still a gap in the literature, concerning their statistical formalization and a systematic discussion of the statistical methods suitable for both tasks of inference and forecasting. This work aims to fill the gap between statistical literature and practical applications. Starting from the fundamental distinction between deterministic compartmental models and stochastic compartmental models, we delve into the various challenges encountered in formulating and evaluating the likelihood function associated with the stochastic model. We distinguish two reasons for the intractability of the likelihood

function, the high dimension of missing data and the complexity of the model structure, to discuss suitable methods for addressing the problem both from a frequentist and Bayesian perspective. We overview likelihood-based methods and explore the use of likelihood-free approaches in this framework, focusing on Approximate Bayesian computation algorithms and a method that combines model calibration with a parametric bootstrap procedure. We showcase their feasibility and reliability through a toy example of the Susceptible-Infected-Removed model on simulated data. Finally, we explore the relevance of likelihood-free methods in a real-world framework through an example of a complex compartmental model developed to study smoking dynamics in the Tuscany region (Italy).

Keywords - Approximate Bayesian Computation, Bayesian inference, Bootstrap, Calibration, Compartmental models, Complex model, Frequentist inference, Incomplete data, Likelihood-free inference

2.1 Introduction

Compartmental models are a class of models used to understand and describe the dynamic evolution of a phenomenon of interest in a population. Due to their simple mechanistic nature, they are widely used for modeling infectious diseases. In particular, in the last few years, compartmental models experimented with a vast spreading due to the increasing interest in epidemiological analyses of the SARS-CoV-2 epidemic [1]. However, compartmental models are also useful and widely employed in other fields such as engineering, pharmacology, and the study of social phenomena [2–4].

Compartmental models assume that, at each point in time, the population is divided into non-overlapping groups, called compartments, which are homogeneous concerning some specific characteristics of the individuals in the population – e.g. the health status. Starting from an initial population, the transitions between compartments are allowed and described by a system of ordinary differential equations (ODEs) that define the evolution of the size of each compartment over time. These ODEs are governed by a set of model parameters that tune the transition rates.

Compartmental models are very flexible and can be employed in a forward perspective, to simulate dynamics under different scenarios defined by specific values of transition parameters, or to understand and predict the evolution of a phenomenon by estimating the model parameters based on observed dynamics. Here we will refer to this second case as the backward procedure. However, in both cases, a realistic description of a complex phenomenon cannot rely only on the mathematical rules defined by the system of ODEs but must take into account the uncertainty and variability that characterize reality. The widespread use of mathematical models in epidemiology, often without knowledge of their limitations, has the drawback of leading to

misleading conclusions. It follows that a stochastic extension is essential, especially to predict future scenarios more reliably by providing uncertainty quantification. This extension is based on the formulation of a statistical model, beyond the mathematical model, in which each transition of an individual from a given compartment to another one follows probabilistic rules. The likelihood function associated with the model depends on these rules, as well as on the other model assumptions.

Unfortunately, compartmental models often exhibit an intractable likelihood function. In some cases, it can be complex to specify an analytical form for the likelihood, due to the complexity of the model – e.g., due to the large number of compartments or to challenging definitions of the allowed transitions and related probabilistic rules. In other cases an analytical form of the likelihood is available but its point-wise evaluation is infeasible due to the presence of a large number of unobserved latent variables. In both cases, the estimation of the model parameters requires suitable methods.

In the literature, the existing reviews on the methods to make inferences on compartmental models focus only on very specific approaches. For example McKinley et al. [5] discusses only the most popular Bayesian algorithms used to provide parameter estimates in various epidemiological applications. Tang et al. [6] provides a comprehensive review of frequentist and Bayesian methods, but the considered methods are suitable only for precise mathematical/statistical models and they cannot be compared to each other. Thus, there is still a lack of a critical and comprehensive review.

This paper is aimed at describing and discussing different frequentist and Bayesian estimation strategies in compartmental models, and at providing comparisons among them. Particular attention is paid to the distinction between mathematical and statistical compartmental models and how introducing stochastic components leads to the definition of a likelihood function associated with the model. We describe some of the reasons that make the likelihood intractable and several methods to address such intractability. Finally, for each method, we discuss adequate strategies for the quantification of the uncertainty around point estimates. A special focus is given on the reliability and flexibility of likelihood-free methods in this framework. In particular, we deal with Approximate Bayesian Computation algorithms and a method that combines model calibration with a parametric bootstrap procedure.

The paper is organized as follows: in Section 2.2 we provide a general description of compartmental models and the relation between mathematical and statistical models. In Section 2.2.2 a focus on the Susceptible-Infected-Removed (SIR) model is given. The SIR model will be used as an illustrative example in the rest of the paper. Section 2.3.1 and 2.3.2 describe frequentist and Bayesian methods for conducting inference in three different situations: the case in which the likelihood function is tractable and complete data are available; the case in which the likelihood would be tractable but missing data makes its evaluation infeasible; the case in which

the likelihood is intractable. The results of the SIR example are reported in Section 2.4. Finally, as an example of a model whose likelihood is unavailable, we consider a complex compartmental model designed to describe the evolution over time of smoking dynamics in the population of the Tuscany region (Italy). Section 2.5 discusses and compares the results from all the considered methods, providing final remarks and conclusions.

2.2 Compartmental models

Compartmental models describe the evolution of a phenomenon in a population over time. At each point in time, the population is divided into compartments – i.e. groups of individuals homogeneous concerning some characteristics, such as health status. Starting from an initial condition, individuals can change their status and transit from a given compartment to another one. It follows that the size of each compartment changes over time. Compartmental models formalize a dynamic system relying on a system of ordinary differential equations (ODEs). This simple mathematical model describes the trajectory of the size of each compartment over time as a function of a set of parameters that govern the transition rates. The system of ODEs, for computational reasons, is often transformed into a system of difference equations defined in discrete time [7].

2.2.1 From mathematical to statistical models

Mathematical models describe reality in a deterministic way, however realistic modeling requires a stochastic extension. Statistical models enable us to account for sampling variability and to quantify uncertainty both in the estimation and prediction phases. They integrate a systematic component – i.e. the mathematical model – and a stochastic component. This latter is introduced by establishing that the number of transitions that occur between two compartments are realizations of random variables that follow specific probability distributions.

Let us denote by $X(t) = \big(X_1,(t),...,X_c(t),...,X_C(t)\big)$ the state of a system made of C compartments, where $X_c(t)$ denotes the size of the c-th compartment at time t. A mathematical model defines a function $f(\cdot;\theta)$ that expresses the change of the size of each compartment as a function of a set of parameters θ. In particular, the mathematical model can be specified by ODEs in continuous time: $\frac{d}{dt}X(t) = f(X(t);\theta)$.

Given a vector of parameters, θ, that describes a specific scenario, the solution of the system of ODEs, intended as the dynamic that satisfies the ODEs, provides the evolution of the compartment sizes in a forward perspective. The system of ODEs is often difficult to solve analytically and involves a lot of calculations. A practical solution is given by Euler's method [7]. This method

considers a system of equations defined in discrete time, meaning that t assumes values in a subset of $\mathbb{I}$, the set of integer numbers. When considering discrete time, the mathematical model specifies the size of each compartment at time t as a function of the sizes at the previous point in time: $X(t) = X(t-\delta) + \Delta X(t)$, where $\Delta X(t) = f(X(t-\delta);\theta)$ denotes the variations caused by the transitions between compartments that occur in the time interval $[t-\delta, t)$. δ is an integer number that represents a time increment and is often assumed to be equal to 1. The smaller δ corresponds to a better approximation from a continuous to a discrete process.

Mathematical models can be used also in a backward perspective, to understand the dynamic underlying the observed phenomenon by learning the transition parameters. In such a case, optimization strategies are implemented to find the optimal value θ^*, i.e. the value that minimizes a distance function between the observed data and the trajectories described by the solution of the system of equations with $\theta = \theta^*$.

Both from a forward and a backward perspective, mathematical models do not account for the occurrence of whatever deviation from the solution of the system. They are not intended to capture the randomness of a phenomenon despite, at least at an individual level, they are inherently stochastic, meaning that each individual can experiment (or not) a transition with a certain probability at each point in time. Kurtz [8, 9] proved that the solution of a well-defined mathematical compartmental model is the infinite population limit of a stochastic system. However, when the size of the population is finite, variability must be modeled and evaluated to come up with conscious analyses and avoid misleading conclusions. Statistical models are the proper tools to combine the mathematical function with a stochastic component. They define a stochastic function that expresses the evolution of the size of each compartment as a function of the parameters θ and a random noise. Again, from a forward perspective, the statistical model can be employed to simulate the dynamic corresponding to a scenario described by the vector of parameters θ. However, the model will simulate different dynamics even considering the same vector θ, since they are the output of a stochastic generative process that involves a random noise. The probability of observing a specific dynamic $\mathbf{x} = \{x(0), ..., x(T)\}$ depends on the formulation of the statistical model – i.e. how the random components are integrated into the model. The likelihood function, $L(\theta \mid \mathbf{x})$, comes from the probability mass/density functions of $\mathbf{X} = \{X(0), ..., X(T)\}$ evaluated at $\mathbf{x}$ and viewed as a function of the parameters. It quantifies how likely the scenario described by θ is, in the light of the observed data. The likelihood function plays a key role in the inference process but, in compartmental models, it often results in being intractable.

Let us consider a simple way to come up with an easy-to-handle statistical model: the introduction of additive random noises. Denoting by $\hat{\mathbf{x}}(\theta) = \{\hat{x}(0;\theta), ..., \hat{x}(T;\theta)\}$ the solution of the system of equations evaluated at $t \in \{0, ..., T\}$, we have that $X(t) = \hat{x}(t;\theta) + E(t)$ is a random variable whose randomness comes from $E(t)$, the vector of random perturbation noises at time t. Observed

data are $x(t) = \hat{x}(t;\theta) + \varepsilon(t)$, where $\varepsilon(t)$ denotes a realization of the random vector $E(t)$, for each t. When assuming Gaussian errors $E(t) \sim MVN(\mu = \mathbf{0}, \Sigma)$ – where μ and Σ are a vector and a matrix of sizes C and $C \times C$, respectively – random variables $X(t)$ have MVN distributions with mean $\hat{x}(t;\theta)$ and variance Σ. Under the assumption of independence among the errors over time, the likelihood function is:

$$L(\theta \mid \mathbf{x}) = \prod_{t=0}^{T} \frac{1}{(2\pi)^{C/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x(t) - \hat{x}(t;\theta))'\Sigma^{-1}(x(t) - \hat{x}(t;\theta))\Big). \tag{2.1}$$

Even in this simple case, the point-wise evaluation of the likelihood is based on the computation of $\hat{\mathbf{x}}(\theta)$ that may require a numerical solution of the system of equations. This often makes the inferential process computationally demanding, because both frequentist and Bayesian methods require several evaluations of the likelihood function. This problem becomes even more serious when the random noise depends in turn on the parameter vector θ. For example, this happens when the waiting time (continuous or discrete) between events – i.e. transitions between compartments – is assumed to follow an exponential or geometric distribution [10] and the number of transitions at each point in time are Poisson or Binomial random variables. In such cases, compartmental models are often associated with an intractable likelihood function. These situations will be detailed in the next section resorting to a working example.

2.2.2 Working example: the SIR model

The definition and the evaluation of the likelihood function are of paramount importance in the inferential process, both from a Bayesian and a frequentist point of view. Unfortunately, in many compartmental models, a point-wise evaluation of the likelihood function is prohibitive. This may happen at least in two cases: when the model is so complex that it makes it infeasible to write down an analytical form for the associated likelihood function; or when the analytical form is available but its evaluation is computationally demanding. In what follows we will describe the formulation of the likelihood function and the reasons that make its evaluation infeasible using a simple working example. An example addressing a much more complex model is deferred to Section 2.4.2.

Let us consider the well-known Susceptible-Infected-Removed (SIR) model [11–13]. The SIR model subdivides the population into three compartments: susceptible individuals (S) are those who can potentially contract the disease when they come into contact with an infectious individual since they are not immunized; infectious individuals (I) are those who are currently infected and infective, thus they can transmit the disease to susceptible individuals; removed individuals (R)

have been infected and have either recovered from the disease or have died. In this model, once an individual recovers, she/he is assumed to have immunity to the disease.

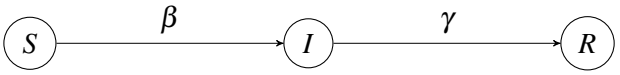


FIGURE 2.1: Graphical representation of the SIR model. Each node represents a compartment and edges indicate the allowed transitions between compartments.

Figure 2.1 shows the transitions allowed by the model and the parameters governing the transition rates. We denote by γ the resolution rate defined as $1/\tau$, where τ is the average time spent by an infected individual in the compartment I. The instantaneous rate of transmission of the infection is denoted by β. From β and γ, we can compute the basic reproduction number – i.e. the expected number of secondary infections caused by a single infected individual at the beginning of the epidemic – as $R_0 = \beta/\gamma$. Here, we denote by $\theta = (\tau, R_0)$ the vector of parameters to be inferred, and by $S(t)$, $I(t)$ and $R(t)$ the sizes of the three compartments at time t.

The following system of differential equations describes the dynamic of the SIR model:

$$\begin{cases} \frac{dS(t)}{dt} = & -\beta \frac{I(t)}{S(0)} S(t) \\ \frac{dI(t)}{dt} = & \beta \frac{I(t)}{S(0)} S(t) - \gamma I(t) \\ \frac{dR(t)}{dt} = & \gamma I(t). \end{cases} \tag{2.2}$$

Note that in Equation (2.2), $\beta \frac{I(t)}{S(0)}$ is the infection rate at time t. It depends on $\frac{I(t)}{S(0)}$, which is the fraction of infectious individuals with whom a susceptible individual can come into contact at time t, and on β, which represents the instantaneous rate at which an infectious individual infects a susceptible one. The system of ODEs in Equation (2.2), may be replaced by the following system of equations in discrete time:

$$\begin{cases} S(t) = & S(t-1) - \pi_{\mathrm{SI}}(t-1) S(t-1) \\ I(t) = & I(t-1) + \pi_{\mathrm{SI}}(t-1) S(t-1) - \pi_{\mathrm{IR}} I(t-1) \\ R(t) = & R(t-1) + \pi_{\mathrm{IR}} I(t-1), \end{cases} \tag{2.3}$$

where $\pi_{\mathrm{SI}}(t-1) = 1 - \exp\big(-\beta \frac{I(t-1)}{S(0)}\big)$ and $\pi_{\mathrm{IR}} = 1 - \exp(-\gamma)$ are the probability of being infected and the probability of recovering or dying during a unit time interval, respectively. These probabilities come from the assumption that the waiting time before experimenting with an event (infection or recovery/death) has an exponential distribution. In particular, π_{SI} and π_{IR} are the

probabilities of waiting a time smaller than 1, i.e. the probability of exiting the compartment between t and $t+1$.

A possible way to introduce stochasticity in the SIR model is assuming two independent Binomial distributions for the random variables that count new infections and resolutions, say $i(t)$ and $r(t)$: $i(t) \sim \text{Binom}[S(t), \pi_{\text{SI}}(t)]$ and $r(t) \sim \text{Binom}[R(t), \pi_{\text{IR}}]$ for each $t \in \{0, ..., T\}$ (Figure 2.2). Note that, as shown by Allen and Burgin [14], the solution of the mathematical model can be seen as the expected value of the statistical one. This is apparent looking at Equation (2.3), where the number of new infections and recoveries, at each time t, resemble expected values of two Binomial distributions.

Let us denote by $i_{t_1:t_2} = \big(i(t_1), ..., i(t_2)\big)$ and $r_{t_1:t_2} = \big(r(t_1), ..., r(t_2)\big)$ the random vectors of new infections and new resolutions from t_1 until t_2 (t_1 and $t_2 \in \{0, ...T\}$, $t_2 \geq t_1$), and by $i^*_{t_1:t_2}$ and $r^*_{t_1:t_2}$ their realization. Under the binomial assumption, the likelihood function is:

$$\begin{aligned} L\big(\theta \mid i_{1:T} = i^*_{1:T}, r_{1:T} = r^*_{1:T}\big) &= \prod_{t=1}^{T} \Pr\big(i(t) = i^*(t) \mid i^*_{0:t-1}, r^*_{0:t-1}\big) \Pr\big(r(t) = r^*(t) \mid i^*_{0:t-1}, r^*_{0:t-1}\big) \\ &= \prod_{t=1}^{T} \binom{S(t)}{i^*(t)} \pi_{\text{SI}}(t)^{i^*(t)} [1 - \pi_{\text{SI}}(t)]^{S(t) - i^*(t)} \binom{I(t)}{r^*(t)} \pi_{\text{IR}}^{r^*(t)} [1 - \pi_{\text{IR}}]^{I(t) - r^*(t)}, \end{aligned} \tag{2.4}$$

where $S(t) = S(t-1) - i^*(t-1)$ and $I(t) = I(t-1) + i^*(t-1) - r^*(t-1)$ and the initial condition of the system is assumed to be: $S(0) = N - 1$, with N equal to the size of the population; $I(0) = 1$; $R(0) = 0$. This likelihood is analytically tractable, but different distributional assumptions about the process may complicate its form.

Furthermore, the model described so far is based on several strong assumptions:

- the population is closed to births and deaths (except those due to the studied infectious disease), to immigration and emigration, thus $S(t) + I(t) + R(t) = N$ for each $t \in \{0, ..., T\}$;
- the population is homogeneously mixed;
- all the individuals in the same compartment at the same time t have the same risk of leaving the compartment, regardless of the time already spent in it;
- individuals are infected from the onset of the infection until recovery or death;
- the reinfection rate is equal to 0;
- both the instantaneous infection rate and the resolution rate are constant over time.

To relax the above-mentioned assumptions, in the literature have been proposed several strategies, such as the introduction of further compartments to take into account the incubation period [15] or the availability of vaccines [16]. Some of them relax the homogeneity assumption by assuming that the spreading process of infectious disease occurs over a network structure. Some models relax the assumption of constant transition rates by introducing some dependencies from the calendar time, or other variables (e.g. age) [17, 18]. Most of these extensions complicate the likelihood function in Equation (2.4) making it intractable.

However, even very simple compartmental models may be associated with an intractable likelihood function. This is often due to a problem of missing data. A typical example is when, in the SIR model, we observe only the daily number of new infections, $i^*_{1:T}$. Note that this is a quite realistic situation that occurs when there is no notification of recovery. Indeed, in this framework, the vector $r^*_{1:T}$ is missing, $r_{1:T}$ represents a latent variable, and the evaluation of the probability of the data requires marginalization over it:

$$\begin{aligned} L\big(\theta \mid i_{1:T} = i^*_{1:T}\big) &= \sum_{r^*_{1:T} \in \mathscr{R}} \Pr\big(i_{1:T} = i^*_{1:T}, r_{1:T} = r^*_{1:T} \mid \theta\big) \\ &= \sum_{r^*_{1:T} \in \mathscr{R}} \prod_{t=1}^{T} \binom{S(t)}{i^*(t)} \pi_{\mathrm{SI}}(t)^{i^*(t)} [1 - \pi_{\mathrm{SI}}(t)]^{S(t) - i^*(t)} \binom{I(t)}{r^*(t)} \pi_{\mathrm{IR}}^{r^*(t)} [1 - \pi_{\mathrm{IR}}]^{I(t) - r^*(t)}, \end{aligned} \tag{2.5}$$

where $\mathscr{R}$ is the subset of $\mathbb{I}^T$ that contains all possible sequences $r^*_{1:T}$ that are compatible with the observed series of infections. The structure and the cardinality of $\mathscr{R}$ often make the point-wise evaluation of the likelihood and likelihood-based methods computationally intensive.

To summarize, the main difficulties one can come across are two:

- unavailability of the analytical expression of the likelihood function, due to the complex structure of the model;
- intractability of the likelihood function due to the presence of high dimensional latent variables, despite the availability of an analytical form.

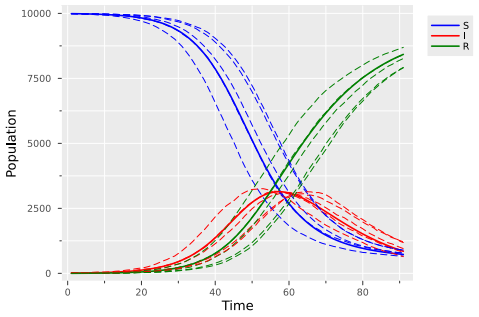


FIGURE 2.2: Evolution of compartment sizes in a mathematical SIR model (continuous line) and in four realizations from a statistical SIR model (dashed line) for $T = 90$ days, setting $\theta = [\tau = 14, R_0 = 3]$, $S(0) = 9,990$, $I(0) = 10$, and $R(0) = 0$.

2.3 Estimation methods

In this section, we review some of the principal frequentist and Bayesian methods for inferring the parameters governing statistical compartmental models defined in discrete time. We describe a suitable method for each of the following situations:

- analytically available likelihood function, with complete data;
- analytically available likelihood function, with incomplete data;
- intractable likelihood due to a too complex model.

TABLE 2.1: Cases and methods addressed in the present work.

	Likelihood	**Frequentist**	**Bayesian**
Complete data	Available, tractable	MLE	MC
Incomplete data	Available, intractable	EM	DA-MCMC
Complex model	Unavailable	Calibration	ABC

Specifically, we focus on the methods reported in Table 2.1. It is worth noting from the beginning that likelihood-free methods, such as calibration and Approximate Bayesian Computation (ABC), can be used also when the likelihood function is available, while they are the only option when the likelihood is unavailable. In Section 2.4.1 we compare the results provided by likelihood-based methods with those provided by likelihood-free methods at work on the SIR example with incomplete data.

2.3.1 Frequentist approaches

From a frequentist point of view, we are interested in providing point estimates and confidence intervals around the estimates, to account for sampling variability. Depending on the availability of the likelihood function and complete data, we suggest three different solutions to come up with point estimates. As regards confidence intervals, we describe a bootstrap procedure suitable for all the considered methods.

2.3.1.1 Maximum likelihood estimation

When the likelihood function has a tractable analytical form as that in Equation (2.4) and complete data are available, one can infer the parameters governing transition rates via numerical maximization of the likelihood function using algorithms such as those described by Nelder and Mead [19]. In the literature, there are several works dealing with the asymptotic theory of these estimators in compartmental models. In some of them, the asymptotic behavior is achieved by considering an increasing number of observations in a given time window – this is possible when working in continuous time – while others consider an increasing time window with $T \to \infty$. We do not deepen into the technicalities of the properties of such estimators and refer the readers to the review in [6]. This review, in discussing different Maximum likelihood (ML) estimators, also reports methods for the estimation of their variance, resorting to composite likelihood strategies or martingale methods. Here, we suggest to implement a flexible bootstrap approach to compute confidence intervals in all the considered cases.

2.3.1.2 Expectation-Maximization algorithm

The Expectation Maximization (EM) algorithm [20] is a computational method for finding ML estimates when the likelihood function is available but intractable, and imputing hidden/missing data simplifies its evaluation. This is the case of simple compartmental models that exhibit a tractable likelihood function when complete data are available, but the incompleteness of the data makes its evaluation infeasible.

Let us denote by $\mathbf{x}$ the observed data, by $\mathbf{z}$ some missing data, and by $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ the complete data. The evaluation of the complete likelihood, $L(\theta \mid \mathbf{y})$, is straightforward but the evaluation of the incomplete likelihood, $L(\theta \mid \mathbf{x})$, is not. The EM algorithm iterates two steps:

- the E-step computes the expected value of the complete log-likelihood, $\ell(\theta \mid \mathbf{x}, \mathbf{z})$, w.r.t. missing variables $\mathbf{Z}$;
- the M-step maximizes this expected value w.r.t. to θ.

When the expected value is difficult to evaluate, its analytical evaluation can be replaced by a Monte Carlo estimate based on a sample of size m from the distribution of the latent variables $p(\mathbf{z} \mid \mathbf{x}, \theta^{(s-1)})$, as proposed by Levine and Casella [21].

Algorithm 1 Expectation Maximization

1: **Inizialize** $\theta^{(0)}$ as random starting value
2: **Set** e
3: **for** s in $1 : S$ **do**
4: **Assign** $\theta^{(s)} = \arg\max_{\theta} E_{Z \sim p(\mathbf{z}|\mathbf{x},\theta^{(s-1)})}[\ell(\theta|\mathbf{x},\mathbf{z})] \approx \arg\max_{\theta} \frac{1}{m} \sum_{i=1}^{m} \ell(\theta|\mathbf{x},\mathbf{z}^{(i)})$
5: **if** $\frac{\theta^{(s)}-\theta^{(s-1)}}{\theta^{(s-1)}} \leq e$ **then**
6: **Break**
7: **end if**
8: **end for**

Algorithm 1 summarizes the EM algorithm implemented in our working example where missing data $\mathbf{z}$ correspond to the series $r^*_{1:T}$, and observed data $\mathbf{x}$ are $i^*_{1:T}$ (see Section 2.4.1).

The main problem of the EM algorithm is that it is highly dependent on the starting points and does not ensure convergence when the expected value of the log-likelihood is too complex to optimize.

2.3.1.3 Calibration

In some cases, compartmental models are so complex that there is no analytical form for the associated likelihood function. In such cases, a possible solution to provide point estimates is the calibration procedure [22]. This method consists of searching for the optimal parameter values that lead to an evolution of the system as close as possible to the observed one. The procedure requires only the availability of a mathematical model that allows for producing forward simulations of the compartment sizes.

Let us denote by $\mathbf{x} = \big(x(0), ..., x(T)\big)$ the observed dynamic of the system, and by $\hat{\mathbf{x}}(\theta) = \big(\hat{x}(0;\theta), ..., \hat{x}(T;\theta)\big)$ the dynamic simulated by the mathematical model when it takes the vector of parameter θ as an input. Given a discrepancy function $\rho(\cdot,\cdot)$, the calibration procedure optimizes the objective function, i.e. minimizes over θ the discrepancy between observed and simulated data:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \rho\big(\mathbf{x}, \hat{\mathbf{x}}(\theta)\big). \tag{2.6}$$

The optimization is performed numerically and the results achieved by minimization algorithms often depend on the values at which they are initialized. To avoid the problem of getting stuck in local minima, we select the initial values through a preliminary optimization over a multidimensional grid [22, 23].

This method is very flexible since it completely disregards the likelihood formulation and relies only on the mathematical model. This fact enables its use regardless of the reason why the likelihood is challenging to handle (e.g. complex model or high dimensional latent variables).

2.3.1.4 Bootstrap procedure

In all of the three cases considered above, we can quantify sampling variability around the point estimate of θ and provide confidence intervals, by using a bootstrap procedure [24]. Specifically, we resort to a parametric bootstrap as proposed by [22]. This choice makes it possible to account for sampling variability according to the distributional assumptions underlying the likelihood function associated with the statistical model in two steps:

For $b \in \{1, ..., B\}$:

1. sample one dynamic $\mathbf{x}^b$ from the stochastic model with input $\hat{\theta}$, i.e. the MLE or the optimal parameters retrieved via calibration;
2. obtain an estimate $\hat{\theta}^b$ of θ, by alternatively:
   a) computing the ML estimate, using $\mathbf{x}^b$ as observed data;
   b) implementing the EM algorithm, using as observed data only the dynamics that are not missing, as derived from $\mathbf{x}^b$;
   c) calibrating the model searching for $\hat{\theta}^b$ minimizing the discrepancy function between simulated and estimated dynamics $\rho\left(\mathbf{x}^b, \hat{\mathbf{x}}(\theta)\right)$.

We use the percentiles of the bootstrap samples $\hat{\theta}^1, ..., \hat{\theta}^B$ to compute confidence intervals.

2.3.2 Bayesian approaches

In the statistical formulation of a compartmental model, observed data $\mathbf{x}$ are realizations of a random variable $\mathbf{X}$. In Bayesian statistics, the set of parameters that governs the probability distribution of $\mathbf{X}$ is, in turn, modeled as a random variable $\theta \in \Theta$ with a prior distribution, here denoted by $\pi(\cdot)$. Thus, given the observed data $\mathbf{x}$, the object of interest for the inference is the posterior distribution derived through Bayes's formula:

$$\pi(\theta \mid \mathbf{x}) = \frac{\pi(\theta)L(\theta \mid \mathbf{x})}{\int\limits_{\Theta} \pi(\theta)L(\theta \mid \mathbf{x})d\theta}, \tag{2.7}$$

where the denominator is the marginal likelihood, which is a normalizing constant. Often, the computation of this normalizing constant is infeasible and requires a numerical approximation.

When the model involves high dimensional latent variables, $\mathbf{Z}$, they should be integrated out to derive the posterior distribution:

$$\pi(\theta \mid \mathbf{x}) = \frac{\pi(\theta) \int\limits_{\mathscr{Z}} L(\theta \mid \mathbf{x}, \mathbf{z})d\mathbf{z}}{\int\limits_{\Theta} \int\limits_{\mathscr{Z}} \pi(\theta)L(\theta \mid \mathbf{x}, \mathbf{z})dz\,d\theta}, \tag{2.8}$$

This is also the case of the SIR model with incomplete data, which requires the solution of several high-dimensional summations, as those in Equation (2.5). In the literature, several methods for addressing these problems and conducting Bayesian inference via simulations are available (see [5] for a comprehensive discussion of the use of these methods in the epidemiological framework).

2.3.2.1 Monte Carlo methods

When the likelihood function is tractable and we observe complete data, the only hurdle is the computation of the normalizing constant in Equation (2.7). A possible strategy to overcome this problem is resorting to Monte Carlo methods, a class of algorithms aimed at solving inferential or optimization problems through stochastic simulations [25]. In particular, here we consider Importance Sampling (IS) [26], although other solutions are available (e.g., Accept-Reject and Markov Chain Monte Carlo methods) – see [27] for a survey of these methods. The main idea of this algorithm is to get samples of the model parameter from an easy-to-sample distribution, and then convert them into a sample from the posterior distribution by assigning an importance weight to each parameter proposal. The algorithm is summarized in Algorithm 2, where $q(\cdot)$ denotes the easy-to-sample proposal distribution.

Algorithm 2 Importance Sampling

1: **Draw** $\theta^{(1:S)}$ i.i.d. from $q(\cdot)$

2: **Assign** to each $\theta^{(s)}$ an importance weight $\omega^{(s)} = \frac{\pi(\theta^{(s)})L(\theta^{(s)}|\mathbf{x})}{q(\theta^{(s)})}$

3: **Compute** normalised weights $\tilde{\omega}^{(s)} = \frac{\omega^{(s)}}{\sum\limits_{i=1}^{S} \omega^{(i)}}$

The output is a weighted sample $(\theta^{(1)}, \tilde{\omega}^{(1)}), ..., (\theta^{(S)}, \tilde{\omega}^{(S)})$ drawn as a single batch. It can be used to estimate posterior quantities through weighted averages or by introducing a resampling step that uses normalized weights as probabilities. The final sample can be considered as an i.i.d. sample from the exact posterior distribution but the variability of the posterior estimates depends on the variability of the importance weights. This latter depends in turn on the choice of the proposal distribution that should be as close as possible to the target.

2.3.2.2 Data Augmentation Markov Chain Monte Carlo methods

MC methods and Markov Chain Monte Carlo (MCMC) methods usually rely on point-wise evaluations of the likelihood function. In the presence of missing data, the evaluation of the likelihood function also requires the solution of high dimensional integrals/summations, as shown in Equation (2.8). To avoid their computation, a possible solution is to provide a sample from a posterior distribution defined on the augmented space $\Theta \times \mathscr{Z}$:$\pi(\theta, \mathbf{z} \mid \mathbf{x})$.

To this aim, Data Augmentation Markov Chain Monte Carlo methods (DA-MCMC) can be implemented [28, 29]. Usually, these algorithms rely on a Gibbs sampling scheme and require the ability to get samples from the full conditional distributions $\pi(\theta \mid \mathbf{x}, \mathbf{z})$ and $p(\mathbf{z} \mid \mathbf{x}, \theta)$, or the collapsed distribution $p(\mathbf{z} \mid \mathbf{x})$. Sometimes, in compartmental models, their definition is not straightforward. Different strategies may be the implementation of a Metropolis-within-Gibbs algorithm, in which samples from unavailable full conditional distributions are obtained through Metropolis steps in the Gibbs sampling scheme, or the implementation of a Metropolis-Hastings (MH) algorithm to get samples directly on the joint space $\Theta \times \mathscr{Z}$, as displayed in Algorithm 3 (see [30] for details on the above mentioned MCMC algorithms).

Algorithm 3 Metropolis-Hastings

Inizialize $\theta^{(0)}, \mathbf{z}^{(0)}$
for s in $1 : S$ **do**
  Propose missing data $\mathbf{z}^* \sim q_z(\cdot \mid \mathbf{z}^{(s-1)})$
  Propose $\theta^* \sim q_\theta(\cdot \mid \theta^{(s-1)})$
  Compute $\alpha = \min\left\{1, \frac{\pi(\theta^*)L(\theta^* \mid \mathbf{z}^*, \mathbf{x})q_z(\mathbf{z}^{(s-1)} \mid \mathbf{z}^*)q_\theta(\theta^{(s-1)} \mid \theta^*)}{\pi(\theta^{(s-1)})L(\theta^{(s-1)} \mid \mathbf{z}^{(s-1)}, \mathbf{x})q_z(\mathbf{z}^* \mid \mathbf{z}^{(s-1)})q_\theta(\theta^* \mid \theta^{(s-1)})}\right\}$
  Sample $u \sim U(0,1)$
  if $u < \alpha$ **then**
    Set $\theta^{(s)} = \theta^*$ and $\mathbf{z}^{(s)} = \mathbf{z}^*$
  else
    set $\theta^{(s)} = \theta^{(s-1)}$ and $\mathbf{z}^{(s)} = \mathbf{z}^{(s-1)}$
  end if
end for

The output of the algorithm is a realization of a Markov chain, $(\boldsymbol{\theta}^{(0)}, \mathbf{z}^{(0)}), ..., (\boldsymbol{\theta}^{(S)}, \mathbf{z}^{(S)})$. The algorithm satisfies the detailed balance condition (see Section B.1 in Supplemental Materials) thus the limiting distribution of the chain is $\pi(\boldsymbol{\theta}, \mathbf{z} \mid \mathbf{y})$. This means that, as always in MCMC methods, samples are only asymptotically distributed according to the joint posterior, and checks for the convergence of the chain are needed.

The sampling scheme in Algorithm 3 is very general as it requires only the ability to evaluate the prior probability and the complete likelihood, and samples from the target distribution $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ can be easily retrieved by disregarding the sequence $\mathbf{z}^{(0)}, ..., \mathbf{z}^{(S)}$ from the final output. Problems may be related to the autocorrelation of the chain and the choice of good proposal distributions $q_{\boldsymbol{\theta}}(\cdot)$ and $q_z(\cdot)$. To accelerate the convergence and improve the efficiency of the algorithm, many alternatives are available in the literature. Examples are algorithms that define adaptive proposal distributions [31] or the Hamiltonian version of the MH algorithm [32], which is efficient in the case of smooth density functions [33].

In the case of compartmental models, particular attention should be paid to the definition of the proposal distribution for the missing data $q_z(\cdot)$. A possible strategy is suggested in [10] (see Section 2.4). To overcome this problem, in some specific cases, one can also resort to Particle Marginal Metropolis-Hastings samplers which avoid the definition of a proposal distribution $q_z(\cdot)$ and approximate $p(\mathbf{z} \mid \mathbf{y})$ through a Sequential Monte Carlo algorithm [34].

2.3.2.3 Approximate Bayesian Computation

All the Bayesian methods described so far require at least an analytical form for the complete likelihood function. In the case of complex models, the likelihood function may be unavailable and its analytical form cannot be written down. Furthermore, even in the case of a simple model with incomplete data, when the size of the latent variables is very large, DA-MCMC is computationally demanding. In such cases, likelihood-free methods, such as Approximate Bayesian Computation (ABC), result in being convenient due to their flexibility.

ABC is a broad class of methods allowing Bayesian inference on parameters governing complex models with intractable likelihood functions. The original intuition can be traced back to an explanation of the Bayes' Theorem provided by Rubin in the 80s [35], but primal ABC algorithms have been formalized by Pritchard et al. [36], Tavaré et al. [37]. ABC methods dispense with exact likelihood computation and only require the ability to simulate pseudo-data by sampling observations from the assumed statistical model employing a computer program that reproduces the stochastic data generative mode, usually called a "simulator", here denoted by $\Pr(\cdot \mid \cdot)$. The underlying idea of ABC methods is to convert samples from the prior distribution into samples from the posterior through three simple steps:

1. draw S parameter proposals from the prior distribution $\pi(\cdot)$;
2. give each parameter $\theta^{(s)}$ as an input to the simulator to sample pseudo-data $\hat{\mathbf{x}}(\theta^{(s)})$;
3. retain only parameter proposals $\theta^{(s)}$ such that $\hat{\mathbf{x}}(\theta^{(s)}) = \mathbf{x}$.

The output is a sample from the exact posterior distribution. However, in practice, several sources of approximation are introduced by replacing the equality constraint at the third step with a "closeness" constraint: parameter proposals are retained when $\rho(\hat{\mathbf{x}}(\theta^{(s)}); \mathbf{x}) \leq e$, where $\rho(\cdot;\cdot)$ is a distance function and e is positive tolerance threshold. Note that the output becomes an i.i.d. sample from an approximate posterior distribution, the closeness of which to the true one depends on the choice of $\rho(\cdot;\cdot)$ and the magnitude of e. In the literature, several advanced sampling schemes have been proposed (we refer the reader to [38] for a comprehensive description of the method). Most of them are sequential methods based on a decreasing sequence $e_1 \geq e_2 \geq ... \geq e_S$ of thresholds, rather than a fixed tuning parameter. Usually, these algorithms rely on the definition of a sequence of tempered target distributions based on the sequence of thresholds, and get samples from each of them using an importance sampling step. Examples are the Population Monte Carlo ABC presented in [39] and some adaptive versions inspired by it (see Algorithm 4). Here, we relied on one of the strategies proposed in [40], where new thresholds are automatically selected during the execution of the algorithm in a way that ensures a decreasing level of approximation from the given iteration to the next one.

Algorithm 4 Adaptive Population Monte Carlo ABC

Initialize e_1
for j in $1:M$ **do**
  Simulate $\theta_j^{(1)} \sim \pi(\cdot)$ and $\hat{\mathbf{x}}(\theta_j^{(1)}) \sim \Pr(\cdot \mid \theta_j^{(1)})$ until $\rho(\hat{\mathbf{x}}(\theta_j^{(1)}); \mathbf{x}) < e_1$
  Set $\omega_j^{(1)} = \frac{1}{M}$
end for
Select e_2 using an adaptive strategy.
for s in $2:S$ **do**
  Set Σ_s to twice the empirical covariance matrix of $\theta_1^{(s-1)}, ..., \theta_M^{(s-1)}$
  for j in $1:M$ **do**
    Pick θ_j^* from $(\theta_1^{(s-1)}, ..., \theta_M^{(s-1)})$ with probabilities $(\omega_1^{(s-1)}, ..., \omega_M^{(s-1)})$
    Generate $\theta_j^{(s)} | \theta_j^* \sim MVN(\theta_j^*, \Sigma_s)$ and $\hat{\mathbf{x}}(\theta_j^{(s)}) \sim \Pr(\cdot \mid \theta_j^{(s)})$
    Set $\omega_j^{(s)} \propto \dfrac{\pi(\theta_j^{(s)})}{\sum_{m=1}^{M} \omega_m^{(s-1)} \phi\{\tau_s^{-1}(\theta_m^{(s)} - \theta_m^{(s-1)})\}} \mathbb{1}\{\rho(\hat{\mathbf{x}}(\theta_j^{(s)}); \mathbf{x}) < e_s\}$
      where ϕ represents the density of the Standard Normal distribution
  end for
  Select e_{s+1} using an adaptive strategy.
end for

Note that the ABC procedure is very similar in spirit to the calibration. However, ABC algorithms use the statistical model and take into account two sources of uncertainty: the uncertainty on the parameter and the one related to the sampling space.

2.4 Results

In this section, we show the results of the presented frequentist and Bayesian methods on the working example presented in Section 2.2.2. Furthermore, we present the results of likelihood-free methods in the case of a complex model whose likelihood is unavailable.

2.4.1 Working example: the SIR model

We applied all the methods described in Section 2.3 on simulated data obtained from the SIR model, after setting $T = 90$, $\tau^{\text{true}} = 14$, $R_0^{\text{true}} = 3$, $S(0) = 9{,}990$, $I(0) = 10$, and $R(0) = 0$. In particular, we simulated the two-time series $i^*_{1:T}$ and $r^*_{1:T}$, the knowledge of which is sufficient to reconstruct compartment sizes $S^*_{1:T}$, $I^*_{1:T}$ and $R^*_{1:T}$, given $S(0)$, $I(0)$ and $R(0)$. We focused on the estimate of R_0 and τ considering the initial size of the compartments as known. Regarding Bayesian inference, it has been conducted using uniform prior distributions: $\tau \sim U[7,21]$ and $R_0 \sim U(0,6]$. We considered both the case of complete and incomplete data.

2.4.1.1 Complete data

Let us consider the case in which we observe the entire evolution of the compartment sizes, equivalent to observing both $i^*_{1:T}$ and $r^*_{1:T}$. We can compute MLE for R_0 and τ and their posterior distributions via IS, respectively from a frequentist and a Bayesian point of view.

Regarding MLE, the maximization of the complete likelihood in Equation (2.4) has been performed using Nelder's procedure implemented in the `Optim` package of JULIA software [19, 41]. Then, the bootstrap procedure described in Section 2.3.1 has been implemented with $B = 1{,}000$. More precisely, to be consistent with the likelihood function in Equation (2.4), at each bootstrap iteration b, we draw samples from two Binomial distributions: $\big(S^b(t-1) - S^b(t)\big) \sim Binom[S^b(t-1), 1-\exp(-\frac{\hat{R}_0}{\hat{\tau}}\frac{I^b(t-1)}{S(0)})]$ and $\big(R^b(t) - R^b(t-1)\big) \sim Binom[R^b(t-1), 1-\exp(-\frac{1}{\hat{\tau}})]$, with $\hat{R}_0$ and $\hat{\tau}$ the maximum likelihood estimates of the model parameters. As regards the IS implementation, we used the joint prior distribution of the parameters as proposal distribution.

Table 2.2 reports the results in terms of ML estimates with 90% bootstrap intervals, and Maximum A Posteriori (MAP) estimates with 90% Highest Posterior Density (HPD) intervals. These results,

as well as Figure 2.3, show that both methods provide point estimates very close to the true parameters.

TABLE 2.2: Frequentist versus Bayesian inference on the parameters of the SIR model, in the case of complete data. Results are reported in terms of point estimate and 90% confidence interval in the frequentist case, and in terms of maximum a posteriori and 90% highest posterior density in the Bayesian case.

	Frequentist	**Bayesian**
Complete Data	**Maximum likelihood**	**Importance Sampling**
$\boldsymbol{\tau}$	14.00 (13.73 - 14.25)	13.96 (13.75 - 14.25)
$\boldsymbol{R_0}$	3.00 (2.92 - 3.08)	3.01 (2.93 - 3.08)

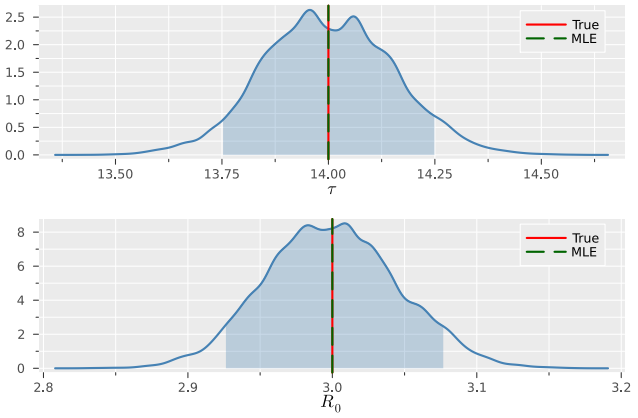


FIGURE 2.3: Marginal posterior distributions of SIR model parameters obtained through Importance Sampling, with shaded areas representing the 90% highest posterior density interval. Vertical lines indicate the true parameter value, along with its maximum likelihood estimate (MLE).

2.4.1.2 Incomplete data

Let us consider the case in which we observe only the series $i^*_{1:T}$, meaning that information about newly resolved infections is not available. We can resort to EM and DA-MCMC algorithms, respectively in the frequentist and Bayesian settings. In our implementation, computing the expected value in the E-step would involve a complex summation over $\mathscr{R}$, thus at each iteration s we use an MC estimate based on $m = 1{,}000$ simulations from $p(r_{1:T} \mid i^*_{1:T}, \theta^{(s-1)})$ (see B.2 in Supplemental Materials). As regards the DA-MCMC proposal distributions, we resorted to Gaussian multivariate proposal distributions for $\theta = (\tau, R_0)$, and to a mixture of proposal distributions for $r_{1:T}$. In particular, we sample from the mixture of the three proposal distributions as described in [10]: at each iteration s of the DA-MCMC algorithm, we select at random one of the following small perturbations:

1. add a resolution: select at random t and propose a series in which $r^{(s)}(t)$ is set to $r^{(s-1)}(t)+1$;

2. subtract a resolution: select at random t and propose a series in which $r^{(s)}(t)$ is set to $r^{(s-1)}(t)-1$;

3. move a resolution: select at random (t_1,t_2) and propose a series in which $r^{(s)}(t_1)$ is set to $r^{(s-1)}(t_1)-1$ and $r^{(s)}(t_2)$ is set to $r^{(s-1)}(t_2)+1$.

We evaluate the probability of this new proposal and use it in the computation of the MH acceptance ratio.

Note that the longer is the observed series, the higher is the cardinality of the space of the latent variables. This makes proposal distributions based on small perturbations inefficient and the chain strongly autocorrelated. In such a case, likelihood-free algorithms should be implemented to obtain more efficient posterior estimates.

This illustrative example provides the opportunity to compare the implementation of likelihood-based and likelihood-free methods in the case of incomplete data, as both of these approaches are feasible. Hence, we tested also calibration and ABC algorithms for inferring model parameters when only $i^*_{1:T}$ is observed. In both the algorithms, we compared observed and simulated trajectories through the Euclidean distance:

$$\rho(\hat{\mathbf{x}}(\theta),\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\left(i^*(t)-\hat{i}(t;\theta)\right)^2}, \tag{2.9}$$

where $\hat{i}(1;\theta),...,\hat{i}(T;\theta)$ denotes the series simulated through the mathematical or the statistical SIR model, in the calibration and ABC procedure respectively, when the vector θ is given as an input (for a description of the procedure for simulating data from the statistical model see B.2 in Supplemental Materials).

The calibration procedure is performed using the optimization strategy implemented in the JULIA package `Optim` [41]. To avoid the problem of getting stuck in local minima, we performed several optimizations using different starting points, then we selected the solution that minimized the distance function in Equation (2.9) [22, 23]. To consider sampling variability and compute confidence intervals, we implement the parametric bootstrap procedure described in Section 2.3.1. Here, at each bootstrap iteration, the optimization algorithm has been initialized at random starting values.

As regards the ABC procedure, we implemented the Algorithm 4. At each iteration, we compared the posterior distributions approximated by ABC with those provided by the DA-MCMC after the assessment of the convergence of the chain (see Section B.3 in Supplemental Materials for further details). Looking at the Kullback-Leibler divergence and the Hellinger distance between these distributions, it turned out that after 80 iterations they are quite stable.

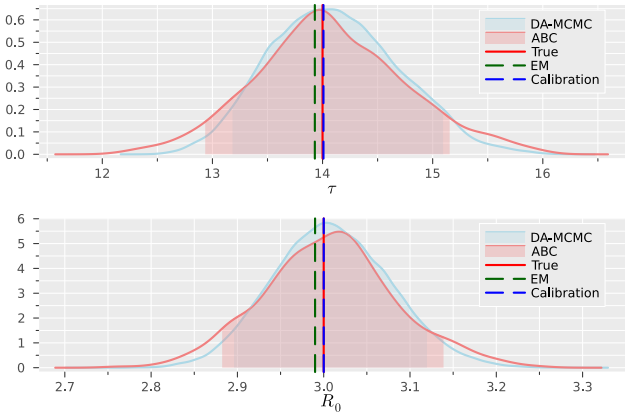


FIGURE 2.4: Marginal posterior distributions of SIR model parameters obtained through DA-MCMC and ABC, with shaded areas representing the 90% highest posterior density intervals. Vertical lines indicate the true parameter value, along with its estimates from EM and calibration methods.

TABLE 2.3: Frequentist versus Bayesian inference on the parameters of the SIR model in the case of incomplete data. Results are reported in terms of point estimate and 90% confidence interval in the frequentist case and in terms of maximum a posteriori and 90% highest posterior density intervals in the Bayesian case.

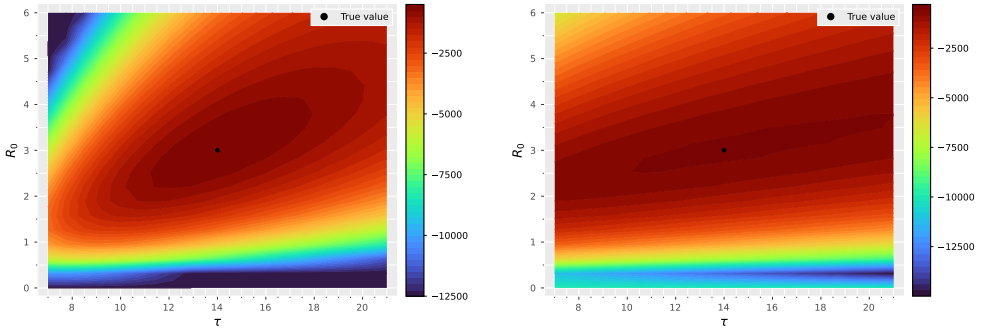
	Frequentist	**Bayesian**
Likelihood-based	**EM**	**DA-MCMC**
$\boldsymbol{\tau}$	13.93	14.08 (13.18 - 15.09)
$\boldsymbol{R_0}$	2.99	3.00 (2.90 - 3.12)
Likelihood-free	**Calibration**	**ABC**
$\boldsymbol{\tau}$	14.01 (13.61 - 14.40)	13.97 (12.93 - 15.15)
$\boldsymbol{R_0}$	3.00 (2.95 - 3.06)	3.02 (2.88 - 3.14)

Figure 2.4 and Table 2.3 show that all the results of the implemented methods are coherent with R_0^{true} and τ^{true}. From a frequentist perspective, point estimates provided by EM and calibration are very close to each other and to true values. It is worth noting that likelihood-based methods are challenging when the shape of the likelihood function makes the optimization difficult. Figure 2.5 shows the contour plot of the log-likelihood function (a) and one of its expected values (b), respectively in the case of complete and incomplete data. It is apparent that the function in panel

(b) assumes approximately the same value whatever is τ when R_0 is close to its true value. Thus, the optimization algorithm solution strongly depends on the starting values. To overcome this problem and get reliable point estimates, we run each M-step several times with different starting values. This makes the EM algorithm inefficient and the bootstrap procedure infeasible, that is the reason why we do not provide confidence intervals in Table 2.3.

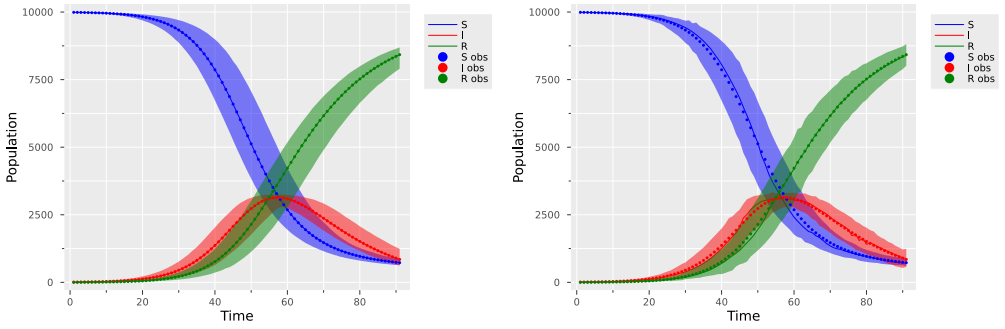
From a Bayesian perspective, we can see that posterior distributions concentrate around the true values of the parameters and that ABC can retrieve an approximate posterior distribution close to the true posterior computed via DA-MCMC (see Section B.3 in Supplemental Materials for further details). We can consider the negligible approximation introduced by likelihood-free methods as the price to pay for having more general and flexible methods that avoid the definition of the likelihood function, its evaluation, and the imputation of missing data. Further details about the implementation of the algorithms are in Section B.3 in Supplemental Materials.

As a general comment, we can conclude that both from a frequentist and Bayesian point of view, likelihood-free methods have some advantages. The bootstrap procedure appears more feasible and fast using the calibration rather than the EM algorithm, while the ABC algorithm overcomes problems related to the autocorrelation of the Markov chain and took only 9 minutes to get posterior quantities very close to those based on the DA-MCMC, the running time of which was equal to 11 minutes. Furthermore, likelihood-free methods allow a straightforward evaluation of the predictive distributions through forward simulations, appropriately accounting for all sources of uncertainty. Figure 2.6 shows the evolution of the compartment sizes over time until $T = 90$, estimated via calibration and ABC. Solid lines are day-by-day punctual estimates of the compartment sizes – i.e. the trajectories computed using MLE parameters (a) and Maximum a Posteriori (MAP) day-by-day predictions. Their closeness to the observed data (dotted lines) suggests a proper fit of the model when the inference is performed via likelihood-free approaches. Confidence and credible bands are retrieved by calculating day-by-day the bootstrap quantiles and the highest posterior density intervals of the predictive distributions, respectively.



a) b)

FIGURE 2.5: Contour plots of the loglikelihood for the SIR model parameters in the case of complete data (a), and of its expected value with respect to the distribution of missing variables in the case of incomplete data (b).



a) b)

FIGURE 2.6: Trajectories of the SIR model in frequentist (a) versus Bayesian (b) inference provided respectively via calibration and ABC. Results are reported in terms of point estimate and 90% confidence interval in the frequentist case and in terms of posterior mean and 90% highest posterior density intervals in the Bayesian case

Note that, in the case of complete data, the frequentist and Bayesian approaches provide similar results also in terms of uncertainty – see confidence and credible intervals in Table 2.2. On the contrary, when the inference is based on incomplete data, Bayesian credibility intervals of the two model parameters are wider than the bootstrap confidence intervals (see Table 2.1), possibly because in the Bayesian approach the imputation of missing data occurs from a distribution that incorporates both the sampling variability and the uncertainty on the parameter space.

2.4.2 A real-world example: the SHC model

In this section, we consider the Smoking Habits Compartmental (SHC) model, developed by Lachi and colleagues [42], as an example of a complex compartmental model where the analytical form of the likelihood function is unavailable. The SHC model has been designed to describe the evolution of smoking habits in a population over the years. Here it is implemented to estimate smoking dynamics in Tuscany, a region of Central Italy, from 1993 to 2019, and forecasts them until 2043.

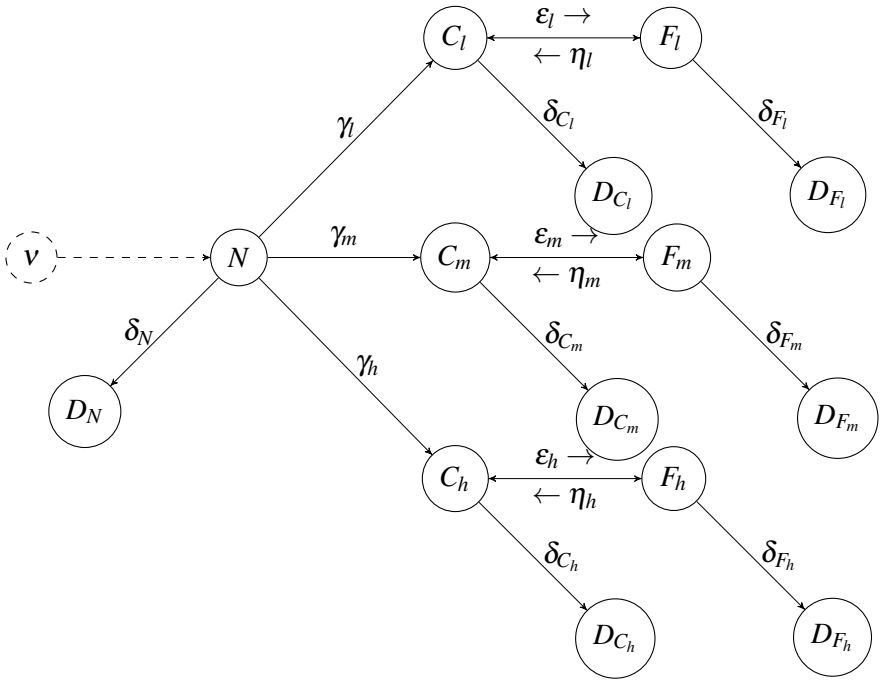


FIGURE 2.7: Smoking Habits Compartmental Model in its simplest form.

The model assumes that, at each point in time, the alive population is divided into the following non-overlapping compartments: never (N), current (C), and former (F) smokers. The compartments C and F are further divided into sub-compartments denoted by C_i and F_i, where $i \in \{l, m, h\}$ indicates the level of smoking intensity, corresponding to low ($<$10 cigarettes/day), medium ($\geq$10 and $<$20 cigarettes/day), and high ($\geq$20 cigarettes/day) smoking intensity, respectively. From each compartment, subjects can transit to a deceased compartment denoted by the letter D and a subscript corresponding to the compartment of origin. New births at time t, denoted by $\nu(t)$, increase the size of the compartment N. Transitions of the individuals from a given compartment to another one are governed by the probabilities of starting smoking (γ_i), stopping smoking (ε_i), and relapsing into smoking after having stopped (η_i). Death happens with different probabilities for never (δ_N), current (δ_{C_i}), and former (δ_{F_i}) smokers belonging to the smoking level category i. In Figure 2.7 a simplified version of the SHC model, which does not consider subjects' age, is depicted. Considering discrete time on an annual scale, $t \in \{1, ..., T\}$, and introducing separate compartments for each age, $a \in \{1, ..., 100\}$, as well as stratification by years since smoking cessation (c) for former smokers, the following system of difference equations arises:

$$
\begin{cases}
N(t;a) = \nu(t-1)\Big(1-\delta_N(a)\Big) & \text{if } a=0 \\
N(t;a) = N(t-1;a-1)\Big(1-\delta_N(a)\Big)\Big(1-\gamma(a)\Big) & \text{if } a>0 \\
C_i(t;a) = 0 & \text{if } a=0 \\
C_i(t;a) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\Big(1-\varepsilon(a)\Big)+ & \\
\qquad N(t-1;a-1)\Big(1-\delta_N(a)\Big)\pi_i\gamma(a)+ & \\
\qquad \sum_{c>0} F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\eta(c) & \text{if } a>0 \\
F_i(t;a,c) = 0 & \text{if } a=0,\ c\geq 0 \\
F_i(t;a,c) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\varepsilon(a) & \text{if } a>0,\ c=0 \\
F_i(t;a,c) = F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\Big(1-\eta(c)\Big) & \text{if } a>0,\ c>0 \\
D_N(t;a) = \nu(t-1)\delta_N(a) & \text{if } a=0 \\
D_N(t;a) = D_N(t-1;a)+N(t-1;a-1)\delta_N(a) & \text{if } a>0 \\
D_{C_i}(t;a) = 0 & \text{if } a=0 \\
D_{C_i}(t;a) = D_{C_i}(t-1;a)+C_i(t-1;a-1)\delta_{C_i}(a) & \text{if } a>0 \\
D_{F_i}(t;a,c) = 0 & \text{if } a=0,\ c\geq 0 \\
D_{F_i}(t;a,c) = 0 & \text{if } a>0,\ c=0 \\
D_{F_i}(t;a,c) = D_{F_i}(t-1;a,c)+F_i(t-1;a-1,c-1)\delta_{F_i}(a,c) & \text{if } a>0,\ c>0
\end{cases}
\tag{2.10}
$$

where $\boldsymbol{\pi} = (\pi_{C_l}, \pi_{C_m}, \pi_{C_h})$ denotes the distribution of the level of smoking intensity among the new current smokers. Note that the probability of starting smoking $\gamma(a)$, as well as the risks of dying $\delta_N(a)$ and $\delta_{C_i}(a)$ depend on the age a, while the probability of relapsing into smoke $\eta(c)$ depends on the years from smoking cessation c and the risk of dying for former smokers $\delta_{F_i}(a,c)$ depends both on a and c. From these dependencies follows that also the number of compartments in the model depends on the values of a and c.

More specifically, $\gamma(a)$ and $\varepsilon(a)$ are modeled through natural cubic regression splines of age with 2 equidistant internal knots, having parameters $\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3)$ and $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3)$, respectively. Regarding the probability of relapsing into smoke $\eta(c)$, it was assumed to vary with time since cessation, according to a negative exponential function governed by positive parameters $\boldsymbol{\omega} = (\omega_0, \omega_1)$. Details on these functions are provided in Section B.4 in Supplemental Materials.

Mortality risks are $\delta_{C_i}(a) = RR_{C_i} \times \delta_N(a)$ and $\delta_{F_i}(a,c) = RR_{F_i}(c) \times \delta_N(a)$, with RR_{C_i} and $RR_{F_i}(c)$ the relative risks of dying for current smokers belonging to the smoking level category i and for people who stopped smoking since c years, belonging to the same smoking level category, versus never smokers.

The distribution of the level of smoking intensity among the new current smokers, $\boldsymbol{\pi}$, is assumed as fixed at values obtained from the National Institute of Statistics (ISTAT) Multipurpose Surveys "Aspect of Daily Life" (`www.istat.it/it/archivio/91926`). Observed data are assumed to be the prevalence of current, never, and former smokers in the observed age classes a^* (14-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75+), for the years in the interval 1993-2019, denoted by $p_C^{\text{obs}}(t;a^*)$, $p_N^{\text{obs}}(t;a^*)$, $p_F^{\text{obs}}(t;a^*)$. Thus, the goal of the inference is to estimate, using two separate models by sex, the vector of unknown model parameters $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\omega})$, given observed data and fixed quantities. Note that in our model formulation, the age-specific risk of dying for never-smokers is unknown, but in this analysis, we treat it as an ancillary quantity. Specifically, we preliminary computed $\delta_N(a)$ as described in [42][1].

The system of equations (2.10) formalizes the assumed mathematical model, but to perform both Bayesian and frequentist analyses, we need to formalize the likelihood function that relates unknown quantities, $\boldsymbol{\theta}$, to the observed prevalence. To this end, we must specify the stochastic generative process associated with the SHC model. For the sake of a straightforward description of the statistical model, 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 allowed to transition and q_x is the probability of that transition, whatever is the compartment. 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 $(\gamma_l(a), \gamma_m(a), \gamma_h(a))$.

Despite the simplicity of the assumed distributions, it is apparent that the analytical form of the likelihood function is infeasible to write down due to the complexity of the mathematical model: the equations governing the SHC model (2.10) are complex and involve a high number of compartments, the transition probabilities are complex functions of age and time from cessation. Only the prevalence $p_C^{\text{obs}}(t;a^*)$, $p_N^{\text{obs}}(t;a^*)$, $p_F^{\text{obs}}(t;a^*)$ are observed and the number of transitions that occur at each t, as well as the size of all compartments at each point in time, represent latent variables. However, the simulation of the stochastic data generative process is straightforward, thus likelihood-free methods such as calibration or ABC are the best solution to infer $\boldsymbol{\theta}$.

[1] If $a \in a^*$, we calculated

$$\delta_N(a) = \frac{1}{T}\sum_t \frac{\delta_{pop}(t;a)}{p_N^{\text{obs}}(t;a^*) + RR_C p_C^{\text{obs}}(t;a^*) + RR_F p_F^{\text{obs}}(t;a^*)},$$

where $\delta_{\text{pop}}(t;a)$ is the mortality rate in the population of age a at time t, obtained from ISTAT (`www.istat.it`), while RR_C and RR_F are the relative risks of dying for current smokers and former smokers versus never smokers obtained from the literature [43].

2.4.2.1 SHC model results

Let $p(t;a^*,\boldsymbol{\theta}) = \big(p_C(t;a^*,\boldsymbol{\theta}), p_N(t;a^*,\boldsymbol{\theta}), p_F(t;a^*,\boldsymbol{\theta})\big)$ be the vector of the prevalence of never, current and former smokers belonging to the class of age a^* at time t, predicted by the model in Equation (2.10), given a specific value of the parameters $\boldsymbol{\theta}$. Within the frequentist framework, we calibrate the model by searching for the value of $\boldsymbol{\theta}$ that minimized the Hellinger distance between the predicted trajectories and the observed ones, defined as follows [44]:

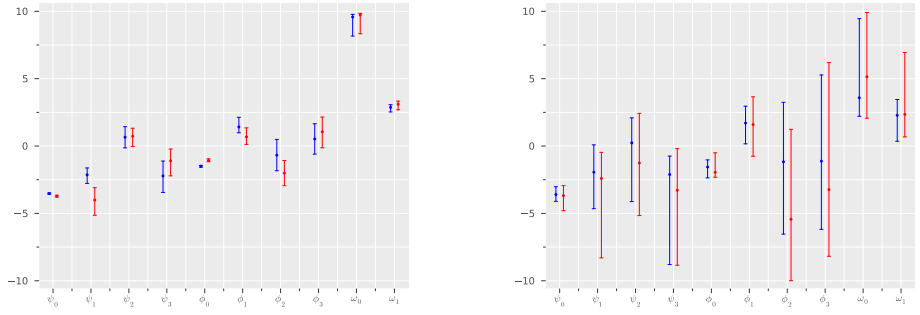
$$Obj(\boldsymbol{\theta}) = \frac{1}{T \times A^* \times \sqrt{2}} \sum_{t,a^*} \sqrt{\sum_{k \in \{C,N,F\}} \left(\sqrt{p_k(t;a^*,\boldsymbol{\theta})} - \sqrt{p_k^{obs}(t;a^*)} \right)^2}. \tag{2.11}$$

We quantify sampling variability around point estimates by using the parametric bootstrap procedure described in Section 2.3.1. As in the previous example, at each bootstrap iteration, we initialized the optimization algorithm at random starting values. Within the Bayesian framework, we implemented the ABC Algorithm 4 by using the same distance function in Equation (2.11), and uniform prior distributions on the parameters. In particular, we specified uniform priors $U[-10,10]$ on the spline parameters and $U(0,10]$ on the exponential function parameters. Details on the implementation of the two algorithms are reported in Section B.4 in Supplemental Materials.

Figure 2.8 presents the results of the inference on $\boldsymbol{\theta}$ obtained via calibration and ABC. Figures 2.9 displays the corresponding estimates of the transition probabilities ($\gamma(a)$, $\varepsilon(a)$, and $\eta(c)$). Confidence bands are obtained by evaluating point-wise the quantiles of the distributions resulting from computing $\gamma(a)$,$\varepsilon(a)$, and $\eta(c)$ using the bootstrap parameter samples. Credible bands correspond to the HPD intervals computed point-wise using samples from the posterior distributions of $\boldsymbol{\theta}$. Solid lines correspond to the values obtained by considering the MLE of θ or the MAP evaluated year-by-year. The curves indicate that males are more likely to start and quit smoking than females and that the probability of starting smoking has a peak around 19 and 20 years of age. The probability of stopping smoking increases after 50 years of age, while the probability of smoking relapse becomes negligible after 2-3 years since cessation. In Figure 2.10, the estimated prevalences for never, current, and former smokers are reported together with the observed ones. The model fit appears to be adequate, with the predicted values close to the observed ones. The forecasts suggest that the smoking prevalence will decrease in the coming years.

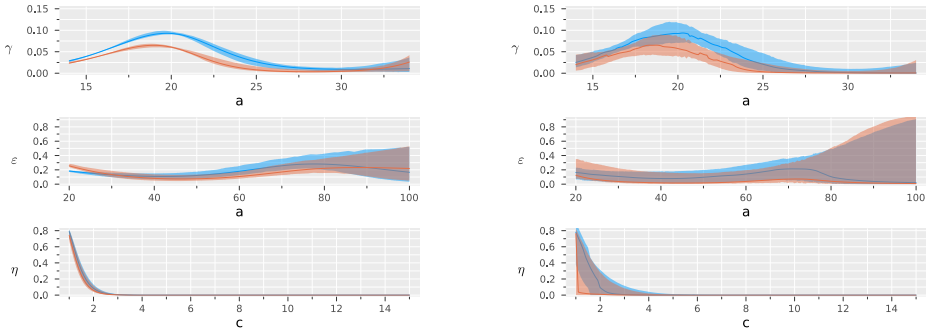
As in the SIR model, the uncertainty around the quantities of interest is greater in the Bayesian framework than in the frequentist one, due to the incorporation of the uncertainty around parameter values in the predictive distribution of the latent variables. From a comparison between the

two methods, it is apparent that ABC avoids problems related to the optimization procedures, such as the dependence on starting values, that could make the estimate unstable. Moreover, it allows the computation of point estimates and the evaluation of all the sources of uncertainty in a single procedure. However, in our implementation, ABC took a longer time with respect to calibration and bootstrap (For further details see B.4 in Supplemental Materials).



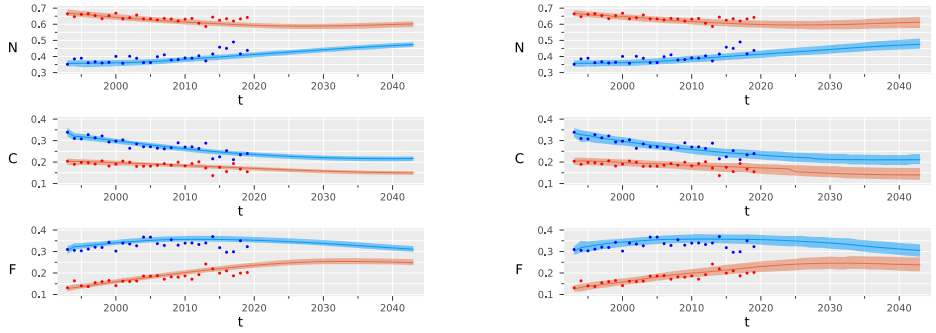
a) b)

FIGURE 2.8: Estimates of SHC model parameters obtained through calibration (a) and ABC (b), for males (blue) and females (pink). Results are reported in terms of point estimate and 90% confidence intervals in the case of calibration and in terms of maximum a posteriori and 90% highest posterior density intervals in the case of ABC.



a) b)

FIGURE 2.9: Probabilities of starting ($\gamma(a)$), quitting ($\varepsilon(a)$), and relapsing into smoke ($\eta(c)$) estimated via calibration (a) and ABC (b), for males (blue) and females (pink). 90% pointwise confidence bands are reported in the case of calibration and 90% pointwise highest posterior density bands in the case of ABC.



a) b)

FIGURE 2.10: Prevalence of never (N), current (C), and former (F) smokers estimated through calibration (a) and ABC (b), for males (blue) and females (pink), with projections up to 2043. 90% pointwise confidence bands are reported in the case of calibration and 90% pointwise highest posterior density bands in the case of ABC. The estimates are compared with the observed prevalences.

2.5 Discussion and conclusions

In this work, we described and compared several Bayesian and frequentist methods that can be used for tasks of inference and prediction in compartmental models. This paper aims to discuss compartmental models from a statistical point-of-view, to fill a gap between the statistical literature and the state-of-the-art in applied fields, and to orient practitioners in the formulation of a proper statistical model and the choice of adequate estimation methods. To the best of our knowledge, few review articles on this topic have been published. In [6] the authors provided a comprehensive review of frequentist and Bayesian methods. They focused on models for dealing with data related to the SARS-CoV-2 epidemic, but most of the considered methods are suitable only for very specific mathematical/statistical models. Furthermore, the review does not cover likelihood-free methods.

Here, we aim to describe a possible way to introduce randomness in a given mathematical compartmental model in a realistic manner, formulate the likelihood function, and provide a description and a comparison among estimation methods. Particular attention has been paid to problems that one may encounter in the evaluation of the likelihood function. We recognized two main reasons for its intractability: the presence of high dimensional missing data and the complexity of the compartmental model. This classification allowed us to identify proper statistical methods for each of the considered cases, and to compare their performance.

We tested the methods at work both on a toy example based on simulated data and a real-world example. From the simulation study, it turned out that all the considered methods can provide point estimates, or posterior distributions, consistent with the “truth". Frequentist methods are all based on optimization strategies that often suffer from problems of strong dependence on starting

values and require repeating the procedure several times, each one initialized at different points. However, this solution increases the computational cost and sometimes makes the bootstrap procedure, needed to quantify sampling variability and compute confidence intervals, infeasible. In this regard, we would like to stress that in the present work, we applied the bootstrap procedure described in Chowell [22], but, to the best of our knowledge, theoretical results on the coverage of the resulting bootstrap intervals are still missing in the statistical literature. As regards Bayesian methods, problems are mainly related to the choice of proposal distributions and to the computational cost of imputing missing data living in high dimensional spaces. DA-MCMC methods often suffer from problems of slow convergence and strong chain autocorrelation.

In this work, a special focus has been placed on the flexibility and potentiality of likelihood-free approaches: calibration and approximate Bayesian computation. They can be implemented whatever the reason for the intractability of the likelihood function, indeed they only require the ability to produce pseudo-data from the mathematical or statistical model and, in the Bayesian framework, to sample from (or evaluate point-wise) the prior distribution. It is interesting to note that, in these estimation methods, the inference procedure would remain largely unaffected even if one assumes a different distribution, such as Poisson or negative binomial.

In our SIR example, both likelihood-based and likelihood-free methods were feasible. This enabled a comparison between the results. In the Bayesian framework, DA-MCMC uses an analytical form of the likelihood function and iterates a Markov chain whose limiting distribution is the exact posterior distribution. At the same time, ABC methods rely on a simulation-based approach that introduces some sources of approximations. However, from our simulation study, it turned out that likelihood-free methods were able to provide good point estimates and that the approximation could be considered negligible, in comparison with DA-MCMC results. It is also worth noting that ABC methods produce i.i.d. samples from the approximate posterior distribution, thus they are highly and straightforwardly parallelizable compared to DA-MCMC algorithms. Finally, the computational cost of likelihood-free methods is less dependent on the cardinality of the latent variable space since they avoid the imputation of missing data and problems related to the mixing of the chain.

A comparison between ABC and MCMC techniques for Bayesian inference in ODE models was also provided in [45]. There, the authors highlighted that the ABC methods that are usually implemented in this framework often fail in quantifying the uncertainty since they do not include adequately the sampling variability in the generative model. Furthermore, they observe that even if ABC provides some computational advantages, they are minimal compared with MCMC methods. However, their comparisons are different from ours. First of all, they consider an observation model that simply adds a Gaussian random error to the numerical solution of the ODE system, as in Equation (2.1). In such a case MCMC methods take advantage of the use of an easy-to-evaluate likelihood function and the major computational burden is required by

the resolution of the ODE system. This approach simplifies the MCMC strategy which is a standard MH algorithm and does not require missing data imputation. However, more realistic observation models, such as the one described in Section 2.4.1 prevent the use of that approach. The second difference is given by the fact that our ABC algorithm uses a simulator that reproduces a data-generative process whose underlying likelihood function is the same as the one involved in the corresponding MCMC method. This means that the comparison among the methods is based on the same observation model and ABC properly includes the sampling variability.

The real-world example fully highlights the great potential of likelihood-free methods, that can retrieve estimates and forecasts even when dealing with very complex models that prevent the use of whatever likelihood-based method. From a comparison between calibration and ABC, we concluded that the results are coherent with each other, thus they give support to the reliability of both methods.

As a general comment, we can note that the calibration and ABC are very similar in spirit. The main difference between the two methods is that, in the point estimation phase, the calibration uses the mathematical model, while ABC resorts only to the statistical model. The presented ABC strategy allows us to consider two sources of variability in a single procedure: the uncertainty over the parameter space described by prior distributions and the sampling variability reproduced by the simulator. Instead, the calibration must be combined with an adequate bootstrap procedure to quantify the sampling variability. Another strength of ABC methods, compared with the calibration, is that they do not rely on optimization strategies, thus avoiding problems of dependence on the starting values.

ABC is in some sense related to the Generalized Likelihood Uncertainty Estimation (GLUE) approach, a very common technique in the hydrological literature that represents one of the first attempts at overcoming standard calibration procedures by providing an uncertainty assessment [46]. The connection between the two methods has been discussed in [47], where the authors showed that the GLUE approach can be interpreted as a particular ABC algorithm, known as Importance Sampling ABC. However, there are some relevant differences between the two approaches. First of all, in GLUE techniques the use of a (uniform) prior distribution over the parameter space must be intended as a way to introduce uncertainty in the estimation rather than a formalization of prior belief in a fully Bayesian spirit. Moreover, GLUE procedures often produce pseudo-data from the mathematical model. Thus, they can be seen as an implementation of an ABC algorithm that uses a simulator that does not account for sampling variability but is just a complex mathematical function of the parameters – see the discussion about model misspecification in ABC methods in [45]. Thus, the output of the algorithm is a sample from a distribution that should not be interpreted as a posterior distribution in a strict Bayesian sense, since it includes only the variability induced by the prior distribution. That distribution may be seen as a redefinition of the prior distribution over the subset of parameter values that are

coherent with the observed data. It follows that, in the limit of the ABC threshold going to 0, this distribution converges to a point of mass over parameter values that lead to pseudo-data equal to observed data. Accordingly, GLUE/ABC results converge to that of a calibration procedure and the GLUE/ABC can be interpreted as a “stochastic search" of the optimal parameters. This conclusion suggests that, in a complex model with a high number of parameters, such as the SHC model, ABC can help to find global minima by overcoming the difficulties related to optimization algorithms. However, the efficiency in ABC strongly depends on the computational cost of the simulator and the adequacy of the proposal distributions.

The present work has several limitations. First of all, we restricted ourselves to discrete-time models, but models for continuous-time Markov processes can be implemented in this field. In such cases, the formulation of the likelihood function is usually based on the assumption that the waiting time between two events is exponentially distributed. Several algorithms for performing Bayesian inference under this assumption have been proposed [48]. Other possible approaches are based on state space models (e.g., [49]) for which posterior distribution can be inferred by resorting to Particle Markov Chain Monte Carlo methods [50]. It is also worth noting that here we considered simple and easy-to-reproduce sampling methods in the implementation of the ABC method, as well as DA-MCMC methods. The state-of-the-art includes more sophisticated gradient-based algorithms such as Metropolis Adjusted Langevin Algorithm (MALA) [51], Hamiltonian Monte Carlo methods (HMC). Moreover, likelihood-free algorithms that exploit some approximations of the generative model, or can cleverly orient the simulation procedure, have been proposed (see [52] for a discussion). Many of them rely on neural networks or other machine learning approaches such as normalizing flows – see [53, 54], among others. Finally, this work does not claim to discuss exhaustively all the statistical methods that can be potentially used in the described scenario. Further work should be done to investigate the applicability of methods such as Indirect Inference [55] or Variational Bayes [56] and to compare their performance with that of the considered methods.

Authors' contributions: C.V. and M.B. conceptualized the project; A.L. wrote the codes and conducted the statistical analysis under the statistical supervision of C.V. and the epidemiological supervision of M.B.; C.V. and A.L. wrote the first version of the paper; M.B. provided critical feedback. All the authors read and approved the final version of the paper.

Bibliography

[1] L.D. Broemeling. *Bayesian analysis of infectious diseases: COVID-19 and beyond*. Chapman and Hall/CRC biostatistics series. CRC Press, Taylor and Francis Group, Boca Raton London New York, 2021. ISBN 978-0-367-64724-7 978-0-367-63386-8.

[2] N. Jourdan, T. Neveux, O. Potier, M. Kanniche, J. Wicks, I. Nopens, U. Rehman, and Y. Le Moullec. Compartmental Modelling in chemical engineering: A critical review. *Chemical Engineering Science*, 210:115196, 2019. ISSN 00092509. doi: 10.1016/j.ces.2019.115196. URL https://linkinghub.elsevier.com/retrieve/pii/S0009250919306888.

[3] V. Booth, J. Rinzel, and O. Kiehn. Compartmental Model of Vertebrate Motoneurons for Ca $^{2+}$ -Dependent Spiking and Plateau Potentials Under Pharmacological Treatment. *Journal of Neurophysiology*, 78(6):3371–3385, 1997. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.1997.78.6.3371. URL https://www.physiology.org/doi/10.1152/jn.1997.78.6.3371.

[4] L. Cao, H. Zhao, X. Wang, and X. An. Competitive information propagation considering local-global prevalence on multi-layer interconnected networks. *Frontiers in Physics*, 11:1293177, 2023. doi: https://doi.org/10.3389/fphy.2023.1293177.

[5] T.J. McKinley, I. Vernon, I. Andrianakis, N. McCreesh, J.E. Oakley, R.N. Nsubuga, M. Goldstein, and R.G. White. Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models. *Statistical Science*, 33(1):4 – 18, 2018. doi: 10.1214/17-STS618.

[6] L. Tang, Y. Zhou, L. Wang, S. Purkayastha, L. Zhang, J. He, F. Wang, and P.X.K. Song. A review of multi-compartment infectious disease models. *International Statistical Review*, 88(2):462–513, 2020. doi: 10.1111/insr.12402.

[7] J.C. Butcher. *Numerical methods for ordinary differential equations*. Wiley, Chichester, West Sussex, United Kingdom, third edition edition, 2016. ISBN 978-1-119-12150-3.

[8] T.G. Kurtz. Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. *Journal of Applied Probability*, 7(1):49–58, 1970. ISSN 00219002. doi:

10.2307/3212147. URL http://www.jstor.org/stable/3212147. Publisher: Applied Probability Trust.

[9] T.G. Kurtz. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *Journal of Applied Probability*, 8(2):344–356, 1971. ISSN 00219002. doi: 10.2307/3211904. URL http://www.jstor.org/stable/3211904. Publisher: Applied Probability Trust.

[10] P.D. O'Neill and G.O. Roberts. Bayesian Inference for Partially Observed Stochastic Epidemics. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 162 (1):121–129, 1999. ISSN 09641998, 1467985X. doi: 10.1111/1467-985X.00125. URL http://www.jstor.org/stable/2680471. Publisher: [Wiley, Royal Statistical Society].

[11] A.G. Mckendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925. doi: 10.1017/S0013091500034428.

[12] W. O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927. doi: 10.1098/rspa.1932.0063.

[13] R. Anderson. The Kermack-McKendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53(1-2):3–32, 1991. ISSN 00928240. doi: 10.1016/S0092-8240(05)80039-4. URL http://link.springer.com/10.1016/S0092-8240(05)80039-4.

[14] L.J.S. Allen and A.M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163(1):1–33, January 2000. ISSN 00255564. doi: 10.1016/S0025-5564(99)00047-4. URL https://linkinghub.elsevier.com/retrieve/pii/S0025556499000474.

[15] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong. Epidemic analysis of covid-19 in china by dynamical modeling. *medRxiv*, 2020. doi: 10.1101/2020.02.16.20023465.

[16] R. Schlickeiser and M. Kröger. Analytical modeling of the temporal evolution of epidemics outbreaks accounting for vaccinations. *Physics*, 3(2):386–426, 2021. doi: 10.3390/physics3020028.

[17] B. Canto, C. Coll, and E. Sanchez. Estimation of parameters in a structured sir model. *Advances in Difference Equations*, 2017:1–13, 2017. doi: 10.1186/s13662-017-1078-5.

[18] M. Baccini, G. Cereda, and C. Viscardi. The first wave of the SARS-CoV-2 epidemic in Tuscany (Italy): A SI2R2D compartmental model with uncertainty evaluation. *PLOS*

ONE, 16(4):e0250029, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0250029. URL https://dx.plos.org/10.1371/journal.pone.0250029.

[19] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965. doi: 10.1093/comjnl/7.4.308.

[20] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977. ISSN 00359246. doi: 10.1111/j.2517-6161.1977.tb01600.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1977.tb01600.x.

[21] R.A. Levine and G. Casella. Implementations of the Monte Carlo EM Algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439, 2001. ISSN 10618600. doi: 10.1198/106186001317115045. URL http://www.jstor.org/stable/1391097.

[22] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042717300234.

[23] W. Zucchini, I.L. MacDonald, and R. Langrock. *Hidden Markov models for time series: an introduction using R*. Number 150 in Monographs on statistics and applied probability. CRC Press, Taylor & Francis Group, Boca Raton London New York, second edition, first issued in paperback edition, 2021. ISBN 978-1-03-217949-0 978-1-4822-5383-2. doi: 10.1201/b20790.

[24] B. Efron and R. Tibshirani. *An introduction to the bootstrap*. Number 57 in Monographs on statistics and applied probability. Chapman & Hall, Boca Raton, Fla., nachdr. edition, 1998. ISBN 978-0-412-04231-7.

[25] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949. doi: 10.1080/01621459.1949.10483310.

[26] A.W. Marshall. *The Use of Multi-stage Sampling Schemes in Monte Carlo Computations*. Rand Corporation. Rand Corporation, 1954. URL https://books.google.it/books?id=60kLYAAACAAJ.

[27] C. P. Robert and W. Changye. *Markov Chain Monte Carlo Methods, Survey with Some Frequent Misunderstandings*, pages 1–28. John Wiley & Sons, Ltd, 2021. ISBN 9781118445112. doi: 10.1002/9781118445112.stat08285. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat08285.

[28] M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540, 1987. doi: 10.1080/01621459.1987.10478458.

[29] J.S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994. doi: 10.1080/01621459.1994.10476829.

[30] C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013. ISBN 978-1441919397.

[31] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. ISSN 13507265. URL http://www.jstor.org/stable/3318737.

[32] R.M. Neal. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2 (11):2, 2011. doi: 10.1201/b10905-6.

[33] H.M. Afshar and J. Domke. Reflection, refraction, and hamiltonian monte carlo. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/8303a79b1e19a194f1875981be5bdb6f-Paper.pdf.

[34] C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):269–342, 2010. doi: 10.1111/j.1467-9868.2009.00736.x.

[35] D.B. Rubin. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, 12(4), 1984. ISSN 0090-5364. doi: 10.1214/aos/1176346785.

[36] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun, and M.W. Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, 1999. ISSN 0737-4038, 1537-1719. doi: 10.1093/oxfordjournals.molbev.a026091. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a026091.

[37] S. Tavaré, D.J. Balding, R.C. Griffiths, and P. Donnelly. Inferring Coalescence Times From DNA Sequence Data. *Genetics*, 145(2):505–518, 1997. ISSN 1943-2631. doi: 10.1093/genetics/145.2.505. URL https://academic.oup.com/genetics/article/145/2/505/6018089.

[38] S.A. Sisson, Y. Fan, and M.A. Beaumont, editors. *Handbook of approximate Bayesian computation*. CRC Press, Taylor and Francis Group, Boca Raton, 2019. ISBN 978-1-4398-8150-7.

[39] M.A. Beaumont, J.M. Cornuet, J.M. Marin, and C.P. Robert. Adaptive approximate bayesian computation. *Biometrika*, 96(4):983–990, 2009. doi: 10.1093/biomet/asp052. URL http://www.jstor.org/stable/27798882.

[40] M. Lenormand, F. Jabot, and G. Deffuant. Adaptive approximate bayesian computation for complex models. *Computational Statistics*, 28(6):2777–2796, 2013. doi: doi.org/10.1007/s00180-013-0428-3.

[41] P.K. Mogensen and A.N. Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615, 2018. ISSN 2475-9066. doi: 10.21105/joss.00615. URL http://joss.theoj.org/papers/10.21105/joss.00615.

[42] A. Lachi, C. Viscardi, G. Cereda, G. Carreras, and M. Baccini. A compartmental models for smoking dynamics in Italy: A pipeline for inference, validation, and forecasting under hypothetical scenarios. preprint, In Review, 2023. URL https://www.researchsquare.com/article/rs-3303111/v1.

[43] M.J. Thun, B.D. Carter, D. Feskanich, N.D. Freedman, R. Prentice, A.D. Lopez, P. Hartge, and S.M. Gapstur. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine*, 368(4):351–364, 2013. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMsa1211127. URL http://www.nejm.org/doi/10.1056/NEJMsa1211127.

[44] E. Hellinger. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136):210–271, 1909. ISSN 1435-5345, 0075-4102. doi: 10.1515/crll.1909.136.210. URL https://www.degruyter.com/document/doi/10.1515/crll.1909.136.210/html.

[45] A.A. Alahmadi, J.A. Flegg, D.G. Cochrane, C.C Drovandi, and J.M. Keith. A comparison of approximate versus exact techniques for Bayesian parameter inference in nonlinear ordinary differential equation models. *Royal Society Open Science*, 7(3):191315, 2020. ISSN 2054-5703. doi: 10.1098/rsos.191315. URL https://royalsocietypublishing.org/doi/10.1098/rsos.191315.

[46] K. Beven and A. Binley. The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, 6(3):279–298, 1992. doi: 10.1002/hyp.3360060305.

[47] D.J. Nott, L. Marshall, and J. Brown. Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection?: TECHNICAL NOTE.

Water Resources Research, 48(12), 2012. ISSN 00431397. doi: 10.1029/2011WR011128. URL http://doi.wiley.com/10.1029/2011WR011128.

[48] C.M. Pooley, S.C. Bishop, and G. Marion. Using model-based proposals for fast parameter inference on discrete state space, continuous-time markov processes. *Journal of The Royal Society Interface*, 12(107):20150225, 2015. doi: 10.1098/rsif.2015.0225.

[49] L. Wang, Y. Zhou, J. He, B. Zhu, F. Wang, L. Tang, M. Kleinsasser, D. Barker, M.C. Eisenberg, and P.X.K. Song. An epidemiological forecast model and software assessing interventions on the covid-19 epidemic in china. *Journal of Data Science*, 18(3):409–432, 2020. doi: doi.org/10.6339/JDS.202007_18(3).0003.

[50] E. Akira, V.L. Edwin, and B. Marc. Introduction to particle markov-chain monte carlo for disease dynamics modellers. *Epidemics*, 29:100363, 2019. ISSN 1755-4365. doi: doi.org/10.1016/j.epidem.2019.100363. URL https://www.sciencedirect.com/science/article/pii/S1755436519300301.

[51] G.O Roberts and O. Stramer. Langevin diffusions and metropolis-hastings algorithms. *Methodology and computing in applied probability*, 4:337–357, 2002. doi: 10.1023/A:1023562417138.

[52] K. Cranmer, J. Brehmer, and G.Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020. doi: 10.1073/pnas.1912789117.

[53] G. Papamakarios and I. Murray. Fast ε-free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016. doi: 10.48550/arXiv.1605.06376.

[54] D. Prangle and C. Viscardi. Distilling importance sampling for likelihood free inference. *Journal of Computational and Graphical Statistics*, pages 1–11, 2023. doi: 10.1080/10618600.2023.2175688.

[55] C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118, 1993. doi: 10.1002/jae.3950080507.

[56] M.N. Tran, D. Nott, and R. Kohn. Variational bayes. *Wiley StatsRef: Statistics Reference Online*, pages 1–9, 2014. doi: 10.1002/9781118445112.stat08387.

CHAPTER 3

Smoking dynamics in Tuscany (Italy) under alternative tobacco control policies: Robustifying inference and forecasting via uncertainty propagation and Global Sensitivity Analysis

Draft paper to be submitted to *peer review journal*

Alessio Lachi[1,2], Giulia Carreras[3], Cecilia Viscardi[1,4], Andrea Saltelli[5], Michela Baccini[1,4]

[1] Department of Statistics, Computer Science, and Applications “Giuseppe Parenti" (DiSIA), University of Florence, Viale Giovan Battista Morgagni 59-65 50134 - Florence, Italy

[2] Epidemiology and Health Research, Institute of Clinical Physiology of the Italian National Institute Research Council (IFC-CNR), Via Giuseppe Moruzzi 1 56124 - Pisa, Italy

[3] Oncologic Network, Prevention and Research Institute (ISPRO), Health Service of Tuscany, Via Cosimo il Vecchio 2 50139 - Florence, Italy

[4] Florence Center for Data Science, University of Florence, Viale Giovanni Battista Morgagni 59-65 50134 - Florence, Italy

[5] UPF Barcelona School of Management (UPF-BSM), University Pompeu Fabra of Barcelona, Carrer de Balmes 132-134 08008 - Barcelona

Abstract

Cigarette smoking has still a significant impact on population morbidity and mortality. This study introduces an innovative approach to enhance the reliability of inferences and forecasts produced by the Smoking Habits Compartmental (SHC) model, a compartmental model for simulating smoking dynamics. While compartmental models like SHC offer a valuable framework for understanding complex systems and projecting public health dynamics, they suffer from limitations related to stringent model assumptions and parameter identifiability. The proposed methodology aims to robustify inference, forecasting, and the assessment of tobacco control

policies by systematically incorporating uncertainties present in model definition and data. This involves a strategy of error propagation analysis and a variance-based Global Sensitivity Analysis (GSA). The GSA provides insights into how the variance of model outputs can be attributed to uncertainties in model inputs, utilizing sensitivity indexes. The study underscores the importance of considering all sources of uncertainty in the modeling process, especially when crafting forecasts under hypothetical scenarios for guiding public health policies. The proposed robustification procedure, incorporating GSA, contributes to a more comprehensive assessment of variability, aiding in the identification of influential data subsets and variables. The paper confirms previous findings regarding model parameters but also highlights the dependence of inference on the calibration window used, signaling potential issues with assumptions about the constancy of transitions between compartments over time. The research concludes with an evaluation of alternative tobacco control policies for Tuscany, emphasizing the substantial impact of uncertainty on policy effectiveness. Overall, the study advocates for the integration of uncertainty analysis and GSA in modeling processes, providing a more nuanced understanding of the robustness of public health projections and guiding the development of intricate models.

Keywords - Compartmental Models, Smoking Dynamics, Tobacco Control Policies, Global Sensitivity Analysis, Sobol Indexes

3.1 Introduction

Smoking is a substantial global health issue due to its impact on morbidity and mortality, and the related economic burden [1]. As reported by the Global Burden of Disease Study 2019 [2], tobacco smoking caused in 2019 more than 7 million deaths and 200 million disability-adjusted life-years worldwide, and remains the leading mortality risk factor among males. The WHO Framework Convention on Tobacco Control (WHO FCTC), which formally entered into force for state parties in 2005 [3], contains guidelines and requirements for the implementation of tobacco control policies (TCP), summarized by the MPOWER acronym (see Section C.1 in Supplemental Materials): Monitoring tobacco use and prevention policies (M), Protecting people from tobacco smoke (P), Offering help to quit tobacco use (O), Warning about the dangers of tobacco (health warning labels (W-L) and mass media (W-MM)), Enforcing bans on tobacco advertising, promotion, and sponsorship (E), and Raising taxes on tobacco (R) [4](`www.drugsandalcohol.ie/37050/1/WHO-TobaccoControl-Roadmap_of_actions_220592.pdf`). Also thanks to the implementation of WHO FCTC, tobacco use decreased across the European region during the last decade. Still, the rate of this reduction has been relatively slow, and Europe continues to have one of the highest rates of tobacco use globally (`www.drugsandalcohol.ie/37050/1/WHO-TobaccoControl-Roadmap_of_actions_220592.pdf`).

Understanding the complex dynamics of tobacco use and its effects on the population thus remains paramount for crafting effective public health policies, and the development of models that simulate smoking dynamics under different policy scenarios is useful for comparing hypothetical future interventions. Among the approaches proposed in the literature for simulating smoke dynamics, compartmental models offer a structured framework. The SimSmoke model is the most used compartmental model in this field. Developed by Levy and Friend [5] to capture smoking dynamics, it has been implemented in a wide number of countries including Italy [6–9]. The SimSmoke model estimates the relevant parameters that govern the transitions between compartments, such as age-specific probabilities of starting and stopping smoking, makes projections, and explores short and long-term impacts on smoking dynamics and related health outcomes of different hypothetical tobacco control policies (TCP), deriving their effects from literature reviews conducted with the advice of an expert panel [5–7, 9–22].

However, despite the significant development and widespread use of this complex modeling framework, formal procedures capable of accounting for all, or at least a significant portion, of the sources of variability involved have not been proposed. Some authors introduced methods to account for the variability around the model parameters in the SimSmoke model, including Probabilistic Sensitivity Analysis and bootstrapping [15, 23, 24], but these are only partial solutions that do not treat comprehensively uncertainties associated with model formulation, inference procedure, and sampling variability in the data. This is a significant limitation, especially when modeling is intended to generate forecasts under hypothetical future scenarios and aims to serve as a tool for providing public health guidance.

In this paper, we propose a procedure aimed at robustifying inference, forecasting, and TCPs assessment produced by the Smoking Habits Compartmental (SHC) model, which has been developed by Lachi et al. [25] grounding on the SimSmoke model experience to model smoking dynamics in the Tuscany region (Italy). This robustification procedure, which can also be reproduced in different models and contexts, takes into account uncertainties involved in model definition, inference, and forecasting, and consists of an uncertainty assessment and a variance-based Global Sensitivity Analysis (GSA) [26, 27]. In statistical modeling, uncertainty assessment is the process of quantifying the uncertainty associated with model assumptions, parameters, and predictions. Uncertainty may arise from various sources, such as measurement errors, sampling variability, model misspecification, and unobserved variables. Sensitivity analysis represents a general statistical concept to evaluate the robustness of parameter estimates and the sensitivity of the results to model assumptions [26]. Uncertainty assessment can be integrated with GSA, which studies how the variance of results (outputs) can be decomposed into contributions from uncertain model inputs, through the use of sensitivity indexes [28]. Especially when the overall uncertainty around the output is high, performing GSA is paramount to understanding which factors mostly influence the answer to our research question. This can help to identify model assumptions

that are most critical and quantify the relative impact of different sources of uncertainty on the research findings.

The main goal of this paper can be summarised in the following points:

- assessing the robustness of the inference parameters produced by the SHC model [25] as the assumptions governing its functioning change;
- providing a forecast of smoking dynamics in the Tuscany population, both in terms of smoking prevalence and Smoking-Attributable Deaths (SAD);
- comparing hypothetical TCPs and evaluating their relative effectiveness over time.

Through the inclusion inside the estimation and forecasting process of the variability induced by the compartmental model's assumptions, observed data used for calibration, and the effect of different TCPs, we robustify the inference and produce predictions that take into account all these sources of uncertainty in the modeling process.

3.2 Uncertainty Analysis and Global Sensitivity Analysis

In statistical modeling, uncertainty analysis is the process of quantifying all uncertainties that intervene at each step of the study – including those associated with model assumptions, parameters definition and inference, and predictions – and assessing their impact on the results. Let $(X_1, X_2, ..., X_{K_X})$ be K_X mutually independent inputs and $f(\cdot)$ a function which, given the inputs, returns K_Y outputs $(Y_1, Y_2, ..., Y_{K_Y})$. Here, we refer to the function $f(\cdot)$, as the Sensitivity Analysis (SA) function. After eliciting a probability distribution on the inputs, Monte Carlo simulations can be performed to propagate the uncertainty from the inputs to the outputs. In each of these simulations, the value of each input factor is sampled from its distribution and the SA function is used to calculate the corresponding outputs, obtaining a sample from their joint distribution, that can be described and analyzed.

Sometimes in modeling studies, it is appropriate to consider multiple candidate models that differ in their assumptions or specifications. In this case, we can perform Monte Carlo simulations, as discussed in [29], by defining random "triggers" that determine the model to be followed in each simulation. In this way, we account also for the uncertainty associated with the model choice, considering the trigger as an additional input.

Uncertainty analysis works in tandem with GSA, which involves examining the relative importance of model inputs or assumptions in determining uncertainty in the model outputs. GSA aims to provide sensitivity measures that are global in the sense that they are concerned with the whole

space of variability of the inputs. GSA quantifies the relative importance of each input on the outcomes by computing variance indices. The first-order index (S_j) represents the main effect contribution of each input factor to the variance of the output and it is defined as:

$$S_j = \frac{V_{X_j}(E_{X\sim j}(Y|X_j))}{V(Y)}.$$

The first order index represents the expected reduction of the variance that would be achieved if the factor X_j could be fixed [28]. It does not account for the variance explained by the interaction of X_j with the other inputs. When we are interested in evaluating also iterations, we can focus on the total variance index. For each input X_j, the so-called total variance index (S_j^{tot}) measures the overall effect of the j-th input on the output Y, including all the interactions of X_j with the other inputs. This index corresponds to the expected variance of Y that would be left on average when all the parameters but X_j, $X_{\sim j}$, are fixed:

$$S_j^{tot} = \frac{E_{X\sim j}(Var_{X_j}(Y|X_{\sim j}))}{Var(Y)}.$$

The total effect index accounts for the total contribution to the output variation due to factor X_j, i.e. its first-order effect expressed by the first-order index plus all higher-order effects due to interactions. A total variance index close to zero indicates that the parameter X_j does not influence Y. Conversely, a large total variance index indicates that the parameter does have an impact on the output. Note that, while $\sum_j S_j \leq 1$, the sum of the total variance indexes over all inputs can be larger than 1.

3.3 Model specification, inference procedure, and policy assessment

The SA function defines the analysis conducted and determines the inputs allowed to vary and the outputs we are interested in. Our application, includes compartmental model specification, inference on the unknown parameters of the model, forecasting of compartment sizes, and health impact assessment under different TCP scenarios. In the upcoming sections, we detail the components of the function $f(\cdot)$.

3.3.1 Smoking habits compartmental model

The SHC model proposed by Lachi et al. [25] attempts to describe the evolution of smoking habits in Tuscany from 1993 to 2019 and forecast them until 2043.

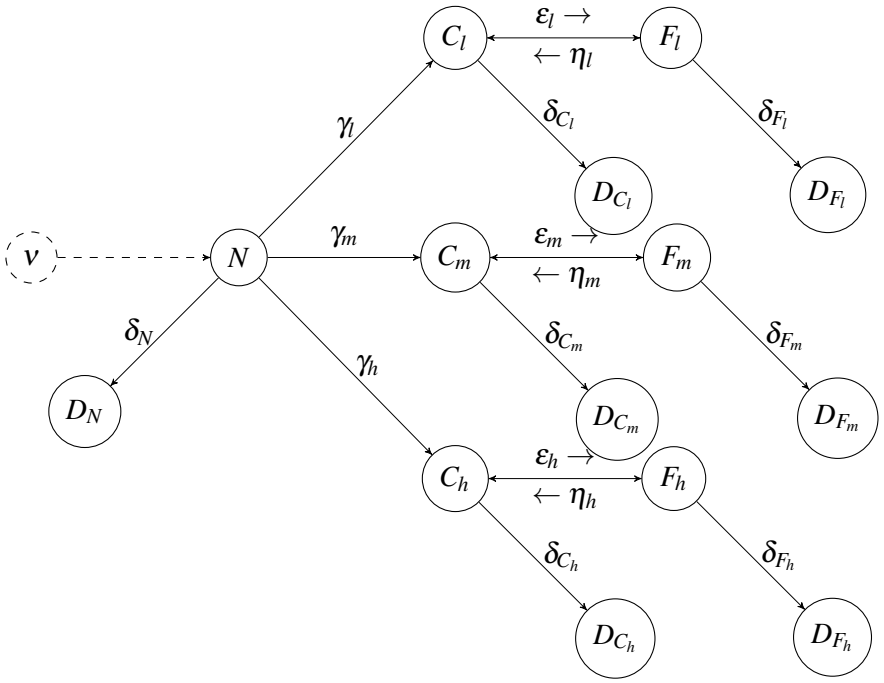


FIGURE 3.1: Smoking Habits Compartmental Model in its simplest form.

Figure 3.1 shows a simplified version of the SHC model, which does not account for the subject's age. At each time the alive population is divided into the following non-overlapping compartments: never (N), current (C), and former (F) smokers. The compartments C and F are further divided into sub-compartments denoted by C_i and F_i, where $i \in \{l, m, h\}$ indicates the level of smoking intensity, corresponding to low ($<$10 cigarettes/day), medium ($\geq$10 and $<$20 cigarettes/day), and high ($\geq$20 cigarettes/day) smoking intensity, respectively. From each compartment, subjects can transit to a deceased compartment denoted by D and a subscript corresponding to the compartment of origin. New births, ν, increase the size of the compartment N. Transitions of the individuals from a given compartment to another one are regulated by the probabilities of starting smoking (γ_i), stopping smoking (ε_i), and relapsing into smoking after having stopped (η_i). Death happens with different probabilities for never (δ_N), current (δ_{C_i}), and former (δ_{F_i}) smokers.

Considering discrete time on an annual scale, $t \in \{1, ..., T\}$, and introducing separate compartments for each discrete age, $a \in \{1, ..., 100\}$, as well as stratification by years since smoking cessation (c) for former smokers, the following system of difference equations arises:

$$
\begin{cases}
N(t;a) = \nu(t-1)\Big(1-\delta_N(a)\Big) & \text{if } a=0 \\
N(t;a) = N(t-1;a-1)\Big(1-\delta_N(a)\Big)\Big(1-\gamma(a)\Big) & \text{if } a>0 \\
C_i(t;a) = 0 & \text{if } a=0 \\
C_i(t;a) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\Big(1-\varepsilon(a)\Big)+ & \\
\qquad N(t-1;a-1)\Big(1-\delta_N(a)\Big)\pi_i\gamma(a)+ & \\
\qquad \sum_{c>0} F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\eta(c) & \text{if } a>0 \\
F_i(t;a,c) = 0 & \text{if } a=0,\ c\geq 0 \\
F_i(t;a,c) = C_i(t-1;a-1)\Big(1-\delta_{C_i}(a)\Big)\varepsilon(a) & \text{if } a>0,\ c=0 \\
F_i(t;a,c) = F_i(t-1;a-1,c-1)\Big(1-\delta_{F_i}(a,c)\Big)\Big(1-\eta(c)\Big) & \text{if } a>0,\ c>0 \\
D_N(t;a) = \nu(t-1)\delta_N(a) & \text{if } a=0 \\
D_N(t;a) = D_N(t-1;a)+N(t-1;a-1)\delta_N(a) & \text{if } a>0 \\
D_{C_i}(t;a) = 0 & \text{if } a=0 \\
D_{C_i}(t;a) = D_{C_i}(t-1;a)+C_i(t-1;a-1)\delta_{C_i}(a) & \text{if } a>0 \\
D_{F_i}(t;a,c) = 0 & \text{if } a=0,\ c\geq 0 \\
D_{F_i}(t;a,c) = 0 & \text{if } a>0,\ c=0 \\
D_{F_i}(t;a,c) = D_{F_i}(t-1;a,c)+F_i(t-1;a-1,c-1)\delta_{F_i}(a,c) & \text{if } a>0,\ c>0
\end{cases}
\tag{3.1}
$$

In the system 3.1, $\delta_N(a)$ represents the probability of dying for never smokers of age a, $\delta_{C_i}(a)$ the one for current smokers belonging to the smoking level category i, $\delta_{F_i}(a,c)$ the probability of dying for former smokers of age a belonging to the smoking level category i after c years from smoking cessation, and $\boldsymbol{\pi} = (\pi_{C_l}, \pi_{C_m}, \pi_{C_h})$ is the distribution of the level of smoking intensity among the new current smokers. We assumed the following functions on the probabilities of starting and stopping smoking:

$$
\gamma(a) = \begin{cases} 0 & 0 \leq a \leq 13 \cup\ a \geq 35 \\ \frac{\exp(s(a;\boldsymbol{\psi}))}{1+\exp(s(a;\boldsymbol{\psi}))} & 14 \leq a \leq 34 \end{cases} \qquad \varepsilon(a) = \begin{cases} 0 & 0 \leq a \leq 19 \\ \frac{\exp(s(a;\boldsymbol{\phi}))}{1+\exp(s(a;\boldsymbol{\phi}))} & a \geq 20, \end{cases}
$$

where $s(a;\boldsymbol{\psi})$ and $s(a;\boldsymbol{\phi})$ are two natural cubic regression splines of age, with 2 equidistant internal knots, governed by the parameter vectors $\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3)$ and $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3)$.

The relapsing probability, $\eta(c)$, was modeled as a negative exponential function of time since cessation, with parameters $\boldsymbol{\omega} = (\omega_0, \omega_1)$:

$$\eta(c) = \begin{cases} 0 & c = 0 \\ 1 - exp(-\omega_0 \omega_1 \exp(-\omega_1 c)) & 1 \leq c \leq 15 \\ 1 - exp(-\omega_0 \omega_1 \exp(-\omega_1 15)) & c \geq 16. \end{cases}$$

Note that ω_0 governs the lifetime probability of no relapse and ω_1 tunes how fast the probability of smoking relapse declines with the time from cessation [13–15, 30]. Both ω_0 and ω_1 are assumed to be positive so that $\eta(c)$ is a positive and decreasing function of c. Regarding mortality risks, we assumed $\delta_{C_i}(a) = RR_{C_i} \times \delta_N(a)$ and $\delta_{F_i}(a,c) = RR_{F_i}(c) \times \delta_N(a)$, with RR_{C_i} and $RR_{F_i}(c)$ relative risks of dying for current smokers belonging to the smoking level category i and for people who stopped smoking since c years, belonging to the same smoking level category, versus never smokers. More details about the assumptions governing the SHC model are reported in Lachi et al. [25].

We assumed that the transition probabilities remain constant over calendar time except when starting from a specific year, they are modified by TCP effects (see Section 3.3.4).

3.3.1.1 Stochastic version of the SHC model

To account for intrinsic uncertainty affecting the transitions between model compartments, we consider a stochastic version of the SHC model. Let us denote by S the number of individuals who transit from a generic compartment to another one. We assume $S \sim Binomial(n_S, q_S)$, where n_S is the number of individuals allowed to transition and q_S 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_S is the number of low-intensity smokers of age a that do not die during the year t, and q_S is equal to $\varepsilon(a)$. The same reasoning applies to the number of individuals relapsing smoking and to the number of deaths. We assume that the number of individuals who start smoking at age a is distributed according to a Multinomial distribution with a vector of probabilities $(\gamma_l(a), \gamma_m(a), \gamma_h(a))$.

3.3.2 Data and calibration

According to Lachi et al. [25], we focus the inference on $\delta_N(a)$ and $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\omega})$. For the strict purpose of inference, all other parameters involved in the system of difference equations 3.1 are assumed as given. Inference is based on demographic data from the National Institute of Statistics (`http://dati.istat.it`) and smoking habit data obtained from two primary sources: the Italian Multiscope Survey on Families (AVQ survey, `https://www.istat.it/it/archivio/91926`), which gathers essential information related to the daily life of

individuals and families, and the European Health Interview Survey (EHIS, `https://www.istat.it/it/archivio/167485`). The smoking habit data consist of the prevalence of never, current, and former smokers, categorized by year, age group ($a^* \in \{14-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75+\}$), and smoking intensity category. Observed data are denoted by 'obs' at the apex. Inference is performed in two steps:

1. We estimate the age-specific risks of mortality for never smokers $\delta_N(a)$ as a combination of observed values of prevalence and relative risks. In particular, we apply the following formula:
$$\delta_N(a) = \frac{1}{T}\sum_{t=0}^{T} \frac{\delta^{\text{obs}}(t;a)}{p_N^{\text{obs}}(t;a^*) + RR_C p_C^{\text{obs}}(t;a^*) + RR_F p_F^{\text{obs}}(t;a^*)}, \tag{3.2}$$
where a^* is the age class to which a belongs, and $\delta_{\text{pop}}(t;a)$ is the mortality rate in the population of age a at time t (`http://dati.istat.it`), while RR_C and RR_F are the relative risks of dying for current smokers and former smokers versus never smokers. A similar procedure is proposed also in Ngalesoni et al. [23].

2. Let $p(t;a^*,\boldsymbol{\theta}) = \big(p_C(t;a^*,\boldsymbol{\theta}), p_N(t;a^*,\boldsymbol{\theta}), p_F(t;a^*,\boldsymbol{\theta})\big)$ be the vector of the prevalence of never, current and former smokers belonging to the class of age a^* at time t, predicted by the model in Equation (3.1), given a specific value of the parameters $\boldsymbol{\theta}$. We perform a calibration procedure to find the value of $\boldsymbol{\theta}$ that leads to predicted prevalence values as close as possible to the observed ones derived from the AVQ survey. To compare observed and simulated prevalence, we consider the following objective function, where $H(\cdot,\cdot)$ denotes the Hellinger distance [31] between two discrete probability distributions:
$$\begin{aligned} Obj(\boldsymbol{\theta}) &= \frac{1}{T \times A^*}\sum_{t,a^*} H\left(p(t;a^*,\boldsymbol{\theta}), p^{obs}(t;a^*)\right) \\ &= \frac{1}{T \times A^* \times \sqrt{2}}\sum_{t,a^*}\sqrt{\sum_{k\in\{C,N,F\}}\left(\sqrt{p_k(t;a^*,\boldsymbol{\theta})} - \sqrt{p_k^{obs}(t;a^*)}\right)^2}. \end{aligned} \tag{3.3}$$

Calibration is conducted by comparing observed and predicted prevalence across three distinct time frames: the entire period from 1993 to 2019, and the two sub-periods 1993-2004, and 2005-2019.

3.3.3 Bootstrap

The sampling variability around point estimates is quantified by using a parametric bootstrap procedure [32, 33]. Let $\hat{\boldsymbol{\theta}}$ be the vector of parameters minimizing the objective function in Equation (3.3) and $p(t;a^*,\hat{\boldsymbol{\theta}})$ the corresponding estimated vector of prevalence for never, current

and former smokers of age a^* in the population at time t. Let $n(t;a^*)$ be the number of subjects belonging to the age class a^*, enrolled in the AVQ survey in the year t in Tuscany (i.e. the denominator of the observed prevalence $p^{obs}(t;a^*)$). The bootstrap procedure consists in:

1. Sampling, for each a^* and t, a vector of prevalence from a Dirichlet distribution: $p^b(t;a^*) \sim Dirichlet\Big(p_C(t;a^*,\hat{\boldsymbol{\theta}})n(t;a^*), p_N(t;a^*,\hat{\boldsymbol{\theta}})n(t;a^*), p_F(t;a^*,\hat{\boldsymbol{\theta}})n(t;a^*)\Big)$;
2. Considering this sampled vector as the observed one, computing $\boldsymbol{\delta}^b(a,c)$ and finding $\boldsymbol{\theta}^b$ that minimizes the objective function, as described in Section 3.3.2.

In [25], a sample of 1,000 bootstrap estimates has been used to calculate the 90% confidence intervals around model parameters and other quantities of interest, including compartment size trajectories.

3.3.4 Tobacco control policies assessment and forecasting

We assess and compare the impact of four TCPs, assuming their full implementation since 2023:

- **TCP0, No policy implemented.** We assume that in the absence of policies, smoking habit dynamics would remain unchanged;
- **TCP1, Tobacco-free generation.** Smoking consumption is banned for children born from 2009 onwards [34]. The policy effect depends on the parameter χ_1^{TCP1} denoting the percentage of individuals that circumvent the prohibition;
- **TCP2, Tax policy.** An annual 10% increase in cigarettes price is introduced for 10 years [24]. The policy effect depends on the age-specific price elasticities for the probabilities of starting and stopping smoking. The elasticities for the probability of stopping smoking are denoted by the parameter vector $\chi^{\mathbf{TCP2}} = (\chi_{<18}^{TCP2}, \chi_{18-24}^{TCP2}, \chi_{25-34}^{TCP2}, \chi_{35-64}^{TCP2}, \chi_{65+}^{TCP2})$; the elasticities for the probability of starting smoking by the vector $-\chi^{\mathbf{TCP2}}$;
- **TCP3, Cessation treatment policies.** Pharmaceutical treatments, quitlines, and brief interventions for smoking cessation are made completely available and reimbursed. The policy effect is described by the parameter vector $\chi^{\mathbf{TCP3}} = (\chi_1^{TCP3}, \chi_2^{TCP3})$, where χ_1^{TCP3} is the probability of stopping smoking induced by the pharmaceutical intervention, and χ_2^{TCP3} is the percentage of smokers, assumed to be constant over the years, that adheres to the program;
- **TCP4, Marketing restrictions.** Bans on all forms of direct and indirect advertising are enforced. The policy effect is described by the parameter vector $\chi^{\mathbf{TCP4}} = (\chi_1^{TCP4}, \chi_2^{TCP4})$, where χ_1^{TCP4} is the reduction in the probability of starting smoking induce by the policy and χ_1^{TCP4} is the increase in the probability of smoking cessation.

For each policy, we calculate the evolution of smoking prevalence and Smoking-Attributable Deaths (SADs). SADs in the year t are defined as the difference between the number of deaths we would observe under a counterfactual condition where current smokers and former smokers in the year t were never smokers and the number of deaths occurring assuming the policy implementation. We also calculate the number of avoided deaths expected from the implementation of each policy with respect to TCP0, defined as the difference between the number of deaths occurring under TCP0 and the one we would observe under the specific counterfactual condition defined by the policy.

3.4 Robustification procedure

3.4.1 Probability distributions of inputs

Eliciting the inputs' probability distributions is preliminary to uncertainty analysis and GSA. We define the distribution on the basis of information from the epidemiological literature, substantive knowledge of the topic, and national surveys, as reported in Table 3.1. Three different kinds of inputs are identified: parameters of the model 3.1, parameters that define the TCPs, described in Section 3.3.4, and two triggers introduced to assess the impact of the calibration period and of the policy effect. Note that the distributions of mortality RRs for former smokers do not distinguish among levels of smoking intensity. The Uniform distributions defined on the TCPs parameters are centered around the effects and price elasticities used in a recent paper by Sánchez-Romero et al. [21].

TABLE 3.1: Distribution of the input parameters used in the GSA and related data sources.

	Parameters	Model	Distribution	Source
Parameters of compartmental model	$\boldsymbol{\pi} = (\pi_{C_l}, \pi_{C_m}, \pi_{C_h})$	males	$\boldsymbol{\pi} \sim \text{Dirichlet}(205^1 \times (0.19, 0.40, 0.41))$	AVQ survey
		females	$\boldsymbol{\pi} \sim \text{Dirichlet}(216^1 \times (0.19, 0.40, 0.41))$	
	ν	males	$\nu \sim \text{Poisson}(14,701)$	ISTAT
		females	$\nu \sim \text{Poisson}(13,895)$	
	$\boldsymbol{RR}_C = (RR_{C_l}, RR_{C_m}, RR_{C_h})$	males	$\log RR_{C_l} \sim \mathcal{N}_{trun}(\log 1.91, 0.03, 0, \infty)$	Thun et al. [35]
			$\log RR_{C_m} \mid \log RR_{C_l} = a \sim \mathcal{N}_{trun}(\log 2.05, 0.03, a, \infty)$	
			$\log RR_{C_h} \mid \log RR_{C_m} = a \sim \mathcal{N}_{trun}(\log 2.42, 0.02, a, \infty)$	
		females	$\log RR_{C_l} \sim \mathcal{N}_{trun}(\log 1.47, 0.03, 0, \infty)$	
			$\log RR_{C_m} \mid \log RR_{C_l} = a \sim \mathcal{N}_{trun}(\log 1.87, 0.03, a, \infty)$	
			$\log RR_{C_h} \mid \log RR_{C_m} = a \sim \mathcal{N}_{trun}(\log 2.36, 0.03, a, \infty)$	
	$RR_F = RR_{F_l}(c) = RR_{F_m}(c) = RR_{F_h}(c) = RR_F(c)$	males	$\log RR_F(c) \sim \mathcal{N}_{trun}(\log 2.53, 0.03, 0, \infty),\ c = 1$	Thun et al. [35]
			$\log RR_F(c) \mid \log RR_F(1) = a \sim \mathcal{N}_{trun}(\log 2.35, 0.03, 0, a),\ c \in \{2,3,4\}$	
			$\log RR_F(c) \mid \log RR_F(2) = a \sim \mathcal{N}_{trun}(\log 1.90, 0.02, 0, a),\ c \in \{5,\ldots,9\}$	
			$\log RR_F(c) \mid \log RR_F(5) = a \sim \mathcal{N}_{trun}(\log 1.49, 0.02, 0, a),\ c \geq 10$	
		females	$\log RR_F(c) \sim \mathcal{N}_{trun}(\log 2.26, 0.05, 0, \infty),\ c = 1$	Thun et al. [35]
			$\log RR_F(c) \mid \log RR_F(1) = a \sim \mathcal{N}_{trun}(\log 2.22, 0.04, 0, a),\ c \in \{2,3,4\}$	
			$\log RR_F(c) \mid \log RR_F(2) = a \sim \mathcal{N}_{trun}(\log 1.58, 0.03, 0, a),\ c \in \{5,\ldots,9\}$	
			$\log RR_F(c) \mid \log RR_F(5) = a \sim \mathcal{N}_{trun}(\log 1.29, 0.03, 0, a),\ c \geq 10$	
	$\boldsymbol{RR}_{status} = (RR_C, RR_F)$	males	$\log RR_C \sim \mathcal{N}_{trun}(\log 2.43, 0.01, 0, \infty)$	Thun et al. [35]
			$\log RR_F \mid \log RR_C = a \sim \mathcal{N}_{trun}(\log 1.43, 0.01, 0, a)$	
		females	$\log RR_C \sim \mathcal{N}_{trun}(\log 2.08, 0.01, 0, \infty)$	Thun et al. [35]
			$\log RR_F \mid \log RR_C = a \sim \mathcal{N}_{trun}(\log 1.28, 0.01, 0, a)$	
Parameters of TCPs	χ^{TCP1}		$\chi^{TCP1} \sim \mathcal{U}[0, 0.10]$	
	$\boldsymbol{\chi}^{TCP2} = (\chi^{TCP2}_{<18}, \chi^{TCP2}_{18-24}, \chi^{TCP2}_{25-34}, \chi^{TCP2}_{35-64}, \chi^{TCP2}_{65+})$		$\chi^{TCP2}_{<18} \sim \mathcal{U}[0.35, 0.45]$	Sánchez-Romero et al. [21]
			$\chi^{TCP2}_{18-24} \sim \mathcal{U}[0.25, 0.35]$	
			$\chi^{TCP2}_{25-34} \sim \mathcal{U}[0.15, 0.25]$	
			$\chi^{TCP2}_{35-64} \sim \mathcal{U}[0.05, 0.15]$	
			$\chi^{TCP2}_{65+} \sim \mathcal{U}[0.15, 0.25]$	
	$\boldsymbol{\chi}^{TCP3} = (\chi^{TCP3}_1, \chi^{TCP3}_2)$		$\chi^{TCP3}_1 \sim \mathcal{U}[0.2744, 0.3144]$	Sánchez-Romero et al. [21]
			$\chi^{TCP3}_2 \sim \mathcal{U}[0.20, 0.50]$	
	$\boldsymbol{\chi}^{TCP4} = (\chi^{TCP4}_1, \chi^{TCP4}_2)$		$\chi^{TCP4}_1 \sim \mathcal{U}[0.06, 0.10]$	Sánchez-Romero et al. [21]
			$\chi^{TCP4}_2 \sim \mathcal{U}[0.02, 0.06]$	
Triggers	τ^{TCP}		$\tau^{TCP} \sim \mathcal{U}\{TCP0, TCP1, TCP2, TCP3, TCP4\}$	
	τ^{window}		$\tau^{window} \sim \mathcal{U}\{1993-2004, 2005-2019, 1993-2019\}$	

[1] Average sample size of the AVQ survey in the period from 1993 to 2019.

3.4.2 Sensitivity Analysis procedure

In this paragraph, we illustrate the procedure that transforms the inputs of the GSA function into outputs of interest [26]. We denote with $f(\boldsymbol{\pi}, \nu, \boldsymbol{RR}, \tau^{window}, \tau^{TCP}, \chi^{TCP1}, \boldsymbol{\chi}^{TCP2}, \boldsymbol{\chi}^{TCP3}, \boldsymbol{\chi}^{TCP4})$ the GSA function, where $\boldsymbol{RR} = (RR_{C_l}, RR_{C_m}, RR_{C_h}, RR_{F_l}, RR_{F_m}, RR_{F_h}, RR_C, RR_F)$ is the vector of the relative risks. Note that depending on the purpose of the analysis (assessing the global impact of TCPs on the model output or establishing a ranking among TCPs), we sample a specific TCP through the trigger τ^{TCP} or apply since 2023 all five policy scenarios. For simplicity we call the first option *procedure* 1 and the second option *procedure* 2.

In detail, we perform, separately for males and females, the following Monte Carlo procedure, following the philosophy of boot-strapping of the modeling process [36, 37]:

1. We independently sample the following model inputs from the distributions defined in Table 3.1:
   (a) values for $\boldsymbol{\pi}$, ν, and $\boldsymbol{RR}$;
   (b) a trigger τ^{window} that selects the time window of the observed data on which the calibration is performed;

(c) a trigger τ^{TCP} that selects which TCP has to be implemented since 2023 (this step is not performed in *procedure* 2);

(d) values for all the policy parameters χ^{TCP1}, $\boldsymbol{\chi}^{TCP2}$, $\boldsymbol{\chi}^{TCP3}$, $\boldsymbol{\chi}^{TCP4}$ (*procedure* 2) or only for the parameters of the policy selected (*procedure* 1);

2. We make an inference on the unknown vector of parameters of the SHC model, $\boldsymbol{\theta}$, given the sampled values for $\boldsymbol{\pi}$, ν, and $\boldsymbol{RR}$ and given the selected calibration window;

3. Once we have found the optimal $\boldsymbol{\theta}$, we run a bootstrap step, according to the procedure described in Section 3.3.3. We save the bootstrap estimate and use it in the upcoming step;

4. We get a realization from the stochastic SHC model, sampling from the appropriate binomial and multinomial distributions described in Section 3.3.1.1, under the sampled TCP (*procedure* 1) or under all TCPs (*procedure* 2).

We sample the different combinations of the model inputs adopting a quasi-random number sampling which provides a more efficient exploration of the sample space [28, 38]. At the end of this procedure, we obtain K realizations from the SHC model, as well as K estimates of the curves $\gamma(a)$, $\varepsilon(a)$, $\eta(c)$, that describe the transition probabilities between compartments, and of the age-specific mortality rates $\delta_N(a)$. We summarized these quantities in terms of pointwise median and 5th and 95th percentiles.

When the pathways of the compartmental model are generated under all TCPs, as in *procedure*2, we also obtain a distribution of policy rankings, which we summarize in terms of the surface under the cumulative ranking (SUCRA) curve. The SUCRA curve is commonly used in network meta-analysis to numerically present the overall ranking of alternative treatments. It is a single number ranging from 0 to 100. The higher the SUCRA value, the higher the likelihood that a treatment is one of the top ranks; the closer to 0 the SUCRA value, the more likely that the treatment is one of the bottom ranks. Rankings are defined based on the projected prevalence of never and current smokers in 2053 and 2063, as well as on SADs for individuals over 35 years old and 65 years old in the same years.

3.4.3 Outputs definition and Total Index calculation

We compute the total index defined for the following outputs:

- Hellinger distance;
- Maximum probabilities of starting and quitting smoking, and age at which the maxima are reached;

- Maximum probability of relapsing into smoke;
- Mean probability of starting, quitting, and relapsing smoking;
- Mean age of starting and quitting smoking;
- Mean time from smoking cessation to relapse;
- Expected prevalence of never, current, and former smokers, evaluated every ten years from 2013 to 2063;
- Expected minimum and maximum prevalence of current and never smokers over the study period;
- Expected SAD for over 35 and 65 years old, evaluated every ten years from 2013 to 2063.
- Expected shifting in the absolute value between the TCPs ranking for a specific combination of model inputs and a reference ranking (it serves as a summary output to capture the impact of the inputs on the TCPs ranking).

Note that the calculation of the sensitivity indexes is based on the deterministic model, neglecting the binomial or multinomial variability around model transitions. In this way, we capture the impact of the inputs on the expected prevalence and SADs, instead of on the single realization from the stochastic process generating the SHC model transitions.

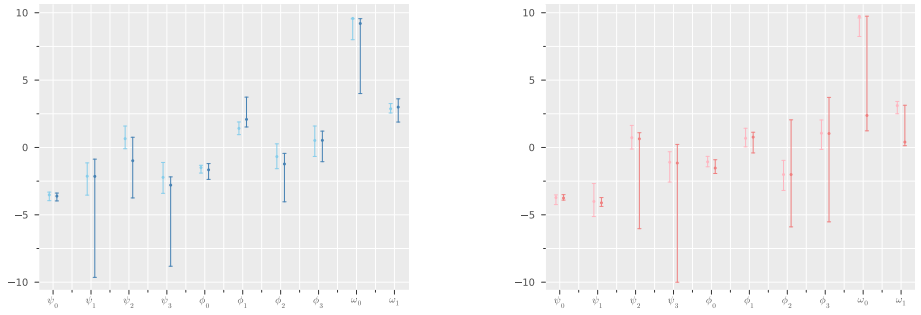
From a computational standpoint, we calculate the Sobol's indexes, starting from the $5'000$ simulations, following the design matrices-based method proposed in [26]. According to this design, we calculate the total effect indices relying on the results of 35,000 and 50,000 uncertainty propagation procedures, respectively for *procedure* 1 and *procedure* 2.

3.5 Results

All the analyses conducted in this paper are performed with JULIA language `https://julialang.org`. In particular, we use the package `Optim.jl` [39] for the calibration of the SHC model and the package `GlobalSensitivity.jl` [40] for the Sobol's total indices calculation.

Firstly, we focus on the unknown vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\omega})$ governing the SHC model. In Figure 3.2 we provide a comparison of the model estimates obtained in Lachi et al. [25], where the only source of uncertainty was the sampling variability, and the estimates obtained via the SA procedure described in Section 3.4.2. Results highlight that central point estimates of the classical inference coincide with the estimates produced by the uncertainty analysis. As expected,

the SA uncertainty intervals are much larger than the confidence intervals found in Lachi et al. [25].



a) b)

FIGURE 3.2: Results of estimation procedure for males in blue and females in red, with their bootstrap 90% confidence intervals (light color) obtained in Lachi et al. [25] in comparison to the same version obtained performing the procedure proposed in Section 3.2 (dark color): parameters tuning the probabilities of starting ($\boldsymbol{\psi}$) and stopping smoking ($\boldsymbol{\phi}$), and the probability of smoking relapse ($\boldsymbol{\omega}$) for male (a) and female (b).

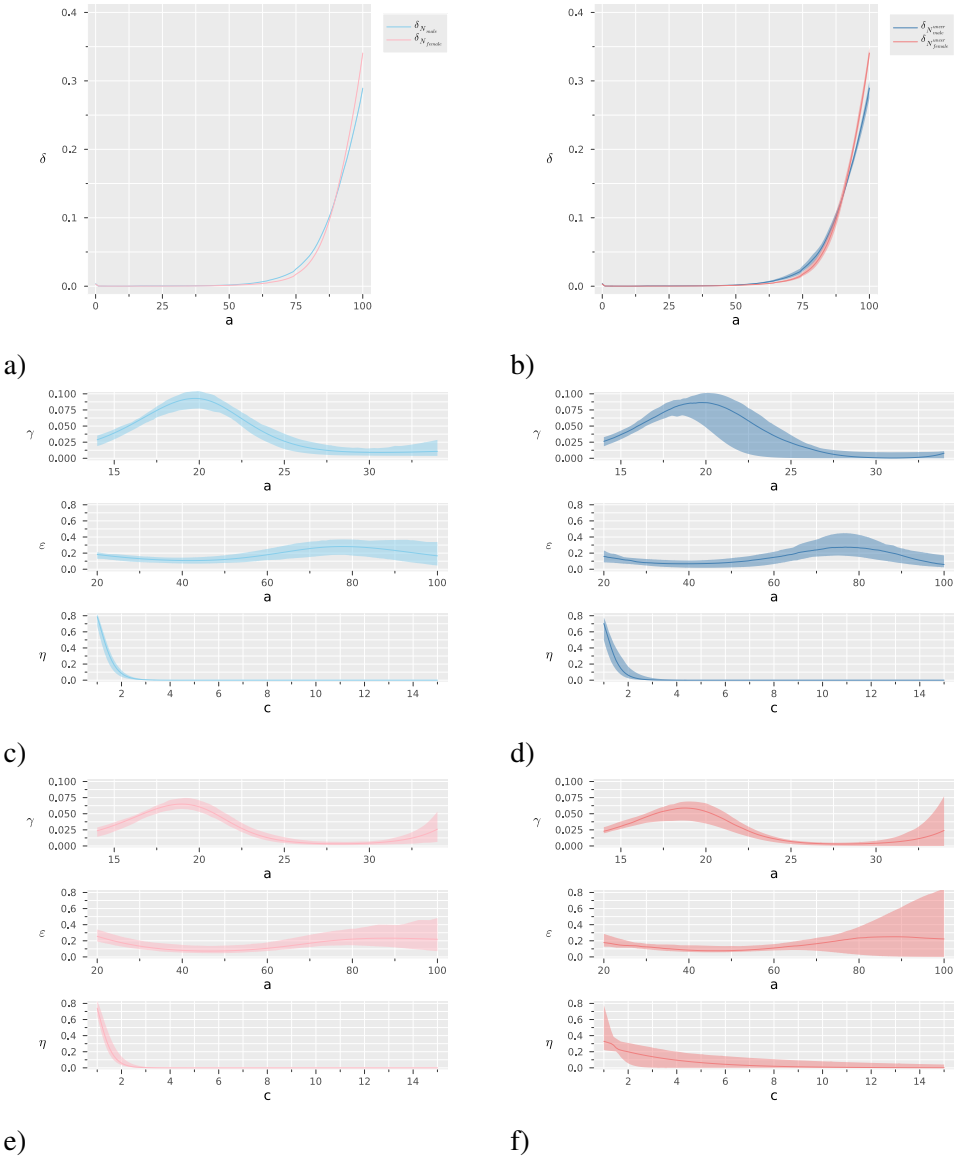
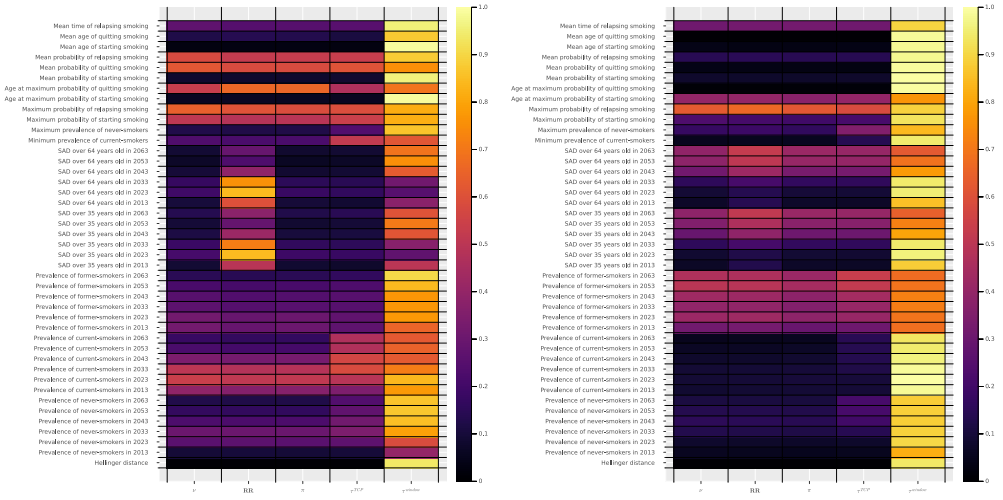


FIGURE 3.3: Results of estimation procedure for males in blue and females in red, with their bootstrap 90% confidence intervals (light color) obtained in Lachi et al. [25] in comparison to the same version obtained performing the procedure proposed in Section 3.2 (dark color): age-specific mortality for never smokers and the general population in (a-b), probabilities of starting ($\gamma(a)$) and stopping smoking ($\varepsilon(a)$), and probability of smoking relapse ($\eta(c)$) for male (c-d) and female (e-f).

In Figure 3.3, we compare the estimates of the transition probabilities of starting and quitting smoking and smoking relapse obtained in [25] with the SA estimates. The error propagation does not affect central point estimates of the transition probabilities but leads to larger confidence bands.

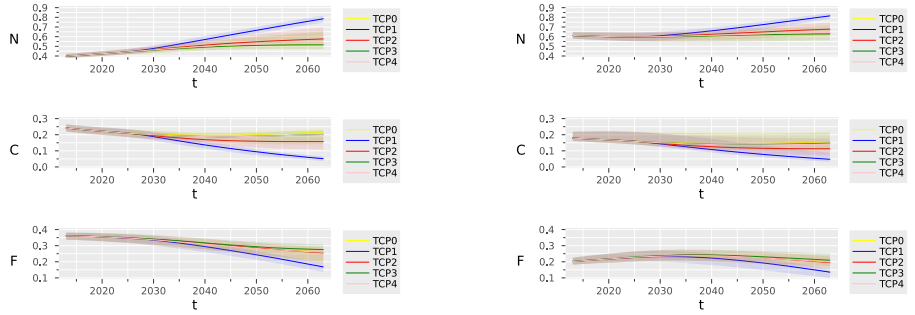


a) b)

FIGURE 3.4: Total variance indices that quantify the contribution of each input on the model output, calculated for the model of males (a) and females (b), based on *procedure* 1 defined in Section 3.2.

The heatmap in Figure 3.4 shows the total variance indexes deriving from *procedure* 1 (see Section 3.4.2), separately for males (a) and females (b). For each output defined in Section 3.4.3, we calculate as many indexes as there are inputs in the SA function. Note that the trigger for TCP captures the global effect of the policies on the outputs. Looking at the Hellinger distance, all the variability can be attributed to the choice of the calibration window. On the one hand, this result indicates that inference is not affected by the values assigned to $\boldsymbol{RR}$, ν, and $\boldsymbol{\pi}$ in the SHC model. On the other hand, suggests that the inference results are very sensitive to the calibration window. The calibration window is the most relevant input also for the outputs regarding the probabilities of starting and quitting smoking and smoking relapse. However, for males, a certain effect of other inputs is also observed. Focusing on the evolution of the prevalence of current smokers from 2013 onward, most of the variability (almost all for females) can be attributed to the choice of the calibration window and the TCPs. For the other prevalence, the calibration window remains the most relevant model input. Regarding the evolution of SAD, in addition to the calibration window, $\boldsymbol{RR}$ represents an important model input in particular for males before 2033. For numerical details see Section C.2 in Supplemental Materials.

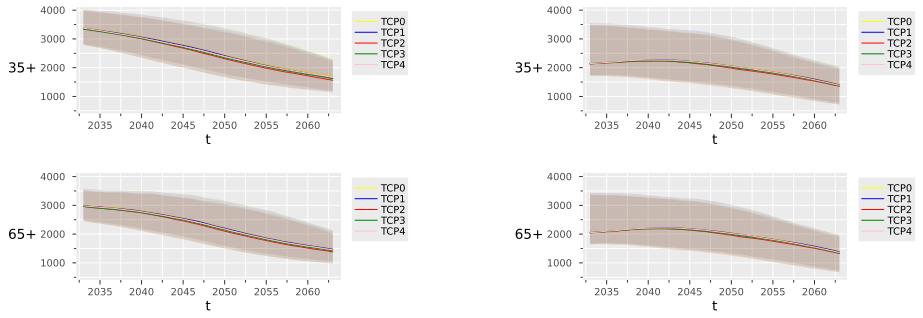
Figure 3.5 compares the evolution of smoking habits among males (a) and females (b) under the five alternative TCPs defined in Section 3.3.4. TCPs have no substantial effect on the prevalence of never, former, and current smokers during the 10 years following their implementation in 2023. After 2033, TCP1 has the largest impact in increasing and reducing respectively the prevalence of never and current smokers, followed by TCP2. TCP3 has the largest impact on increasing the prevalence of former smokers.



a) b)

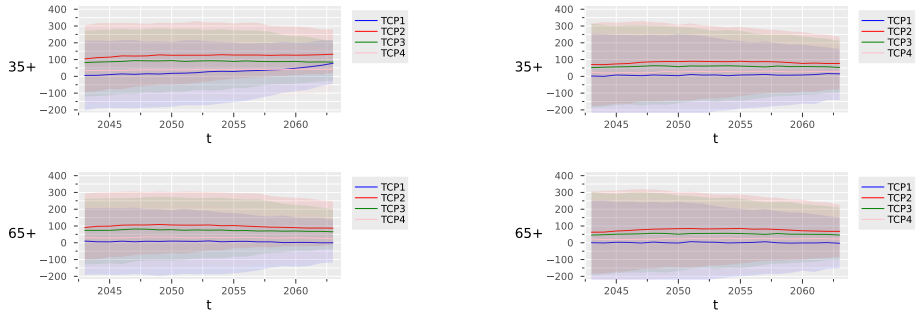
FIGURE 3.5: Estimated prevalence of never (N), current (C), and former (F) smokers under different tobacco control policies (TCP) with their 90% uncertainty intervals, for males (a) and females (b) based on a stochastic model.

Figure 3.6 reports the predicted number of attributable deaths for the age classes 35+ and 65+, both for males (a) and females (b), under different TCPs from 2033 onward. The uncertainty around the SADs is very high and the uncertainty bands largely overlap, indicating that within the time window considered we cannot expect one policy to prevail over the other in terms of reducing attributable mortality. Figure 3.7 reports the saved deaths under TCP1, TCP2, TCP3, and TCP4, in reference to the no-policy scenario TCP0. Examining SADs among individuals over 35 years old, it is evident that TCP1 leads to the prevention of a greater number of deaths over the long term. Conversely, within the considered time window, TCP2 is the policy that, on average, prevents more deaths. However, it is important to note that the uncertainty intervals are very wide and encompass zero.



a) b)

FIGURE 3.6: Estimated Smoking Attributable Deaths (SAD) among people over 35 and 65 years old, with 90% uncertainty intervals, for males (a) and females (b) under different tobacco control policies (TCP) based on a stochastic model.



a) b)

FIGURE 3.7: Decrease of the number of Smoking Attributable Deaths (SAD) among people over 35 and 65 years old, with 90% uncertainty intervals, for males (a) and females (b) under different tobacco control policies (TCP) based on a stochastic model.

Adopting the *procedure* 2 described in Section 3.4.2, for each different combination of the other inputs, we rank the policies from the most efficient (rank 1) to the least efficient (rank 4) based on various outputs. Figures 3.8 and 3.9 show, for males and females respectively, the probability that each policy has a rank$\leq r$, where $r \in \{1,2,3,4\}$ represents all the possible ranking positions. The corresponding SUCRA values are reported in the legend. Figure 3.8 and 3.9 show that TCP1 is always the best when we look at the prevalence of current smokers, while TCP2 is the best in terms of SAD reduction, according to Figure 3.7.

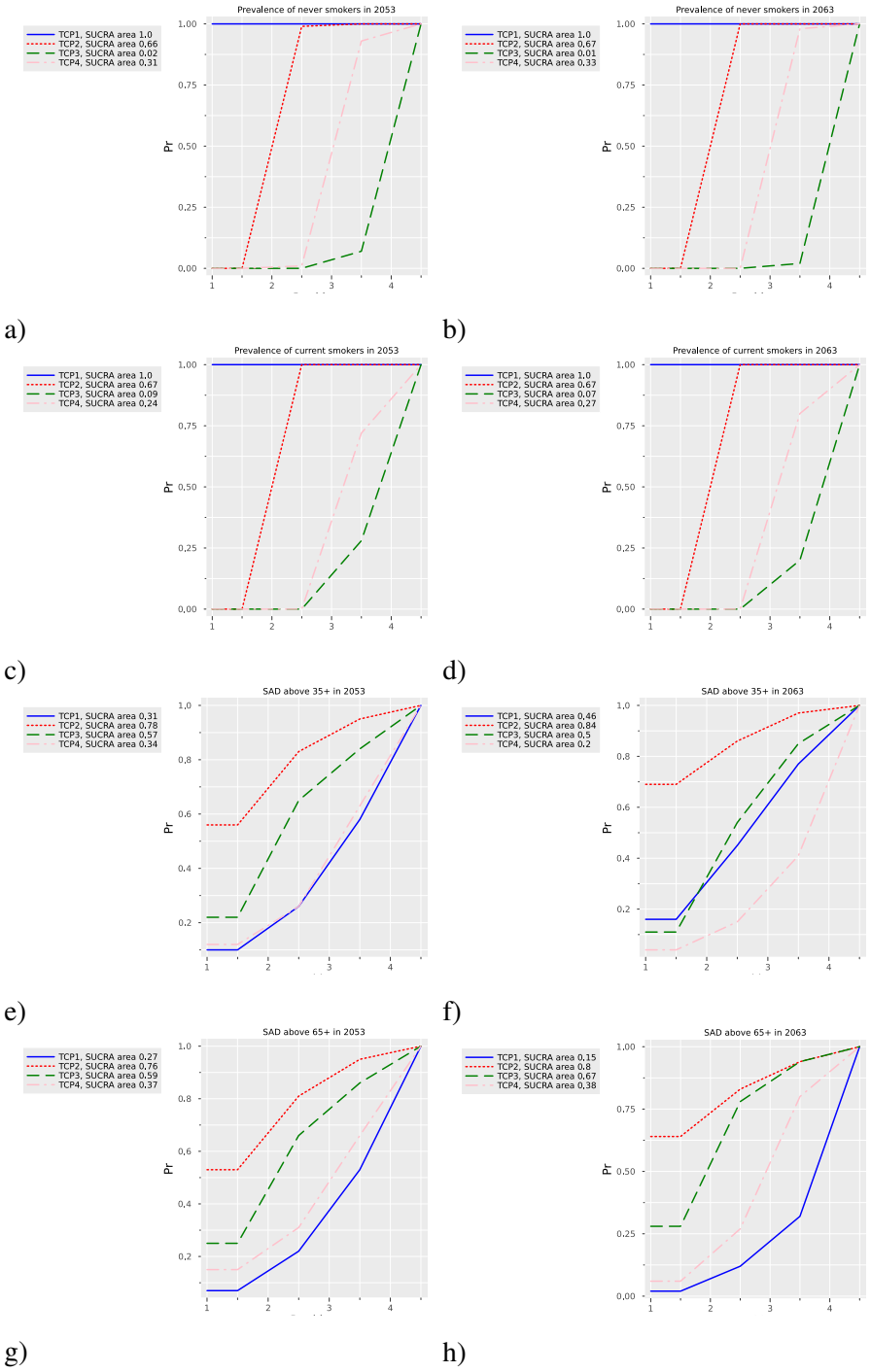


FIGURE 3.8: Ranking of Tobacco Control Policies (TCP) for the prevalence of never smokers (a-b), current smokers (c-d), Smoking Attributable Deaths (SAD) among people over 35 years old (e-f), and SAD among people over 65 years old (g-h), reached in 2053 and 2063 for males based on a stochastic model.

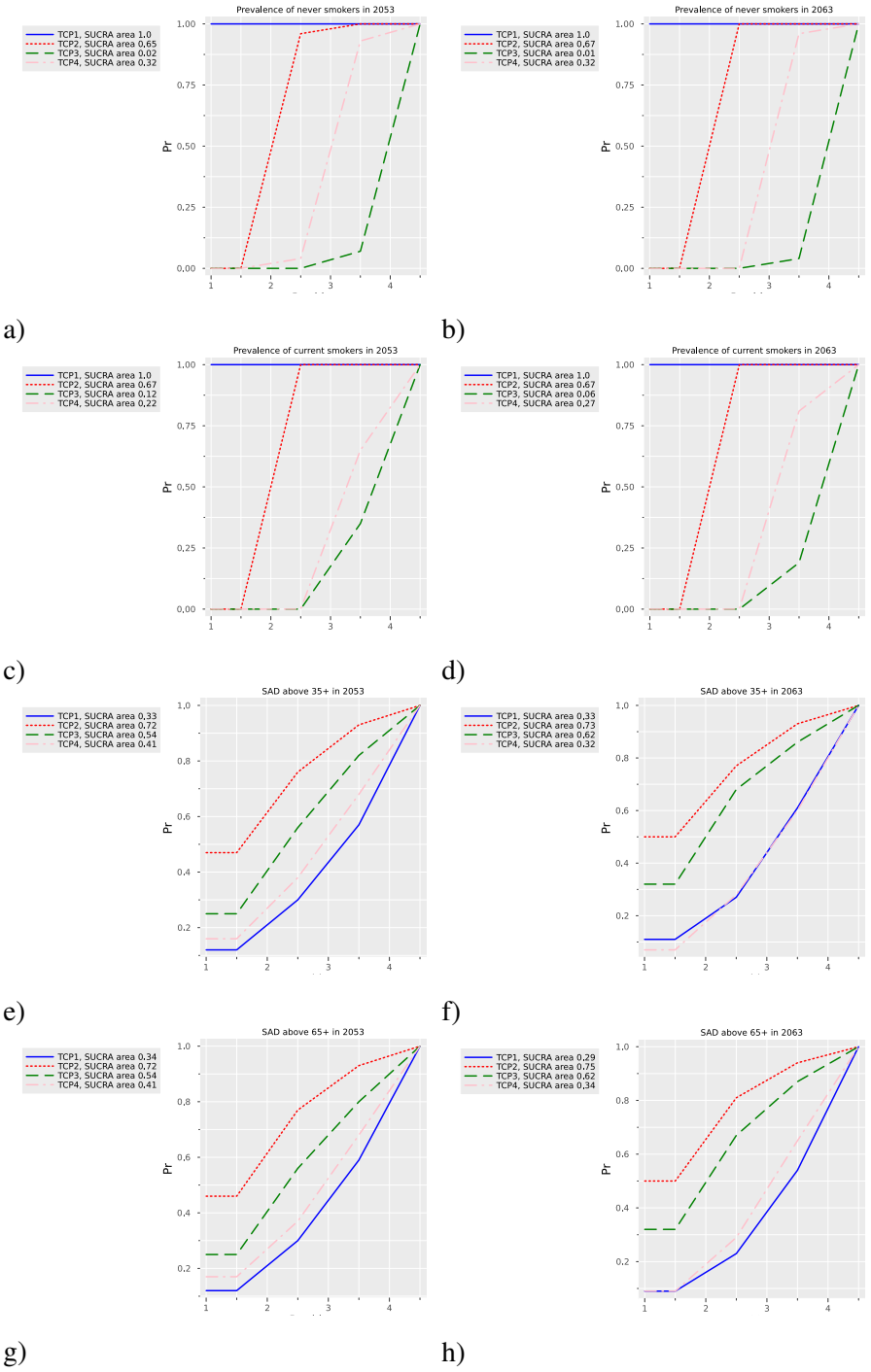
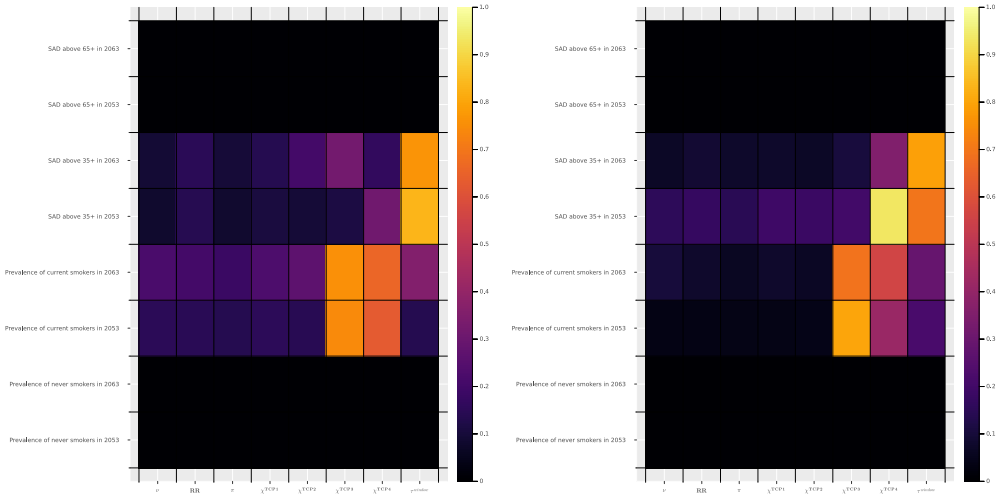


FIGURE 3.9: Ranking of Tobacco Control Policies (TCP) for the prevalence of never smokers (a-b), current smokers (c-d), Smoking Attributable Deaths (SAD) among people over 35 years old (e-f), and SAD among people over 65 years old (g-h), reached in 2053 and 2063 for females based on a stochastic model.

The heatmaps in Figure 3.10 show, for males and females, the total variance indexes measuring the impact of each input, including the policy parameters, on TCP rankings based on current and never smokers prevalence and SADs in 2053 and 2063. The inputs do not affect the rankings based on never smokers prevalence and SADs among individuals over 65 years old. Regarding

the ranking based on the prevalence of current smokers, most of the variability is due to the parameters governing TCP3 and TCP4. For males, the ranking based on SADs among individuals over 35 years old, the variability is greatly due to the calibration window. For females, it is also attributable to the parameters governing TCP4. For numerical details see Section C.2 in Supplemental Materials.



a) b)

FIGURE 3.10: Total variance indices that quantify the contribution of each input on the model output, calculated for the model of males (a) and females (b), based on *procedure* 2 defined in Section 3.2.

3.6 Discussion and conclusion

Compartmental models provide a useful framework for understanding the behavior of complex systems, helping to study the evolution of populations over time. Additionally, due to their mechanistic nature, compartmental models are especially well-suited for projections. However, they suffer from limitations associated with relying on sometimes strong structural assumptions and being typically governed by numerous parameters, posing identification issues. Therefore, while compartmental models are powerful tools for investigating complex phenomena and generating projections to explore future public health dynamics, they may lack robustness when assumptions are misspecified, resulting in highly uncertain outcomes. In this context, uncertainty analysis and GSA can be used to propagate uncertainty, perform an overall assessment of the impact of model assumptions on results and forecasting, as well as detect and quantify the contribution of different sources of variability on outputs of interest [26, 41, 42].

In this paper, we provide a robustification of inference and projections produced by the SHC model, a compartmental model that simulates the evolution of smoking dynamics in Tuscany, a region of central Italy [25]. Note that the model exclusively focuses on the use of conventional

cigarettes and does not take into account other tobacco products such as electronic cigarettes and heated tobacco products.

In order to set up the sensitivity analysis, it is necessary to define the SA model, along with its inputs and outputs. In our case, the SA model involves the compartmental model specified in its stochastic form, calibration, and bootstrap procedures. Following the model taxonomy introduced by [26], our model has both diagnostic and prognostic purposes because it is used to understand a law, but also to predict the behavior of a system given a supposedly understood law. Similarly, it can be considered both a data-driven and a law-driven model, because it uses data to derive the system's properties statistically, and it uses laws that have been attributed to the system to predict its behavior.

Through Monte Carlo simulations, we propagate uncertainty from the analysis inputs to the outputs obtaining a more comprehensive assessment of variability than what is achieved using standard approaches. Then, we detect subsets of data and variables that seriously influence the analysis results, by calculating the total variance indexes provided by GSA. This approach is known in the literature under the name of "modeling of the modeling process", a radical version of sensitivity analysis proposed by Piano and Benini [27]. The point is to make sure that a model that is not meaningful is not eventually used to make a decision. It is crucial to emphasize that these analyses may indicate that the model is unsuitable or unreliable to produce relevant outcomes.

Our analysis substantially confirms the inference results on the model parameters obtained by using a standard approach that considers sampling variability as the unique source of uncertainty Lachi et al. [25]. However, it also highlights that the estimates of the probability of starting and quitting smoking and the probability of smoking relapse, as well, albeit to a lesser extent, the forecasts, strongly depend on the specific calibration window used, indicating that the fundamental assumption that transitions between compartment do not change over time is not correct. In fact, this assumption is not without issues. For example, a decreasing trend in the probability of starting smoking has been reported for both males and females in Europe [43], while evidence of a dependence between age and risk of smoking relapse has been found in the US population [44]. However, if introducing multiple time-axes dependence in the transition probabilities could lead to more realistic results, this would be at the price of further complicating the model by introducing new unknown parameters to be estimated. Future work should try to include multivariate splines to model the dependence of the transition probabilities from calendar years.

Our analysis provides also an evaluation of alternative TCPs to be applied since 2023, as well as their ranking based on mortality and prevalence projections for the upcoming years. In particular, we focus on TCPs acting to promote tobacco demand-reduction, defined in light of Italy's MPOWER status in 2021 (`https://cdn.who.int/media/docs/default-source/`

country-profiles/tobacco/who_rgte_2021_italy.pdf). We do not consider policies aimed at protecting people from tobacco smoke, such as policies that mandate a very low nicotine content standard for cigarettes, or policies that ban or phase out the retail supply of tobacco. The sensitivity analysis results indicate that the TCP effects are influenced by considerable variability when considering all sources of uncertainty in the modeling process. However, it is possible to rank the TCPs based on specific outcomes, even though this ranking is subject to variability related to process stochasticity. The ranking naturally varies depending on the forecasting horizon. In particular, while the smoking-free generation policy is the most effective on average in the long term, policies that also impact smoking cessation prove more effective in the short term, especially when the focus is on reducing attributable deaths.

In conclusion, beyond the specific results related to the considered application, GSA represents an important useful tool, that should be always used to guide the development of increasingly complex models.

Authors' contributions: M.B., A.S., and A.L. conceptualized the project; A.L. wrote the codes and conducted the statistical analysis under the supervision of M.B. and A.S.; M.B. and A.L. wrote the first version of the paper; C.V. and A.S. contributed to the statistical discussion of the results; G.C. supported policies definition and contributed to the epidemiological discussion of the results; M.B. supervised the project. All the authors read and approved the final version of the paper.

Bibliography

[1] S. Rezaei, A.A. Sari, M. Arab, R. Majdzadeh, and A.M. Poorasl. Economic burden of smoking: a systematic review of direct and indirect costs. *Medical journal of the Islamic Republic of Iran*, 30:397, 2016.

[2] M.B. Reitsma, L.S. Flor, E.C. Mullany, V. Gupta, S.I. Hay, and E. Gakidou. Spatial, temporal, and demographic patterns in the prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. *The Lancet Public Health*, 6(7): e472–e481, 2021. ISSN 24682667. doi: 10.1016/S2468-2667(21)00102-X. URL https://linkinghub.elsevier.com/retrieve/pii/S246826672100102X.

[3] R. Roemer, A. Taylor, and J. Lariviere. Origins of the who framework convention on tobacco control. *American Journal of Public Health*, 95(6):936–938, 2005. doi: https://doi.org/10.2105/AJPH.2003.025908.

[4] 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*. Regional Committee for Europe, 65th session, 2015.

[5] D.T. Levy and K. Friend. A Simulation Model of Policies Directed at Treating Tobacco Use and Dependence. *Medical Decision Making*, 22(1):6–17, 2002. ISSN 00000000, 0272989X. doi: 10.1177/02729890222062874. URL http://mdm.sagepub.com/cgi/doi/10.1177/02729890222062874.

[6] D.T. Levy, S. Gallus, K. Blackman, G. Carreras, C. La Vecchia, and G. Gorini. Italy SimSmoke: the effect of tobacco control policies on smoking prevalence and smoking-attributable deaths in Italy. *BMC Public Health*, 12(1):709, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-709. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-709.

[7] A.M. Near, K. Blackman, L.M. Currie, and D.T. Levy. Sweden SimSmoke: the effect of tobacco control policies on smoking and snus prevalence and attributable deaths. *The European Journal of Public Health*, 24(3):451–458, 2014. ISSN 1101-1262, 1464-360X. doi: 10.1093/eurpub/ckt178. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckt178.

[8] D.T. Levy, L.M. Sánchez-Romero, Y. Li, Z. Yuan, N. Travis, M.J. Jarvis, J. Brown, and A. McNeill. England SimSmoke: the impact of nicotine vaping on smoking prevalence and smoking-attributable deaths in England. *Addiction*, 116(5):1196–1211, 2021. ISSN 0965-2140, 1360-0443. doi: 10.1111/add.15269. URL https://onlinelibrary.wiley.com/doi/10.1111/add.15269.

[9] L.M. Sánchez-Romero, L. Zavala-Arciniega, L.M. Reynales-Shigematsu, B.S. De Miera-Juárez, Z. Yuan, Y. Li, Y.K. Lau, N.L. Fleischer, R. Meza, J.F. Thrasher, and D.T. Levy. The Mexico SimSmoke tobacco control policy model: Development of a simulation model of daily and nondaily cigarette smoking. *PLOS ONE*, 16(6):e0248215, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0248215. URL https://dx.plos.org/10.1371/journal.pone.0248215.

[10] D.T. Levy, L. Nikolayev, E. Mumford, and C. 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*, 16(4):359–371, 2005. ISSN 0957-5243, 1573-7225. doi: 10.1007/s10552-004-7841-4. URL http://link.springer.com/10.1007/s10552-004-7841-4.

[11] D.T. Levy, J.E. Bauer, and H.R. Lee. Simulation Modeling and Tobacco Control: Creating More Robust Public Health Policies. *American Journal of Public Health*, 96(3):494–498, 2006. ISSN 0090-0036, 1541-0048. doi: 10.2105/AJPH.2005.063974. URL https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2005.063974.

[12] D.T. Levy, A.L. Graham, P.L. Mabry, D.B. Abrams, and C.T. Orleans. Modeling the Impact of Smoking-Cessation Treatment Policies on Quit Rates. *American Journal of Preventive Medicine*, 38(3):S364–S372, 2010. ISSN 07493797. doi: 10.1016/j.amepre.2009.11.016. URL https://linkinghub.elsevier.com/retrieve/pii/S0749379709008575.

[13] G. Carreras, S. Gallus, L. Iannucci, and G. Gorini. Estimating the probabilities of making a smoking quit attempt in Italy: stall in smoking cessation levels, 1986-2009. *BMC Public Health*, 12(1):183, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-183. URL http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-183.

[14] G. Carreras, G. Gorini, and E. 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*, 5(6):874–882, 2012. ISSN 1940-6207, 1940-6215. doi: 10.1158/1940-6207.CAPR-12-0019. URL https://aacrjournals.org/cancerpreventionresearch/article/5/6/874/49965/Can-a-National-Lung-Cancer-Screening-Program-in.

[15] G. Carreras, G. Gorini, S. Gallus, L. Iannucci, and D.T. Levy. Predicting the future prevalence of cigarette smoking in Italy over the next three decades. *European Journal of Public Health*, 22(5):699–704, 2012. ISSN 1464-360X, 1101-1262. doi: 10.1093/eurpub/ckr108. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckr108.

[16] E.J. Feuer, D.T. Levy, and W.J. McCarthy. Chapter 1: The Impact of the Reduction in Tobacco Smoking on U.S. Lung Cancer Mortality, 1975-2000: An Introduction to the Problem: Introduction: Impact of the Reduction in Tobacco Smoking on U.S. Lung Cancer Mortality. *Risk Analysis*, 32:S6–S13, 2012. ISSN 02724332. doi: 10.1111/j.1539-6924.2011.01745.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.2011.01745.x.

[17] D.T. Levy, H. Fouad, J. Levy, A.D. Dragomir, and F. El Awa. Application of the *Abridged SimSmoke* model to four Eastern Mediterranean countries. *Tobacco Control*, 25(4):413–421, 2016. ISSN 0964-4563, 1468-3318. doi: 10.1136/tobaccocontrol-2015-052334. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tobaccocontrol-2015-052334.

[18] D.T. Levy, R. Borland, A.C. Villanti., R. Niaura, Z. Yuan, Y. Zhang, R. Meza, T.R. Holford, G.T. Fong, K.M. Cummings, and D.B. Abrams. The Application of a Decision-Theoretic Model to Estimate the Public Health Impact of Vaporized Nicotine Product Initiation in the United States. *Nicotine and Tobacco Research*, 19(2):149–159, 2017. ISSN 1462-2203, 1469-994X. doi: 10.1093/ntr/ntw158. URL https://academic.oup.com/ntr/article-lookup/doi/10.1093/ntr/ntw158.

[19] J. Tam, D.T. Levy, J. Jeon, J. Clarke, S. Gilkeson, T. Hall, E.J. Feuer, T.R. Holford, and R. Meza. Projecting the effects of tobacco control policies in the USA through microsimulation: a study protocol. *BMJ Open*, 8(3):e019169, 2018. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2017-019169. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2017-019169.

[20] M. Sanna, W. Gao, Y.W. Chiu, H.Y. Chiou, Y.H. Chen, C.P. Wen, and D.T. Levy. Tobacco control within and beyond WHO MPOWER: outcomes from Taiwan SimSmoke. *Tobacco Control*, 29(1):36–42, 2020. ISSN 0964-4563, 1468-3318. doi: 10.1136/tobaccocontrol-2018-054544. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tobaccocontrol-2018-054544.

[21] L.M. Sánchez-Romero, Z. Yuan, Y. Li, and D.T. Levy. The Kentucky SimSmoke Tobacco Control Policy Model of Smokeless Tobacco and Cigarette Use. *International Journal of Health Policy and Management*, page 1, 2020. ISSN 2322-5939. doi: 10.34172/ijhpm.2020.187. URL https://www.ijhpm.com/article_3936.html.

[22] D.T. Levy, J. Tam, L.M. Sanchez-Romero, Y. Li, Z. Yuan, J. Jeon, and R. Meza. Public health implications of vaping in the USA: the smoking and vaping simulation model. *Population Health Metrics*, 19(1):19, 2021. ISSN 1478-7954. doi: 10.1186/s12963-021-00250-7. URL https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-021-00250-7.

[23] F. Ngalesoni, G. Ruhago, M. Mayige, T.C. Oliveira, B. Robberstad, O.F. Norheim, and H. Higashi. Cost-effectiveness analysis of population-based tobacco control strategies in the prevention of cardiovascular diseases in tanzania. *PloS one*, 12(8):e0182113, 2017. doi: https://doi.org/10.1371/journal.pone.0182113.

[24] H. Kim, S. Park, H. Kang, N. Kang, D.T. Levy, and S. Cho. Modeling the future of tobacco control: Using simsmoke to explore the feasibility of the tobacco endgame in korea. *Tobacco induced diseases*, 21, 2023. doi: https://doi.org/10.18332/tid/174127.

[25] A. Lachi, C. Viscardi, G. Cereda, G. Carreras, and M. Baccini. A compartmental models for smoking dynamics in Italy: A pipeline for inference, validation, and forecasting under hypothetical scenarios. preprint, In Review, 2023. URL https://www.researchsquare.com/article/rs-3303111/v1.

[26] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-05997-5.

[27] S. Lo Piano and L. Benini. A critical perspective on uncertainty appraisal and sensitivity analysis in life cycle assessment. *Journal of Industrial Ecology*, 26(3):763–781, June 2022. ISSN 1088-1980, 1530-9290. doi: 10.1111/jiec.13237. URL https://onlinelibrary.wiley.com/doi/10.1111/jiec.13237.

[28] I.M. Sobol. *A primer for the Monte Carlo method*. CRC Press, Boca Raton, 1994. ISBN 978-0-8493-8673-2.

[29] D.P. Kroese, T. Brereton, T. Taimre, and Z.I. Botev. Why the Monte Carlo method is so important today. *WIREs Computational Statistics*, 6(6):386–392, 2014. ISSN 1939-5108, 1939-0068. doi: 10.1002/wics.1314. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1314.

[30] R.T. Hoogenveen, P.Hm. Van Baal, H.C. Boshuizen, and T.L. Feenstra. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Effectiveness and Resource Allocation*, 6(1):1, 2008. ISSN 1478-7547. doi: 10.1186/1478-7547-6-1. URL http://resource-allocation.biomedcentral.com/articles/10.1186/1478-7547-6-1.

[31] E. Hellinger. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136):210–271, 1909. ISSN 1435-5345, 0075-4102. doi: 10.1515/crll.1909.136.210. URL https://www.degruyter.com/document/doi/10.1515/crll.1909.136.210/html.

[32] B. Efron and R. Tibshirani. *An introduction to the bootstrap*. Number 57 in Monographs on statistics and applied probability. Chapman and Hall, New York, 1993. ISBN 978-0-412-04231-7.

[33] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042717300234.

[34] O.D. Ait Ouakrim, T. Wilson, A. Waa, R. Maddox, H. Andrabi, S.R. Mishra, J.A. Summers, C.E. Gartner, R. Lovett, R. Edwards, N. Wilson, and T. Blakely. Tobacco endgame intervention impacts on health gains and Māori:non-Māori health inequity: a simulation study of the Aotearoa/New Zealand Tobacco Action Plan. *Tobacco Control*, pages tc–2022–057655, 2023. ISSN 0964-4563, 1468-3318. doi: 10.1136/tc-2022-057655. URL https://tobaccocontrol.bmj.com/lookup/doi/10.1136/tc-2022-057655.

[35] M.J. Thun, B.D. Carter, D. Feskanich, N.D. Freedman, R. Prentice, A.D. Lopez, P. Hartge, and S.M. Gapstur. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine*, 368(4):351–364, 2013. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMsa1211127. URL http://www.nejm.org/doi/10.1056/NEJMsa1211127.

[36] C. Chatfield. Model Uncertainty, Data Mining and Statistical Inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158(3):419–466, 1995. ISSN 09641998, 1467985X. doi: 10.2307/2983440. URL http://www.jstor.org/stable/2983440. Publisher: [Wiley, Royal Statistical Society].

[37] N. Breznau, E. Rinke, A. Wuttke, H. Nguyen, M. Adem, J. Adriaans, A. Alvarez-Benjumea, H. Andersen, D. Auer, F. Azevedo, O. Bahnsen, D. Balzer, G. Bauer, P. Bauer, M. Baumann, S. Baute, V. Benoit, J. Bernauer, C. Berning, and T. Żółtak. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, 119:e2203150119, 2022. doi: 10.1073/pnas.2203150119.

[38] S. Kucherenko, D. Albrecht, and A. Saltelli. Exploring multi-dimensional spaces: a comparison of latin hypercube and quasi monte carlo sampling techniques. *arXiv - University of Cornell (USA). JRC98050*, 2015. doi: https://doi.org/10.48550/arXiv.1505.02350.

[39] P.K. Mogensen and A.N. Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615, 2018. ISSN 2475-9066. doi: 10.21105/joss.00615. URL http://joss.theoj.org/papers/10.21105/joss.00615.

[40] V.K. Dixit and C. Rackauckas. GlobalSensitivity.jl: Performant and Parallel GlobalSensitivity Analysis with Julia. *Journal of Open Source Software*, 7(76):4561, 2022. ISSN 2475-9066. doi: 10.21105/joss.04561. URL https://joss.theoj.org/papers/10.21105/joss.04561.

[41] E.E. Leamer. Global sensitivity results for generalized least squares estimates. *Journal of the American Statistical Association*, 79(388):867–870, 1984.

[42] A. Saltelli, S. Tarantola, and F. Campolongo. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*, 15(4):377–395, 2000. ISSN 08834237. URL http://www.jstor.org/stable/2676831. Publisher: Institute of Mathematical Statistics.

[43] A. Marcon, G. Pesce, L. Calciano, V. Bellisario, S.C. Dharmage, J. Garcia-Aymerich, T. Gislasson, J. Heinrich, M. Holm, C. Janson, D. Jarvis, B. Leynaert, M.C. Matheson, P. Pirina, C. Svanes, S. Villani, T. Zuberbier, C. Minelli, S. Accordini, and the Ageing Lungs In European Cohorts study. Trends in smoking initiation in Europe over 40 years: A retrospective cohort study. *PLOS ONE*, 13(8):e0201881, 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201881. URL https://dx.plos.org/10.1371/journal.pone.0201881.

[44] A. Alboksmaty, I.T. Agaku, S. Odani, and F.T. Filippidis. Prevalence and determinants of cigarette smoking relapse among US adult smokers: a longitudinal study. *BMJ Open*, 9(11):e031676, 2019. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2019-031676. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2019-031676.

CHAPTER 4

An Annotated Timeline of Sensitivity Analysis

Paper published on *Environmental Modelling and Software*
doi: `https://doi.org/10.1016/j.envsoft.2024.105977`

Stefano Tarantola[1], Federico Ferretti[1], Samuele Lo Piano[2], Mariia Kozlova[3], Alessio Lachi[4], Rossana Rosati[1], Arnald Puy[5], Pamphile Roy[6], Giulia Vannucci[7], Marta Kuc-Czarnecka[8], Andrea Saltelli[9,10]

[1] European Commission, Joint Research Centre (JRC), Via Enrico Fermi 1 21027 - Ispra, Italy

[2] School of the Built Environment, University of Reading, RG6 6AF - Reading, United Kingdom

[3] LUT Business School, LUT University, Yliopistonkatu 34 53851 - Lappeenranta, Finland

[4] Department of Statistics, Computer Science, and Applications “Giuseppe Parenti" (DiSIA), University of Florence, Viale Giovan Battista Morgagni 59-65 50134 - Florence, Italy

[5] School of Geography, Earth and Environmental Sciences, University of Birmingham, B15 2TT - Birmingham, United Kingdom

[6] Quansight, LLC, 8656 W. Hwy 71 Bldg F200 78735 - Austin, USA

[7] Institute of Cognitive Science and Technology of the Italian National Institute Research Council (ICST-CNR), Via San Martino della Battaglia 44 00185 - Roma, Italy

[8] Department of Statistics and Econometrics Faculty of Management and Economics, Gdansk University of Technology, Traugutta 79 80-233 - Gdańsk, Poland

[9] UPF Barcelona School of Management (UPF-BSM), University Pompeu Fabra of Barcelona, Carrer de Balmes 132-134 08008 - Barcelona

[10] Centre for the Study of the Sciences and the Humanities, University of Bergen, Parkveien 9 PB 7805 5020 - Bergen, Norway

Abstract

The last half a century has seen spectacular progress in computing and modelling in a variety of fields, applications, and methodologies. Over the same period, a cross-disciplinary field known as sensitivity analysis has been making its first steps, evolving from the design of experiments for laboratory or field studies, also called in vivo, to the so-called experiments in silico. Some disciplines were quick to realize the importance of sensitivity analysis, whereas others are still lagging behind. Major tensions within the evolution of this discipline arise from the interplay between local versus global perspectives in the analysis as well as the juxtaposition of the mathematical complexification and the desire for practical applicability. In this work, we retrace these main steps with some attention to the methods and through a bibliometric survey to assess the accomplishments of sensitivity analysis and to identify the potential for its future advancement.

Keywords - Global Sensitivity Analysis, Local Sensitivity Analysis, Monte Carlo, History of Sensitivity Analysis, Design of Experiments

4.1 Introduction

Models simulate the real world by synthesizing a multitude of input configurations in their output, mapping potential present and future system states of interest. Their primary objective is to extract valuable insights regarding the relationship between inputs and outputs. Defining the nature of mathematical models is not easy, due to the variety of contexts and applications (Page 2018). Various authors identified modelling as an art [1] or a craft [2], with models being performative [3], and acting as mediators between theories and the world [4]. What remains undisputed is the remarkable development realised in computing and modelling in recent decades. Computer models are so widely used in a variety of fields, applications, and methodologies that they are seemingly affecting any aspect of our lives [4].

Together with modelling, a new field of research called sensitivity analysis has come to life, moving from the design of experiments for laboratory or field studies to experimental techniques performed by computers, namely the experiments in silico. While uncertainty analysis studies the uncertainty in the output, sensitivity analysis studies how the uncertainty in the output can be allocated to the different sources of uncertainty in the input [5–7]. In other words, sensitivity analysis elucidates the intimate relationship between the system output and its influential factors. It is easy to recognize the strong bond between sensitivity analysis and modelling. For some, sensitivity analysis, namely the drawing of the connection between model output and relevant input, is the very raison d'être of models (`www.theguardian.com/education/2020/mar/06/uk-universities-face-cash-black-hole-coronavirus-crisis`).

Sensitivity analysis can effectively tackle a multitude of issues, serving a dual role in the model development phase as well as during its utilization by users to enhance decision-making processes. Sensitivity analysis serves various purposes, including model validation, dimensionality reduction, prioritization of research efforts, pinpointing critical regions within the space of uncertainties under investigation, and aiding decision-making by quantifying how input variations impact outcome uncertainty [8–10].

At present, sensitivity analysis is evolving toward an independent discipline recognised also by institutional guidelines [11, 12]. However, whilst some disciplines promptly embraced sensitivity analysis, its potential has not yet been exploited in other fields, or its full adoption proceeds with hesitation.

The present concise historical account of sensitivity analysis attempts to chart the evolution of the field and gain insights into its contemporary challenges. In this study, we will prioritize interpretations of sensitivity analysis that emphasize the global exploration of uncertain inputs. This global understanding started in the 70s with the pioneering work of [13] who recognised that simultaneous variation of the parameters over a wide range of uncertainty is necessary to give reliable results.

As Global Sensitivity Analysis techniques have advanced in recent decades, becoming capable of handling complex models alongside the growing computational power of computers, user-friendly tools and software have been developed to broaden accessibility for a wide spectrum of researchers and practitioners and contribute to the wider dissemination of the discipline. Nevertheless, the sensitivity analysis panorama is still dominated in practically all disciplines by the so-called local approaches. To make an example, in operation research, where the objective is the optimal allocation of tasks and resources, sensitivity analysis is mostly pursued by looking at factors one at a time [14], ignoring possible crucial interactions of factors that may change the optimal solution only when changed jointly, but not one at a time.

4.2 Evolution of Sensitivity Analysis

Sensitivity analysis has undergone remarkable development over time, achieving several historical milestones that have significantly shaped its evolution. These crucial advancements 4.1, show the progressive journey of sensitivity analysis and highlight its growing importance as a fundamental tool within various scientific disciplines.

4.2.1 Early developments

Sensitivity analysis is, after all, finding things that have an effect on a certain phenomenon out of many things that could potentially be causes. So, if we look at sensitivity analysis as to a science of the causes, all the scientific revolution can be taken as anticipating sensitivity analysis. So, one sensitivity author compares Leonardo's experiment leading to laws of sliding friction to an early sensitivity experiment [15]. If we remain instead to the realm of the causes that can be discovered in mathematical constructs – rather than in the real, then perhaps a good date to set the start of sensitivity analysis is 1905, when Karl Pearson, the founder of modern statistics, proposed the idea of correlation ratio (known as the η^2 index), to link two variables associated by a non-linear relation.

A further milestone in the development of sensitivity analysis is the formalization of the experimental design in the 1920s and 1930s by the statistician Ronald Fisher. Experimental design is the process of planning and conducting experiments to test a hypothesis, answer a research question, or optimize the use of resources, including measuring or manipulating variables, hence the link with sensitivity analysis. The process whereby statistics managed to adjudicate the authority to assess the realism of causes is well described in the classic book of Desrosières and Desrosières [16].

World War 2 provided a significant impetus for the expansion and application of sensitivity analysis within the field of operational research [14, 17]. During this global conflict, nations were faced with unprecedented challenges in terms of strategic planning, resource allocation, and decision-making. The complexities of managing large-scale military operations, logistics, and supply chains required innovative approaches to optimize resource utilization and maximize efficiency.

Experimental design continued to develop with several important advancements in this field in the 1950s, including the widespread adoption of factorial designs. These designs allowed researchers to investigate the effects of multiple factors or variables on an outcome of interest. In a factorial design, each factor is varied at multiple levels, and the effects of each factor and their interactions are examined, thus allowing us to identify the unique effects of each independent variable and to test complex hypotheses.

Another important development in the 1950s was the introduction of the response surface methodology, which provided a way to optimize a response variable influenced by several input or process variables. The response surface methodology involves the use of mathematical models to describe the relationships between the input variables and the response variable, allowing researchers to identify the optimal values for each input variable to achieve the desired response. Notably, the Polynomial Chaos Expansions (PCE) [18] and Kriging, also called Gaussian Processes (GP) [19] got some traction.

In the following decades, Cukier et al. [20] developed the Fourier amplitude sensitivity test (FAST) in the early 1970s, one of the most elegant methods of sensitivity analysis. In FAST, the sensitivity of model output is computed by using spectral analysis through Fourier transformation of the input parameters. The method was primarily used in chemistry, but its applications extended to engineering, finance, and environmental modelling.

4.2.2 Transition to computer experiments

In the 1980s, the advancement of computing resources revolutionized sensitivity analysis and significantly expanded its capabilities. Prior to this period, sensitivity analysis was often limited to manual and analytical techniques, which were practical only for simple models with a few input parameters. However, with the increased computing power, researchers could now conduct sensitivity analyses on complex models that involved numerous input parameters and interactions.

One of the key breakthroughs during this time was the adoption of random sampling techniques. Instead of relying solely on analytical methods, researchers began to generate random samples of input parameters within specified ranges and then execute the model for each combination of these sampled inputs. This process allowed them to explore a vast range of possible input combinations, covering a wide spectrum of parameter values and assumptions. As a result, sensitivity analysis became more comprehensive, enabling the identification of critical factors that significantly influenced the model's behaviour. For a detailed overview of the progress made over these decades, see Myers et al. [21].

As computing resources continued to advance throughout the 1990s, sensitivity analysis reached another milestone with the pioneering work of the Russian mathematician Ilya M. Sobol'. In 1993, Sobol introduced an innovative approach to sensitivity analysis based on the decomposition of the output variance [22]. This method, known as Sobol's indices, allowed researchers to quantify the contribution of each input parameter to the variance of the model's output accurately.

The Sobol indices provided a deeper level of insight into the model's behaviour by quantifying the individual and combined effects of input parameters on the output variance. This method not only allowed researchers to rank the importance of different inputs but also enabled them to identify interactions and non-linearity between parameters, which were crucial for understanding complex systems.

Over the years, Sobol's sensitivity indices have become a widely used and well-established tool in various scientific domains, including engineering and environmental sciences [23]. The method's versatility and reliability have contributed significantly to the robustness of sensitivity analysis, making it an essential component in the study of models.

4.2.3 The modern communities

Towards the end of the 1990s, a brand-new community of sensitivity-analysis practitioners emerged, reflecting on the concept of Global Sensitivity Analysis. This approach involves simultaneously varying model inputs across a wide range of values to uncover interactions between parameters [23]. Supplemental Materials D synthetises this concept.

Concomitantly, the concept of uncertainty quantification gained traction in various scientific fields like climate modelling [24], computational physics, and materials science [25], focusing on propagating uncertainties through models to estimate prediction uncertainty.

The global approach to uncertainty quantification and sensitivity analysis garnered interest from institutions and communities worldwide. Efforts from U.S. national laboratories such as Sandia and Los Alamos [26, 27], along with the European Commission's Joint Research Centre (JRC) in Italy through the SAMO community (`https://www.sensitivityanalysis.org/`), played a crucial role in advancing such techniques. The first software started to emerge, such as the PREP (Preprocessor) and SPOP (Statistical Post-processor) codes for uncertainty and sensitivity analysis developed in the context of nuclear waste management where modellers engaged in a model intercomparison program, which included a benchmark on sensitivity analysis [28–31].

The growing enthusiasm led to the formation of new communities of practitioners, such as the UK's MUCM (Managing Uncertainty in Complex Models) community, which developed Bayesian techniques for computer experiments [32].

The Society for Industrial and Applied Mathematics (SIAM) in the United States and the CNRS (Centre national de la recherche scientifique) research group MASCOT-NUM (Methodes d'Analyse Stochastique pour les Codes et Traitements NUMériques) in France further contributed to spread uncertainty quantification and sensitivity analysis in their respective regions.

To support these efforts, software packages emerged from the early 1990s to 2010s, with a notable boom in the 2010s 4.1. The inclusion of SA in notable software like SciPy (`https://scipy.org`) has proven to be challenging as SA is only starting to be seen as a fundamental scientific tool by the scientific software community. Considering the reach of some of these software, it is expected that SA will get lot of attention in these communities.

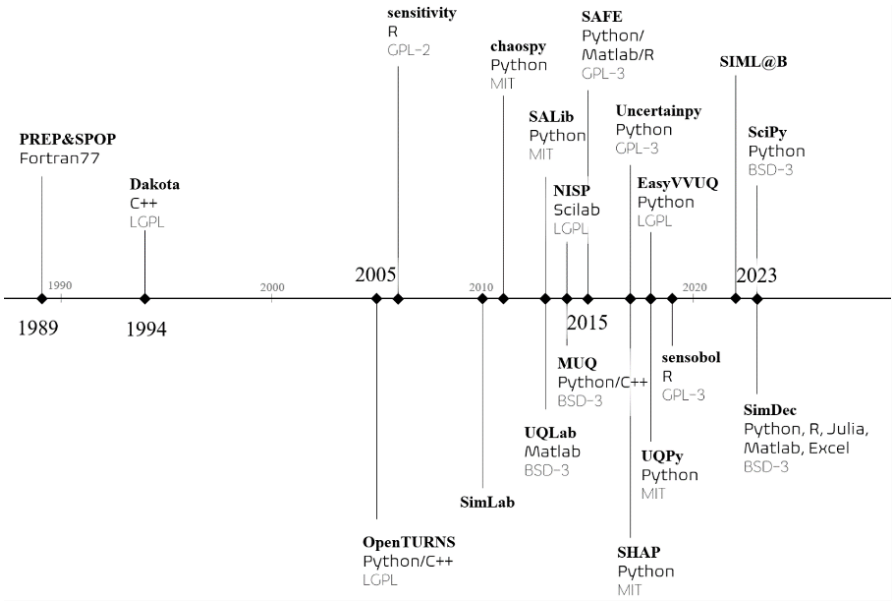


FIGURE 4.1: Introduction of software packages for sensitivity analysis.

Several handbooks on sensitivity analysis [9, 10, 27, 33], also contributed to the mainstreaming of sensitivity analysis by emphasizing its application in various settings like factor prioritization, factor fixing, and variance reduction [5]. These authors advocated for global methods other than local approaches for analysing non-linear and non-additive models.

To foster collaboration and knowledge exchange, the JRC school began the organization of international conferences on Global Sensitivity Analysis in 1995 (`www.sensitivityanalysis.org/conferences/`). Despite these advancements, local sensitivity analysis methods remained prevalent across disciplines. Researchers investigated the underpinning reasons and the practical implications of these trends in several contributions [34–37].

Recent key developments in sensitivity analysis include the introduction of moment-independent methods [38], which do not rely on statistical distribution assumptions of input parameters. Other important developments include a variogram-based method to determine sensitivities at different spatial scales [39, 40], PAWN [41], and others.

Faithful to their own precepts, SA practitioners also started to compare the performance of sensitivity analysis methods using SA itself [42, 43].

Furthermore, due to the increasing complexity of models, researchers focused on developing methods for sensitivity analysis of computationally expensive models. These methods involved constructing simpler surrogate models that could replace the original complex model for sensitivity analysis, with adaptive sampling strategies selecting the most informative input parameter combinations [44].

Moreover, a growing interest in integrating sensitivity analysis with machine learning methods cuts across scientific communities. Both approaches are grounded on the exploration of the parameter space to achieve both interpretable and highly predictive solutions, which is promising towards a fruitful synergy [45–47].

The trend toward model complexification emphasized the importance of using sensitivity analysis to ensure accurate and reliable model outputs. This trend tied back to the programmatic introduction of sensitivity analysis as a tool for model transparency within the framework of post-normal science [48]. In this context, sensitivity auditing [49] and the modelling of the modelling process [50] were introduced, urging modellers to retrace their assumptions and enhance transparency in the modelling process.

In the modelling of the modelling process, it is often advisable to consider multiple candidate models that differ in their assumptions or specifications. The process involves subjecting the various stages of the model-building process to coordinated and simultaneous variation in the modelling assumptions. This exploration can be carried out within a Monte Carlo framework, as discussed by Kroese et al. [51], by introducing random triggers that determine the model to be followed in each simulation. By combining the predictions from these models, we can account for the uncertainty associated with each model's parameter, assumptions and structural components.

4.2.4 The politics of sensitivity analysis

Recent years have seen an extra impetus to sensitivity analysis from policy studies. The COVID-19 pandemic was partly instrumental in this development, leading several authors to question the political use of models [52–54], ((`www.youtube.com/watch?v=_cgCTK17ics`)) with sensitivity analysis being advocate as a tool to make models less opaque [54]. Impact assessment is also a field where sensitivity analysis is seen as a useful lens to peers at models [55], also in conjunction with sensitivity auditing just mentioned. A recent volume devoted to the politics of modelling [56] also includes a relevant discussion of sensitivity analysis. Sensitivity analysis and auditing have recently been proposed as tools to jointly match the double demand for technical and normative quality in modelling [57], echoing a parallel discussion in the field of social statistics [58, 59].

4.3 Bibliometric Survey

Bibliometric tools have recently emerged as valuable instruments for studying the evolutionary dynamics within specific scientific domains [60]. These tools have previously been employed to investigate the trajectories of sensitivity analysis and Global Sensitivity Analysis [35], as well as

the patterns of adoption and utilization of software for uncertainty management in the field of environmental sciences [61, 62].

In this article, a bibliometric analysis was conducted with the explicit aim of exploring the development of sensitivity analysis as a discipline in scholarly literature. The analysis leverages a Scopus dataset containing 16,513 documents including books, book chapters, articles, conference papers, and reviews. These documents were selected[1] based on the presence of the term sensitivity analysis (respectively Global Sensitivity Analysis), within their abstracts or keywords, coupled with model and uncertainty as control fields anywhere in the text body. After data cleansing and processing, leading to the creation of infographics and charts, several observations emerged.

Above all, consolidating the findings of Ferretti et al. [35], the corpus of literature has continued its consistent growth since 2016, as shown in 4.2. As mentioned in the previous section, the penetration of Global Sensitivity Analysis methods into the broader modelling community has not reached its full potential and, to the present, still represents roughly a fifth of the total number of published documents.

As noted in Saltelli et al. [36] the slower uptake of GSA methods might be partly due to their intrinsic complexity. These often involve algorithms and computational processes which can be daunting for researchers who are not well-versed in sensitivity analysis methodologies. Additionally, there might be a reluctance among some practitioners to deviate from familiar and established local sensitivity analysis approaches, even when global methods could offer more comprehensive insights into complex models.

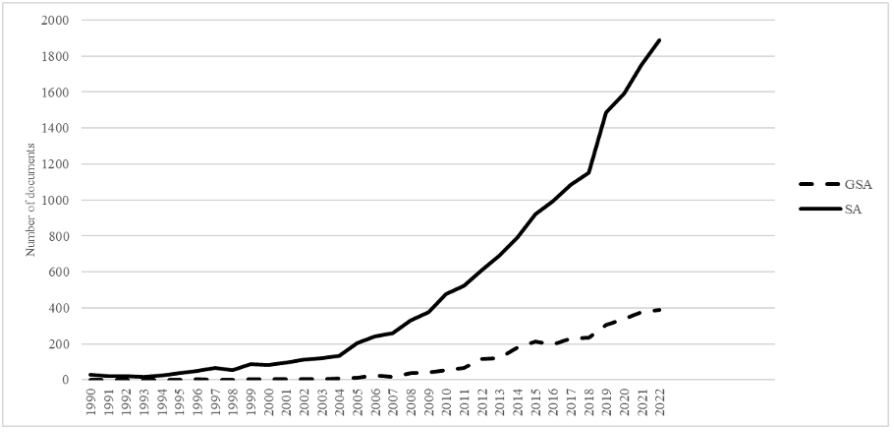


FIGURE 4.2: Publications per year that adopt any kind of sensitivity analysis (SA, solid line) versus those that employ more sophisticated global methods (GSA, dashed line).

[1]Query specification: (ABS(sensitivity analysis) OR KEY (sensitivity analysis) AND ALL (model AND uncertainty) AND REF (sensitivity analysis) AND PUBYEAR > 1900 AND PUBYEAR < 2023) AND (LIMIT-TO (DOCTYPE, bk) OR LIMIT-TO (DOCTYPE, ch) OR LIMIT-TO (DOCTYPE, re) OR LIMIT-TO (DOCTYPE, cp) OR LIMIT-TO (DOCTYPE, ar)). Retrieved on Scopus.com through API calls. June 2023

Figure 4.3 provides insight into the distribution of documents across distinct subject areas, highlighting a concentration of documents within engineering and environmental science.

Minimal disparity emerges between sensitivity analysis and Global Sensitivity Analysis and their fields of application: the distribution of the various disciplines' shares is essentially mirrored in Global Sensitivity Analysis, albeit on a more modest scale.

This trend is further reinforced by Figure 4.4, showing the distribution based on publishing sources. Notably absent fields include finance and economics, and, to a lesser extent, medicine and related fields such as psychology and neuroscience. Considering the relevance of risk within these disciplines, it is rather surprising to observe their substantial absence in the body of literature on sensitivity analysis, a fundamental component in comprehending and managing risk. A tentative explanation could come from the fact that these fields have traditionally been using other statistical tools such as hypothesis testing. In some ways, these methods can be used to answer similar questions.

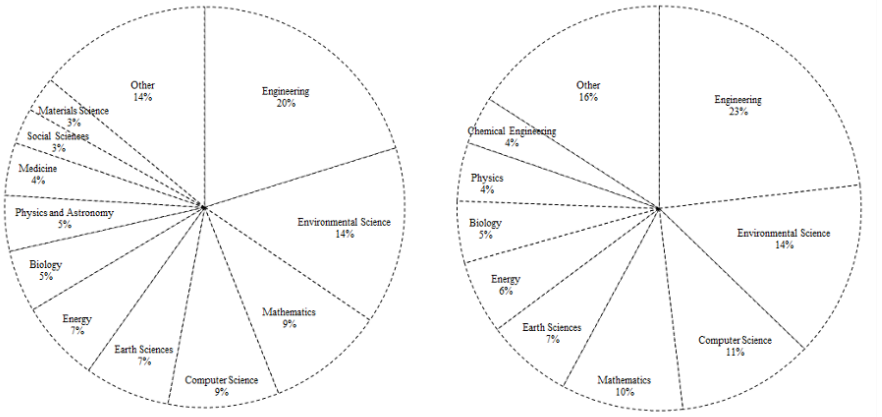


FIGURE 4.3: Subject area segmentation, on the left SA, on the right GSA.

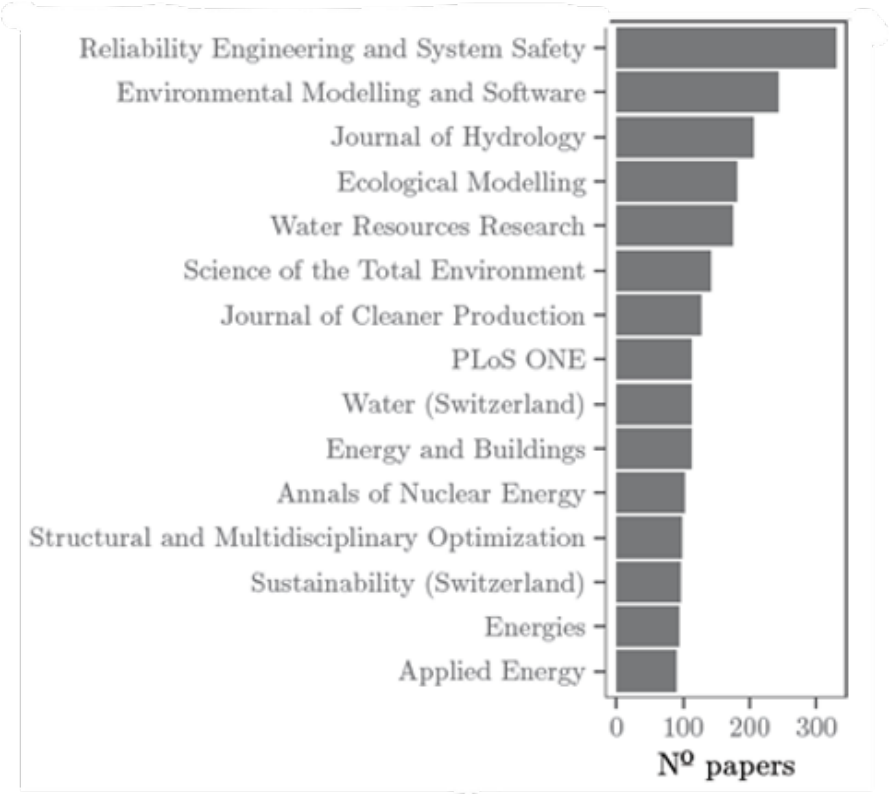


FIGURE 4.4: Outlets that publish sensitivity analysis studies.

The geographical distribution of publishing countries is displayed in Figure 4.5, wherein the United States and China jointly account for a significant proportion of all published material.

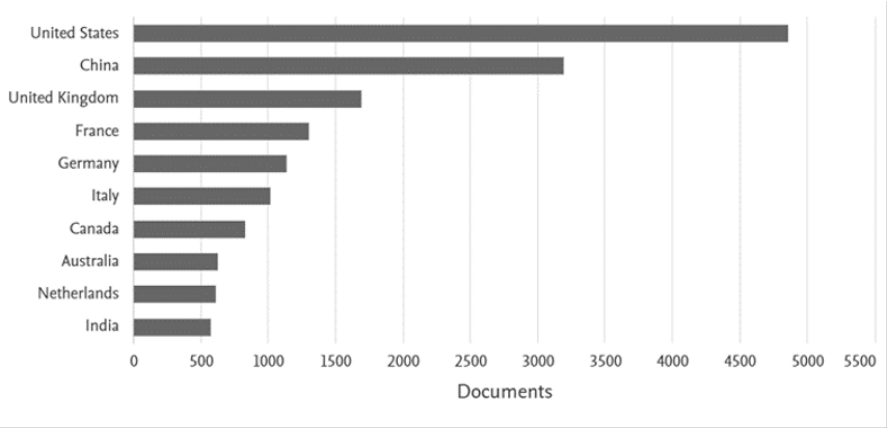


FIGURE 4.5: Geographical profile of sensitivity analysis publications.

Figure 4.6 attempts to reconstruct the methodological landmarks within the field. Specifically, it focuses on documents that account for more than 500 citations, underscoring the pivotal role played by these works in shaping the methodological landscape of Sensitivity Analysis.

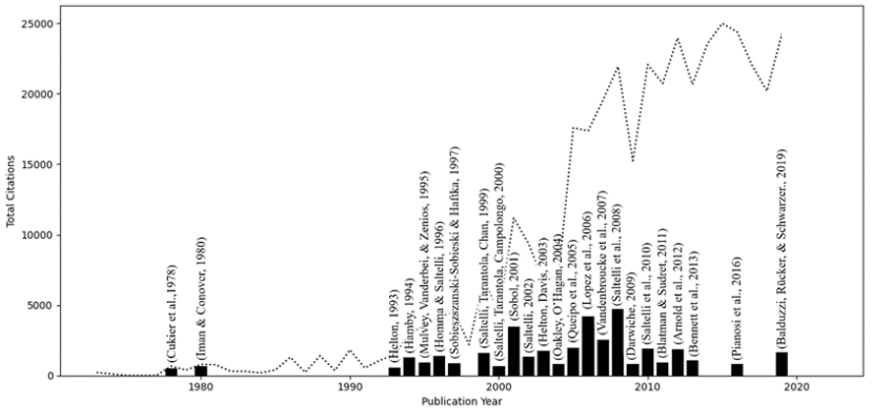


FIGURE 4.6: Most cited documents and total citations (dotted line).

4.4 A fragmented adoption of sensitivity analysis

The historical development of sensitivity analysis is driven by the need to understand the effects of changes in model parameters on model outputs. However, the adoption of sensitivity analysis across various disciplines remains fragmented, with no guarantee that approaches effective in one field will be equally applicable in another. Factors that contribute to this divergence include:

- Research Culture and Tradition: Different scientific disciplines may have distinct research cultures and traditions that influence the preferred methods and practices. If sensitivity analysis has not been widely adopted or promoted within a particular discipline, researchers might be less inclined to explore its potential benefits;
- Computational Resources: Some sensitivity analysis techniques require significant computational resources, making them less feasible for fields with limited access to high-performance computing facilities or where model evaluations are computationally expensive;
- Expertise and Awareness: The level of expertise and awareness of sensitivity analysis methods among researchers in different disciplines can affect their willingness to adopt these techniques. Disciplines with a strong background in statistics might be more likely to embrace sensitivity analysis compared to those less familiar with the concepts.

4.5 Conclusions

Sensitivity analysis has progressed from its origins in laboratory and field experiments to in-silico research. Along its journey, it has grappled with crucial challenges, notably the balance between

local and global analysis approaches. With the increasing sophistication of sensitivity analysis methods, that emerged in order to address the intricacies of increasingly complex models, there is a growing demand for user-friendly tools, aiming to broaden accessibility for researchers and practitioners.

To assess the present landscape of sensitivity analysis, we conducted a bibliometric survey spanning diverse academic disciplines. This analysis offered valuable insights into the spread of documents among different subject areas, underscoring a notable concentration within the fields of engineering and environmental science, whereas absent fields include finance and economics, and, to a lesser extent, medicine and related fields, such as psychology and neuroscience. Interestingly, there is minimal disparity between sensitivity analysis and Global Sensitivity Analysis in terms of their respective applications. The distribution of document shares across various disciplines essentially mirrors that of Global Sensitivity Analysis, albeit on a somewhat smaller scale. This assessment pinpointed areas where further integration and adoption of sensitivity analysis methods is required.

In our perspective, the fragmentation in the adoption of sensitivity analysis across diverse fields can be attributed to a few factors, including:

- divergent research cultures and traditions in the different scientific disciplines;
- heterogeneous computational resources in the various research fields;
- varying degrees of proficiency and familiarity with sensitivity analysis methods among researchers in the different disciplines.

In the future, sensitivity analysis is expected to play a pivotal role in guiding model development and decision-making processes, especially as simulation models become increasingly bigger and more complex. The ongoing innovation and collaboration among researchers and practitioners are key to addressing the adoption challenges, fully harnessing the potential of sensitivity analysis, and enhancing our grasp of complex systems and their uncertainties. This will eventually lead to more dependable and informed decision-making amid the increasing complexity and uncertainty in our world.

Authors' contributions: S.T. was the project administrator that conceptualized the paper. F.F. and M.K. conducted the formal analysis, with the validation and data curation step. A.S. supervised the project. S.T., F.F., S.L.P, A.L., R.R., A.P., P.R., G.V., M.K.C., and A.S. wrote, reviewed, and edited the manuscript.

Bibliography

[1] M.D. Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174, 1991. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.1991.10484804. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.1991.10484804.

[2] R. Rosen. *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. Complexity in ecological systems series. Columbia Univ. Press, New York, 1991. ISBN 978-0-231-07564-0 978-0-231-07565-7.

[3] W.N. Espeland and M.L. Stevens. A Sociology of Quantification. *European Journal of Sociology*, 49(3):401–436, 2008. ISSN 0003-9756, 1474-0583. doi: 10.1017/S0003975609000150. URL https://www.cambridge.org/core/product/identifier/S0003975609000150/type/journal_article.

[4] M.S. Morgan and M. Morrison, editors. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press, 1 edition, 1999. ISBN 978-0-521-65097-7 978-0-521-65571-2 978-0-511-66010-8. doi: 10.1017/CBO9780511660108. URL https://www.cambridge.org/core/product/identifier/9780511660108/type/book.

[5] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002. ISSN 00104655. doi: 10.1016/S0010-4655(02)00280-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0010465502002801.

[6] A. Saltelli and S. Tarantola. On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. *Journal of the American Statistical Association*, 97(459):702–709, 2002. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214502388618447. URL http://www.tandfonline.com/doi/abs/10.1198/016214502388618447.

[7] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, 1 edition, 2002. ISBN 978-0-470-87093-8 978-0-470-87095-2. doi: 10.1002/0470870958. URL https://onlinelibrary.wiley.com/doi/book/10.1002/0470870958.

[8] S. Tarantola, N. Giglioli, J. Jesinghaus, and A. Saltelli. Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making? *Stochastic Environmental Research and Risk Assessment (SERRA)*, 16(1):63–76, 2002. ISSN 14363240. doi: 10.1007/s00477-001-0085-x. URL http://link.springer.com/10.1007/s00477-001-0085-x.

[9] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-05997-5.

[10] A. Saltelli, editor. *Sensitivity analysis*. Wiley paperback series. Wiley, Chichester Weinheim, paperback ed edition, 2008. ISBN 978-0-470-74382-9.

[11] European Commission. Joint Research Centre. *Uncertainty and sensitivity analysis for policy decision making: an introductory guide*. Publications Office, LU, 2020. URL https://data.europa.eu/doi/10.2760/922129.

[12] European Commission. Better Regulation—Joining Forces to Make Better Laws. *European Commission*, 2021. Publisher: European Commission Brussels, Belgium.

[13] C.S. Holling and United Nations Environment Programme, editors. *Adaptive environmental assessment and management*. Number 3 in International series on applied systems analysis. International Institute for Applied Systems Analysis ; Wiley, [Laxenburg, Austria] : Chichester ; New York, 1978. ISBN 978-0-471-99632-3. Meeting Name: Workshop on Adaptive Assessment of Ecological Policies.

[14] F.S. Hillier and G.J. Lieberman. *Introduction to operations research*. McGraw-Hill Higher Education, New York, 9th ed edition, 2010. ISBN 978-0-07-337629-5 978-0-07-729834-0. OCLC: ocn245598960.

[15] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. Lo Piano, T. Iwanaga, W. Becker, S. Tarantola, J.H.A. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H.R. Maier. The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137:104954, 2021. ISSN 13648152. doi: 10.1016/j.envsoft.2020.104954. URL https://linkinghub.elsevier.com/retrieve/pii/S1364815220310112.

[16] A. Desrosières and A. Desrosières. *The politics of large numbers: a history of statistical reasoning*. Harvard University Press, Cambridge, Mass., digital repr. 2011 edition, 2011. ISBN 978-0-674-00969-1 978-0-674-68932-9.

[17] S.I. Gass and A. Assad. *An annotated timeline of operations research: an informal history*. Number 75 in International series in operations research & management science. Kluwer Academic Publishers, New York?, 2005. ISBN 978-1-4020-8112-5 978-1-4020-8116-3.

[18] N. Wiener. The Homogeneous Chaos. *American Journal of Mathematics*, 60(4):897, 1938. ISSN 00029327. doi: 10.2307/2371268. URL https://www.jstor.org/stable/2371268?origin=crossref.

[19] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963. ISSN 1554-0774, 0361-0128. doi: 10.2113/gsecongeo.58.8.1246. URL http://pubs.geoscienceworld.org/economicgeology/article/58/8/1246/17275/Principles-of-geostatistics.

[20] R.I. Cukier, H.B. Levine, and K.E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26(1):1–42, 1978. ISSN 00219991. doi: 10.1016/0021-9991(78)90097-9. URL https://linkinghub.elsevier.com/retrieve/pii/0021999178900979.

[21] R.H. Myers, A.I. Khuri, W.H. Carter, and A.I. Khuri. Response Surface Methodology: 1966-1988. *Technometrics*, 31(2):137, 1989. ISSN 00401706. doi: 10.2307/1268813. URL https://www.jstor.org/stable/1268813?origin=crossref.

[22] I.M. Soboĺ. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.*, 1:407, 1993.

[23] A. Saltelli, S. Tarantola, and F. Campolongo. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*, 15(4):377–395, 2000. ISSN 08834237. URL http://www.jstor.org/stable/2676831. Publisher: Institute of Mathematical Statistics.

[24] C.E. Forest, P.H. Stone, A.P. Sokolov, M.R. Allen, and M.D. Webster. Quantifying Uncertainties in Climate System Properties with the Use of Recent Climate Observations. *Science*, 295(5552):113–117, 2002. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1064419. URL https://www.science.org/doi/10.1126/science.1064419.

[25] J.J. Gabriel, F.Y.C. Congo, A. Sinnott, K. Mathew, T.C. Allison, F. Tavazza, and R.G. Hennig. Uncertainty Quantification for Materials Properties in Density Functional Theory with k-Point Density. *arXiv*, 2020. doi: 10.48550/ARXIV.2001.01851. URL https://arxiv.org/abs/2001.01851.

[26] A. Saltelli and J. Marivoet. Non-parametric statistics in sensitivity analysis for model output: A comparison of selected techniques. *Reliability Engineering & System Safety*, 28 (2):229–253, 1990. ISSN 09518320. doi: 10.1016/0951-8320(90)90065-U. URL https://linkinghub.elsevier.com/retrieve/pii/095183209090065U.

[27] J.C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2-3): 327–367, 1993. ISSN 09518320. doi: 10.1016/0951-8320(93)90097-I. URL https://linkinghub.elsevier.com/retrieve/pii/095183209390097I.

[28] OECD Nuclear Energy Agency. Probabilistic System Assessment Code User Group. *PSACOIN Level O Intercomparison: An International Code Intercomparison Exercise on a Hypothetical Safety Assessment Case Study for Radioactive Waste Disposal Systems*. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, 1987.

[29] A. Saltelli, T.H. Andres, B.W. Goodwin, E. Sartori, and S.G. S.G. Carlyle. PSACOIN level 0 intercomparison-an international verification exercise on a hypothetical safety assessment case study. In *[1989] Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences. Volume II: Software Track*, pages 267–274, Kailua-Kona, HI, USA, 1989. IEEE Comput. Soc. Press. ISBN 978-0-8186-1912-0. doi: 10.1109/HICSS.1989.48001. URL http://ieeexplore.ieee.org/document/48001/.

[30] R.A. Klos, J.E. Sinclair, C. Torres, U. Bergström, and D.A. Galson. Psacoin level 1B intercomparison: An International code intercomparison exercise on a hypothetical safety assessment case study for radioactive waste disposal systems. Technical report, Nuclear Energy Agency, 1993.

[31] P.D.R. WASTE. PSACOIN LEVEL 2. *Citeseer*, 1994. Publisher: Citeseer.

[32] J.E. Oakley and A. O'Hagan. Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):751–769, 2004. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/3647504. Publisher: [Royal Statistical Society, Wiley].

[33] A. Saltelli and S. Funtowicz. The Precautionary Principle: implications for risk management strategies. *International Journal of Occupational Medicine and Environmental Health*, 17 (1):47–57, 2004. ISSN 1232-1087.

[34] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance-based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655.

doi: 10.1016/j.cpc.2009.09.018. URL https://linkinghub.elsevier.com/retrieve/pii/S0010465509003087.

[35] F. Ferretti, A. Saltelli, and Stefano S. Tarantola. Trends in sensitivity analysis practice in the last decade. *Science of The Total Environment*, 568:666–670, 2016. ISSN 00489697. doi: 10.1016/j.scitotenv.2016.02.133. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969716303448.

[36] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, 114:29–39, 2019. ISSN 13648152. doi: 10.1016/j.envsoft.2019.01.012. URL https://linkinghub.elsevier.com/retrieve/pii/S1364815218302822.

[37] S. Lo Piano and L. Benini. A critical perspective on uncertainty appraisal and sensitivity analysis in life cycle assessment. *Journal of Industrial Ecology*, 26(3):763–781, June 2022. ISSN 1088-1980, 1530-9290. doi: 10.1111/jiec.13237. URL https://onlinelibrary.wiley.com/doi/10.1111/jiec.13237.

[38] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007. ISSN 09518320. doi: 10.1016/j.ress.2006.04.015. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832006000883.

[39] S. Razavi and H.V. Gupta. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, 52(1):423–439, 2016. ISSN 0043-1397, 1944-7973. doi: 10.1002/2015WR017558. URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015WR017558.

[40] S. Razavi and H.V. Gupta. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, 52(1):440–455, 2016. ISSN 0043-1397, 1944-7973. doi: 10.1002/2015WR017559. URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015WR017559.

[41] F. Pianosi and T. Wagener. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67:1–11, 2015. ISSN 13648152. doi: 10.1016/j.envsoft.2015.01.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1364815215000237.

[42] A. Puy, S. Lo Piano, and A. Saltelli. Current Models Underestimate Future Irrigated Areas. *Geophysical Research Letters*, 47(8):e2020GL087360, 2020. ISSN 0094-8276, 1944-8007. doi: 10.1029/2020GL087360. URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL087360.

[43] A. Puy, S. Lo Piano, and A. Saltelli. Is VARS more intuitive and efficient than Sobol' indices? *Environmental Modelling & Software*, 137:104960, March 2021. ISSN 13648152. doi: 10.1016/j.envsoft.2021.104960. URL https://linkinghub.elsevier.com/retrieve/pii/S1364815221000037.

[44] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008. ISSN 09518320. doi: 10.1016/j.ress.2007.04.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832007001329.

[45] P. Zhang. A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. *Applied Soft Computing*, 85: 105859, 2019. ISSN 15684946. doi: 10.1016/j.asoc.2019.105859. URL https://linkinghub.elsevier.com/retrieve/pii/S1568494619306404.

[46] C. Bénard, S. Da Veiga, and E. Scornet. Interpretability via Random Forests. In Antonio Lepore, Biagio Palumbo, and Jean-Michel Poggi, editors, *Interpretability for Industry 4.0 : Statistical and Machine Learning Approaches*, pages 37–84. Springer International Publishing, Cham, 2022. ISBN 978-3-031-12401-3 978-3-031-12402-0. doi: 10.1007/978-3-031-12402-0_3. URL https://link.springer.com/10.1007/978-3-031-12402-0_3.

[47] B. Iooss, R. Kenett, and P. Secchi. Different views of interpretability. In *Interpretability for Industry 4.0: Statistical and Machine Learning Approaches*, pages 1–20. Springer, 2022.

[48] S.O. Funtowicz and J.R. Ravetz. Science for the post-normal age. *Futures*, 25(7):739–755, 1993. ISSN 00163287. doi: 10.1016/0016-3287(93)90022-L. URL https://linkinghub.elsevier.com/retrieve/pii/001632879390022L.

[49] A. Saltelli, A. Guimaraes Pereira, J.P. Van Der Sluijs, and S. Funtowicz. What do I make of your latinorumc Sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy*, 9(2/3/4):213, 2013. ISSN 1740-2816, 1740-2824. doi: 10.1504/IJFIP.2013.058610. URL http://www.inderscience.com/link.php?id=58610.

[50] S. Lo Piano, R. Sheikholeslami, A. Puy, and A. Saltelli. Unpacking the modelling process via sensitivity auditing. *Futures*, 144:103041, 2022. ISSN 00163287. doi: 10.1016/j.futures.2022.103041. URL https://linkinghub.elsevier.com/retrieve/pii/S0016328722001410.

[51] D.P. Kroese, T. Brereton, T. Taimre, and Z.I. Botev. Why the Monte Carlo method is so important today. *WIREs Computational Statistics*, 6(6):386–392, 2014. ISSN 1939-5108, 1939-0068. doi: 10.1002/wics.1314. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1314.

[52] C. Caduff. What Went Wrong: Corona and the World after the Full Stop. *Medical Anthropology Quarterly*, 34(4):467–487, 2020. ISSN 0745-5194, 1548-1387. doi: 10.1111/maq.12599. URL https://anthrosource.onlinelibrary.wiley.com/doi/10.1111/maq.12599.

[53] T. Rhodes and K. Lancaster. Mathematical models as public troubles in COVID-19 infection control: following the numbers. *Health Sociology Review*, 29(2):177–194, 2020. ISSN 1446-1242, 1839-3551. doi: 10.1080/14461242.2020.1764376. URL https://www.tandfonline.com/doi/full/10.1080/14461242.2020.1764376.

[54] A. Saltelli, G. Bammer, I. Bruno, E. Charters, M. Di Fiore, E. Didier, W. Nelson Espeland, J. Kay, S. Lo Piano, D. Mayo, R. Pielke Jr, T. Portaluri, T.M. Porter, A. Puy, I. Rafols, J.R. Ravetz, E. Reinert, D. Sarewitz, P.B. Stark, A. Stirling, J. Van Der Sluijs, and P. Vineis. Five ways to ensure that models serve society: a manifesto. *Nature*, 582(7813):482–484, 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/d41586-020-01812-9. URL https://www.nature.com/articles/d41586-020-01812-9.

[55] A. Saltelli, M. Kuc-Czarnecka, S. Lo Piano, M.J. Lőrincz, M. Olczyk, A. Puy, E. Reinert, S.T. Smith, and J.P. Van Der Sluijs. Impact assessment culture in the European Union. Time for something new? *Environmental Science & Policy*, 142:99–111, 2023. ISSN 14629011. doi: 10.1016/j.envsci.2023.02.005. URL https://linkinghub.elsevier.com/retrieve/pii/S1462901123000382.

[56] A. Saltelli and M. Di Fiore, editors. *The politics of modelling: numbers between science and policy*. Oxford University Press, New York, 2023. ISBN 978-0-19-887241-2.

[57] A. Saltelli and A. Puy. What Can Mathematical Modelling Contribute to a Sociology of Quantification? *SSRN Electronic Journal*, 2022. ISSN 1556-5068. doi: 10.2139/ssrn.4212453. URL https://www.ssrn.com/abstract=4212453.

[58] A. Sen. Justice: Means versus Freedoms. *Philosophy & Public Affairs*, 19(2):111–121, 1990. ISSN 00483915, 10884963. URL http://www.jstor.org/stable/2265406. Publisher: Wiley.

[59] R. Salais. "La donnée n'est pas un donné": Statistics, Quantification and Democratic Choice. In A. Mennicken and R. Salais, editors, *The New Politics of Numbers*, pages

379–415. Springer International Publishing, Cham, 2022. ISBN 978-3-030-78200-9 978-3-030-78201-6. doi: 10.1007/978-3-030-78201-6_12. URL https://link.springer.com/10.1007/978-3-030-78201-6_12.

[60] M.W. Neff and E.A. Corley. 35 years and 160,000 articles: A bibliometric exploration of the evolution of ecology. *Scientometrics*, 80(3):657–682, 2009. ISSN 0138-9130, 1588-2861. doi: 10.1007/s11192-008-2099-3. URL http://link.springer.com/10.1007/s11192-008-2099-3.

[61] D. Douglas-Smith, T. Iwanaga, B.F.W. Croke, and A.J. Jakeman. Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. *Environmental Modelling & Software*, 124:104588, 2020. ISSN 13648152. doi: 10.1016/j.envsoft.2019.104588. URL https://linkinghub.elsevier.com/retrieve/pii/S1364815219305845.

[62] S. Lo Piano and L. Benini. A critical perspective on uncertainty appraisal and sensitivity analysis in life cycle assessment. *Journal of Industrial Ecology*, 26(3):763–781, June 2022. ISSN 1088-1980, 1530-9290. doi: 10.1111/jiec.13237. URL https://onlinelibrary.wiley.com/doi/10.1111/jiec.13237.

CHAPTER 5

Other published articles

During the PhD period, we also worked on other projects related to the dissertation topic, which resulted in the following published articles.

Burden of disease from second-hand tobacco smoke exposure at home among adults from European Union countries in 2017: an analysis using a review of recent meta-analyses

G. Carreras, **A. Lachi**, B. Cortini, S. Gallus, M.J. Lopez, A.N. Lopez, J.B. Soriano, E. Fernandez, O. Tigova, G. Gorini, and TackSHS Project Investigators. Burden of disease from second-hand tobacco smoke exposure at home among adults from european union countries in 2017: an analysis using a review of recent meta-analyses. Preventive medicine, 145:106412, 2021. doi: https://doi.org/10.1016/j.ypmed.2020.106412

Abstract: Smoke-free legislation reduced second-hand smoke (SHS) exposure in public places and indirectly promoted private smoke-free settings. Nevertheless, a large proportion of adults are still exposed to SHS at home. The aim of this paper is to quantify the burden of disease due to home SHS exposure among adults in the 28 European Union (EU) countries for the year 2017. The burdens by sex from lung cancer, chronic obstructive pulmonary disease (COPD), breast cancer, ischemic heart disease (IHD), stroke, asthma, and diabetes were estimated in an original research analysis using the comparative risk assessment method. Relative risks of death/diseases by sex for adults exposed to SHS at home compared to those not exposed were estimated by updating existing meta-analyses. The prevalence of home SHS exposure by sex was estimated using a multiple imputation procedure based on Eurobarometer surveys. Data on mortality and disability-adjusted life years (DALYs) were obtained from the Global Burden of Disease, Injuries and Risk Factors Study. In 2017, 526,000 DALYs (0.36% of total DALYs) and 24,000 deaths

(0.46% of total deaths) were attributable to home SHS exposure in the 28-EU countries, mainly from COPD and IHD. South-Eastern EU countries showed the highest burden, with a proportion of DALYs/deaths attributable to SHS exposure on a total higher than 0.50%/ 0.70%, whereas northern EU countries showed the lowest burden, with proportions of DALYs/deaths lower than 0.25%/0.34%. The burden from SHS exposure is still significant in EU countries. More could be done to raise awareness of the health risks associated with SHS exposure at home.

Burden of disease from exposure to secondhand smoke in children in Europe

G. Carreras, **A. Lachi**, B. Cortini, S. Gallus, M.J. Lopez, A.N. Lopez, A. Lugo, M.T. Pastor, J.B. Soriano, E. Fernandez, and TackSHS Project Investigators. Burden of disease from exposure to secondhand smoke in children in europe. Pediatric research, 90(1):216–222, 2021. doi: https://doi.org/10.1038/s41390-020-01223-6

Abstract: Secondhand smoke (SHS) exposure at home and fetal SHS exposure during pregnancy are major causes of disease among children. The aim of this study is to quantify the burden of disease due to SHS exposure in children and in pregnancy in 2006–2017 for the 28 European Union (EU) countries. Exposure to SHS was estimated using a multiple imputation procedure based on the Eurobarometer surveys, and SHS exposure burden was estimated with the comparative risk assessment method using meta-analytical relative risks. Data on deaths and disability-adjusted life years (DALYs) were collected from National statistics and from the Global Burden of Disease Study. Exposure to SHS and its attributable burden stalled from 2006 to 2017; in pregnant women, SHS exposure was 19.8% in 2006, 19.1% in 2010, and 21.0% in 2017; in children, it was 10.1% in 2006, 9.6% in 2010, and 12.1% in 2017. In 2017, 35,633 DALYs among children were attributable to SHS exposure in the EU, mainly due to low birth weight. Comprehensive smoking bans up to 2010 contributed to reducing SHS exposure and its burden in children immediately after their implementation; however, SHS exposure still occurs, and in 2017, its burden in children was still relevant.

Dose-risk relationships between cigarette smoking and cervical cancer: a systematic review and meta-analysis

M.C. Malevolti, A. Lugo, M. Scala, S. Gallus, G. Gorini, **A. Lachi**, and G. Carreras. Dose-risk relationships between cigarette smoking and cervical cancer: a systematic review and meta-analysis. European Journal of Cancer Prevention, 32(2):171–183, 2023. doi: https://doi.org/10.1097/cej.0000000000000773

Abstract: Cervical cancer (CC) is the fourth most frequent cancer worldwide. Cigarette smoking has been shown to influence CC risk in conjunction with human papillomavirus (HPV) infection. The aim of this study is to provide the most accurate and updated estimate of this association and its dose-response relationship. Using an innovative approach for the identification of original publications, we conducted a systematic review and meta-analysis of studies published up to January 2021. Random effects models were used to provide pooled relative risks (RRs) of CC for smoking status. Dose-response relationships were evaluated using one-stage random effects models with linear or restricted cubic splines models. We included 109 studies providing a pooled RR of invasive CC and preinvasive lesions, respectively, of 1.70 [95% confidence interval (CI), 1.53–1.88] and 2.11 (95% CI, 1.85–2.39) for current versus never smokers, and, respectively, 1.13 (95% CI, 1.02–1.24) and 1.29 (95% CI, 1.15–1.46) for former versus never smokers. Considering HPV does not alter the positive association or its magnitude. Risks of CC sharply increased with few cigarettes (for 10 cigarettes/day, RR = 1.72; 95% CI, 1.34–2.20 for invasive CC and RR = 2.13; 95% CI, 1.86–2.44 for precancerous lesions). The risk of CC increased with pack years and smoking duration and decreased linearly with time since quitting, reaching that of never-smokers about 15 years after quitting. This comprehensive review and meta-analysis confirmed the association of smoking with CC, independently from HPV infection. Such association rose sharply with smoking intensity and decreased after smoking cessation.

School-based screening strategies to prevent the spread of COVID-19 in school: a systematic review of the literature

M. Marra, M. Baccini, G. Cereda, M. Culasso, M. De Sario, I. Eboli, **A. Lachi**, Z. Mitrova, R. Saulle, and A. Bena. Strategie di screening per contenere la diffusione del covid-19 nella scuola: una revisione sistematica della letteratura. Epidemiologia e Prevenzione, 47(3):152–171, 2023. doi: https://doi.org/10.19191/EP23.3.A576.054.101

Abstract: To describe studies that evaluated the screening programs implemented in the school during the COVID-19 pandemic. A systematic literature review was conducted according to the PRISMA 2020 Guidelines. Studies published until December 2021 were included. The methodological quality of the studies was assessed with validated scales. Study selection, data extraction, and quality assessment were carried out by two authors independently. Teachers and students belonging to schools of all levels, including universities. After having removed duplicate articles, 2,822 records were retrieved. Thirty-six studies were included (15 used an observational design and 21 modeling studies). Regarding the former, the methodological quality has been rated as high in 2 studies, intermediate in 6, and low in 2; in the remaining ones, it was not evaluated because only descriptive. Screenings were quite different in terms of school study population, types of tests used, methods of submission and analysis, and level

of incidence in the community at the time of implementation. Outcome indicators were also varied, a heterogeneity that, on the one hand, did not allow for meta-analysis of results and, on the other, allowed for testing the performance of the screenings in very different settings. All of the field studies claim that the screenings reduced SARS-CoV-2 exposure and infection among children, adolescents, and college students, curbing at-school transmission and helping to reduce the number of closing school days. Studies that evaluated the cost of the intervention emphasized its cost-effectiveness, while those that focused on the acceptability of the instrument showed a preference among children, adolescents, and parents for minimally invasive, self-administered tests with high sensitivity and lower frequency of repetition. Simulation-based studies are mostly based on compartmental and agent-based models. Their quality is quite high methodologically, although uncertainty quantification and external validation, aimed at verifying the model's ability to reproduce observed data, are lacking in many cases. The contexts to which the simulations refer are all school-based, although 7 studies consider residential situations, which are poorly suited to the Italian context. All simulation-based models indicate the importance of planning repeated testing on asymptomatic individuals to limit contagion. However, the costs of these procedures can be high unless assessments are spaced out or pool testing procedures are used. Obtaining high student adherence to the screening program is extremely important to maximize results. School-based screenings, especially when combined with other preventive measures, have been important public health tools to contain infections during COVID-19 waves to ensure children's and adolescents' right to education and to prevent the fallout in physical and mental health (with strong equity consequences) associated with school closures.

Conclusion

The focus of this dissertation is on the use of Compartmental models in epidemiology and the associated quantification of uncertainty. The articles included in this work contribute to advancing the understanding of the complex dynamics of smoking habits.

In Chapter 1, the paper entitled *A compartmental model for smoking dynamics in Italy: A pipeline for inference, validation, and forecasting under hypothetical scenarios* illustrates the development of a Compartmental model for simulating smoking dynamics. The model presented in this article proposes an extremely versatile tool that can be easily adapted to different contexts other than the Italian one. The incorporation of validation tools, such as cross-validation and Global Sensitivity Analysis, enhances the robustness of the model and supports the credibility of its findings.

In Chapter 2, the paper entitled *Frequentist and Bayesian inference on compartmental models in epidemiology: A critical review with a focus on likelihood-free approaches* highlights the importance of likelihood-free approaches. Calibration and Approximate Bayesian Computation represent powerful tools in the study of intractable likelihood functions. Their flexibility, requiring only the generation of pseudo-data and sampling from the prior distribution, underscores their easy applicability in various research fields.

Chapter 3, the paper entitled *Smoking dynamics in Tuscany (Italy) under alternative tobacco control policies: Robustifying inference and forecasting via uncertainty propagation and Global Sensitivity Analysis* emphasizes the importance of sensitivity analysis in complex models. Since epidemiological models can be extremely complex, sensitivity analysis is destined to become a cornerstone for navigating the intricate terrain of complex systems. Identifying paths that lead to misleading results, is extremely important in a statistical analysis.

In Chapter 4, the paper entitled *An Annotated Timeline of Sensitivity Analysis* tries to explore the evolution of sensitivity analysis over the years. The recognition of sensitivity analysis as a pivotal tool for guiding model development and decision-making processes signals its growing importance, particularly in the face of increasingly complex simulation models.

In conclusion, in Chapter 1 we developed an approach for modeling smoking dynamics in the Tuscany population, that overcomes many of the limitations of previously proposed models. In

Chapter 2, we highlight the great potential of likelihood-free methods, that can retrieve estimates and forecasts even when dealing with very complex models, as in the case of the Smoking Habits Compartmental (SHC) model introduced in Chapter 1, that prevent the use of whatever likelihood-based method. In Chapter 3, we propose a procedure for robustifying the inference produced by the SHC model via Global Sensitivity Analysis, which makes the estimates produced in the model defined in Chapter 1 more robust. In Chapter 4, we emphasize the importance of Global Sensitivity Analysis as a tool to guide the development of increasingly complex models. Finally, in Chapter 5, we also highlight all other works published during the Ph.D. period.

This dissertation represents a significant stride in deepening our comprehension of smoking dynamics and refining methodologies for modeling and inference in epidemiological research. By introducing the Smoking Habits Compartmental (SHC) model and investigating various estimation techniques, we have established a robust framework for analyzing smoking behaviors and evaluating the effectiveness of policy interventions. Furthermore, our examination of uncertainty quantification and sensitivity analysis underscores the imperative of addressing model uncertainties in decision-making processes. Looking forward, the methodologies and insights presented herein hold immense potential to inform public health policies, not only in Tuscany but also in regions grappling with similar smoking-related challenges. Through ongoing refinement of our models and methodologies, we can enhance our capacity to forecast future trends, assess the impact of policy interventions, and mitigate the detrimental health consequences of smoking on individuals and communities. The methodological advancements delineated in this dissertation align with the broader objective of fostering more reliable and informed decision-making amidst the complexities and uncertainties of our world. The collaborative support from the Tuscany region, particularly through the Attributable Cancer Burden in Tuscany (ACAB, `www.acab-toscana.it/#progetto`) project funded by the 2018 Tuscany Health Research Call (`www.regione.toscana.it/-/bando-ricerca-salute-2018`), has played a pivotal role in shaping this research. I extend my heartfelt appreciation to all participants involved in this collective endeavor for their invaluable contributions.

APPENDIX A

Supplemental Materials Chapter 1

A.1 Model assumptions

The assumptions underlying the SHC model are summarized below:

- the probability of starting smoking $\gamma_i(a)$ depends on age and smoking intensity;
- people can start smoking between the ages of 14 and 34;
- the probabilities of starting smoking $\gamma_i(a)$ depend on the age through $\gamma(a)$, while the distribution of the level of smoking intensity $\boldsymbol{\pi}$ is assumed to be constant over time and age;
- the distribution of the new smokers by smoking intensity does not depend on age and calendar time;
- smokers do not change their smoking intensity during their entire life (this also implies that if an ex-smoker goes back to smoking, her/his smoking intensity is the same as when she/he first started smoking);
- the probability of stopping smoking $\varepsilon(a)$ depends only on age;
- people can quit smoking only after 20 years of age;
- the probability of relapsing $\eta(c)$ changes with time since smoking cessation, but does not depend on age and smoking intensity;
- after 15 years since smoking cessation, the probability of smoking relapse becomes constant;
- the rates of quitting depend on age but does not depend on the level of smoking intensity;

- an ex-smoker who first relapses, then stops smoking again, becomes a 0-year former smoker;
- the population is closed to immigration and emigration (but we considered new births and deaths);
- the risk of death depends on the age for never smokers ($\delta_N(a)$), both on smoking intensity and age for current smokers ($\delta_{C_i}(a)$), and on time since smoking cessation and age for former smokers ($\delta_F(a,c)$). For the reason of simplicity, we do not consider the level of smoking intensity in the definition of the mortality for former smokers: $\delta_{F_i}(a,c) = \delta_F(a,c)$ for each $i \in \{l,m,h\}$;
- the mortality rate of current smokers does not depend on the time from starting smoking;
- the mortality rate of former smokers does depend on the time from smoking cessation and on age;
- for each individual, only one event among starting, quitting, relapsing, or dying occurs in the year;
- at each time the probabilities of starting and quitting smoking and the probability of smoking relapse are defined among those who do not die during the year (they are conditional probabilities);
- all the transition rates do not change with t.

A.2 Details on the fixed parameters

The values assigned to the fixed parameters of the SHC model are detailed below:

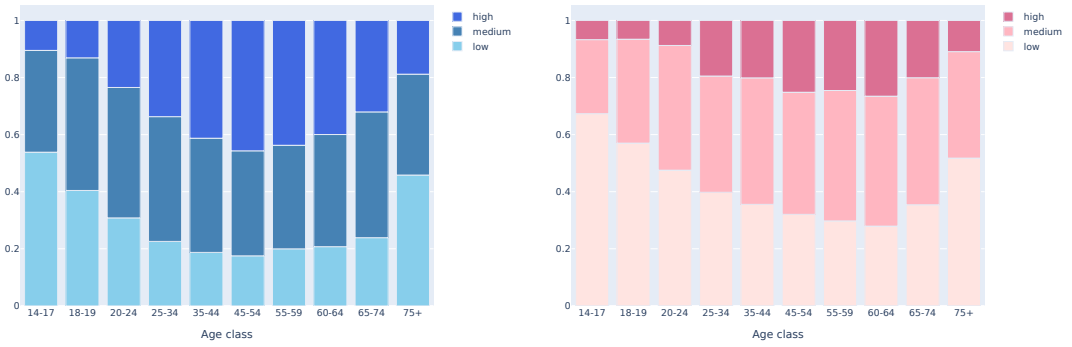
- The vector of proportions $\boldsymbol{\pi}$ was set to the average proportions of low, medium, and heavy smokers estimated for ages 34-44 from the ISTAT AVQ surveys (`www.istat.it/it/archivio/91926`) carried out from 1993 to 2019. Specifically, we set $(\pi_{C_l}, \pi_{C_m}, \pi_{C_h}) = (0.19, 0.40, 0.41)$ for males and $(\pi_{C_l}, \pi_{C_m}, \pi_{C_h}) = (0.36, 0.44, 0.20)$ for females (see Figure A.1).
- $\nu(t)$ was assumed to be constant over time, being the new births quite stable over the years 1993-2019. In particular, we set $\nu(t)$ to the average number of new births in Tuscany from 1993 to 2019, $\nu(t) = 14,701$ for males and $\nu(t) = 13,895$ for females (`http://www.istat.it/`);

- The initial number of never, current, and former smokers, by age and gender, was obtained applying to the resident population in Tuscany on the 1^{st} of January 1993 ($t = 0$) (`http://www.istat.it/`) the prevalence of never, current, and former smokers estimated from the 1993 ISTAT AVQ survey (`www.istat.it/it/archivio/91926`). For details on these quantities, see Figures A.2 and A.3. Note that in the sensitivity analysis described in Section 3.4 of the manuscript, we used also the same quantities referred to the year 2005 (Figures A.4 and A.5).

- In order to quantify $C_i(0;a)$, the initial number of current smokers in 1993, stratified by age, obtained as described at the previous point, has been multiplied by the proportions of low, medium and high-intensity smokers arising from the ISTAT AVQ survey carried out from 1993 to 2019 (Figure A.1). This procedure has been applied separately for males and females.

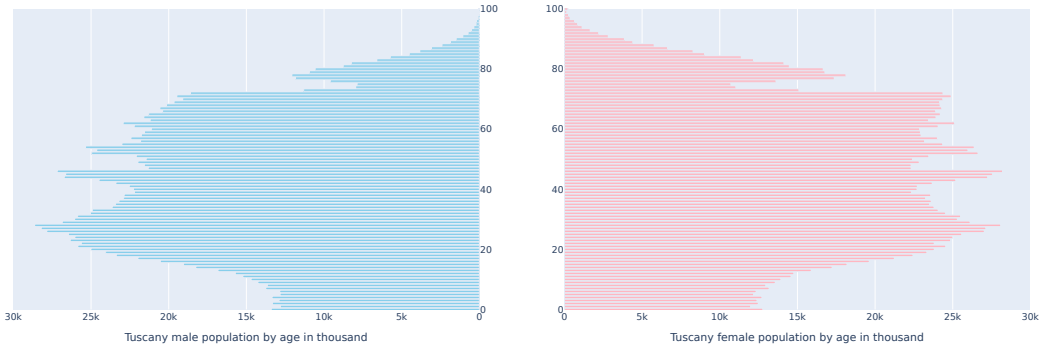
- in order to quantify $F_i(0;a,c)$, first we multiplied the initial number of former smokers in 1993, stratified by age, by the proportions of low, medium, and high-intensity ex-smokers arising from the ISTAT EHIS surveys (`www.istat.it/it/archivio/167485`) carried out in 1994, 1999, 2004, and 2013. Then, we used the distribution of former smokers by time from smoking cessation in 1993 to obtain the initial compartment sizes. This procedure has been applied separately for males and females. For details, see Figures A.6 and A.7.

- The relative risks for current and former smokers versus never smokers were obtained from the literature. Specifically, we used the rates estimated from the US population within the Cancer Prevention Study II [1]. For details see Tables A.1 and A.2.



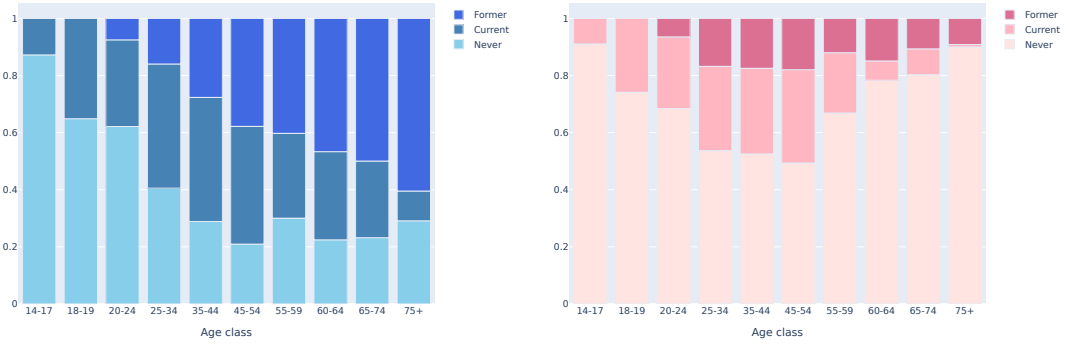
a) b)

FIGURE A.1: Average prevalence of smoking intensity among current smokers evaluated over the period 1993-2019 for males (a) and females (b). Source AVQ survey.



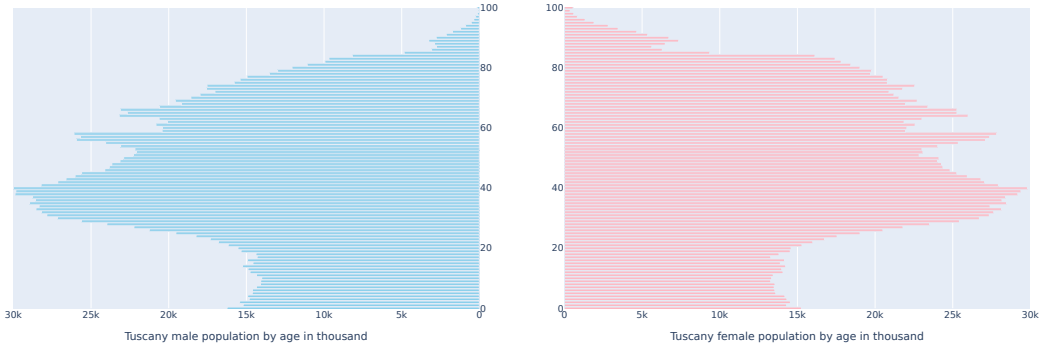
a) b)

FIGURE A.2: Size of the population in 1993 for males (a) and females (b). Source ISTAT.



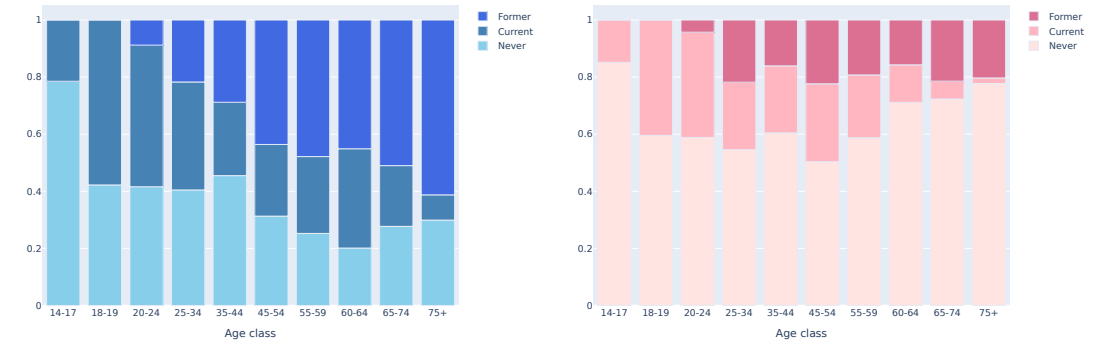
a) b)

FIGURE A.3: Prevalence of smoking habits in 1993 for males (a) and females (b). Source AVQ survey.



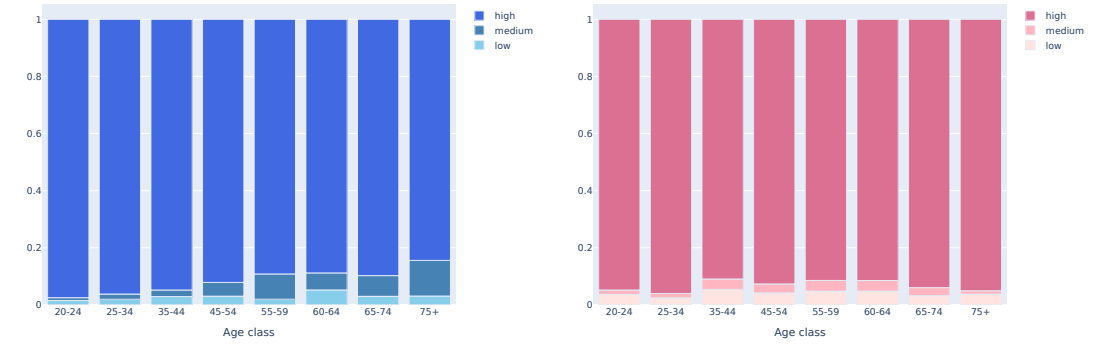
a) b)

FIGURE A.4: Size of the population in 2005 for males (a) and females (b). Source ISTAT.



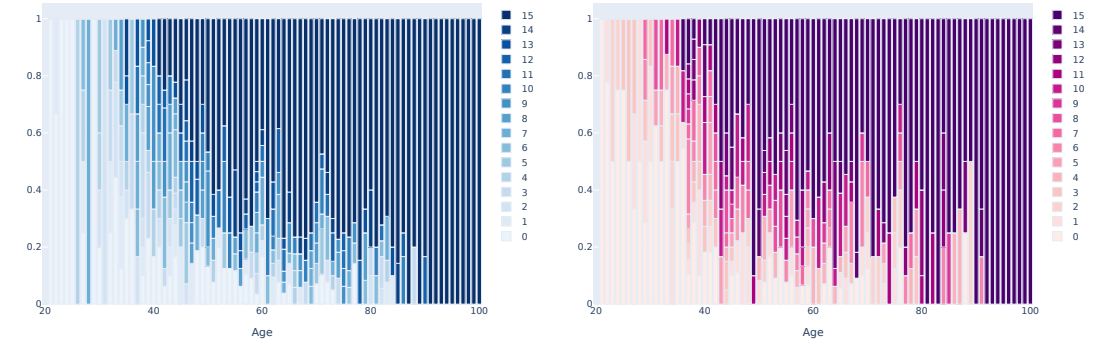
a) b)

FIGURE A.5: Prevalence of smoking habits in 2005 for males (a) and females (b). Source AVQ survey.



a) b)

FIGURE A.6: Average prevalence of smoking intensity among former smokers for males (a) and females (b). Source EHIS survey.



a) b)

FIGURE A.7: Average prevalence of time since smoking cessation among former smokers for males (a) and females (b). Source EHIS survey.

TABLE A.1: Relative Risk of current smokers by smoking intensity (RR_{C_i}) for males and females. Source Thun et al. [1].

Smoke intensity	Male	Female
all	2.43	2.08
low	1.91	1.47
medium	2.05	1.87
high	2.42	2.36

TABLE A.2: Relative Risk of former smokers by time since smoking cessation ($RR_{F_i}(c)$) for males and females. Source Thun et al. [1].

Time from cessation	Male	Female
all	1.43	1.28
<2	2.53	2.26
2-4	2.35	2.22
5-9	1.90	1.58
10-19	1.49	1.29

A.3 Additional details on Global Sensitivity Analysis

TABLE A.3: Distribution of the input parameters used in the GSA and related data sources.

Parameter	Model	Distribution	Source
$\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3)$	all	$\psi_i \sim \mathscr{U}(-10, 10)$ independent	non informative distr.
$\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3)$	all	$\phi_i \sim \mathscr{U}(-10, 10)$ independent	non informative distr.
$\boldsymbol{\omega} = (\omega_0, \omega_1)$	all	$\omega_i \sim \mathscr{U}(0, 10)$ independent	non informative distr.
$\boldsymbol{\pi} = (\pi_{C_l}, \pi_{C_m}, \pi_{C_h})$	males	$\boldsymbol{\pi} \sim \text{Dirichlet}(205^1 \times (0.19, 0.40, 0.41))$	AVQ survey
	females	$\boldsymbol{\pi} \sim \text{Dirichlet}(216^1 \times (0.19, 0.40, 0.41))$	
ν	males	$\nu \sim \text{Poisson}(14,701)$	ISTAT
	females	$\nu \sim \text{Poisson}(13,895)$	
$\boldsymbol{RR}_C = (RR_{C_l}, RR_{C_m}, RR_{C_h})$	males	$\log RR_{C_l} \sim \mathscr{N}_{trun}(\log 1.91, 0.03, 0, \infty)$	Thun et al. [1]
		$\log RR_{C_m} \mid \log RR_{C_l} = a \sim \mathscr{N}_{trun}(\log 2.05, 0.03, a, \infty)$	
		$\log RR_{C_h} \mid \log RR_{C_m} = a \sim \mathscr{N}_{trun}(\log 2.42, 0.02, a, \infty)$	
	females	$\log RR_{C_l} \sim \mathscr{N}_{trun}(\log 1.47, 0.03, 0, \infty)$	
		$\log RR_{C_m} \mid \log RR_{C_l} = a \sim \mathscr{N}_{trun}(\log 1.87, 0.03, a, \infty)$	
		$\log RR_{C_h} \mid \log RR_{C_m} = a \sim \mathscr{N}_{trun}(\log 2.36, 0.03, a, \infty)$	
$RR_F = RR_{F_l}(c) = RR_{F_m}(c) = RR_{F_h}(c) = RR_F(c)$	males	$\log RR_{F_i}(c) \sim \mathscr{N}_{trun}(\log 2.53, 0.03, 0, \infty)$, $c = 1$	Thun et al. [1]
		$\log RR_{F_i}(c) \mid \log RR_{F_i}(1) = a \sim \mathscr{N}_{trun}(\log 2.35, 0.03, 0, a)$, $c \in \{2,3,4\}$	
		$\log RR_{F_i}(c) \mid \log RR_{F_i}(2) = a \sim \mathscr{N}_{trun}(\log 1.90, 0.02, 0, a)$, $c \in \{5,...,9\}$	
		$\log RR_{F_i}(c) \mid \log RR_{F_i}(5) = a \sim \mathscr{N}_{trun}(\log 1.49, 0.02, 0, a)$, $c \geq 10$	
	females	$\log RR_{F_i}(c) \sim \mathscr{N}_{trun}(\log 2.26, 0.05, 0, \infty)$, $c = 1$	Thun et al. [1]
		$\log RR_{F_i}(c) \mid \log RR_{F_i}(1) = a \sim \mathscr{N}_{trun}(\log 2.22, 0.04, 0, a)$, $c \in \{2,3,4\}$	
		$\log RR_{F_i}(c) \mid \log RR_{F_i}(2) = a \sim \mathscr{N}_{trun}(\log 1.58, 0.03, 0, a)$, $c \in \{5,...,9\}$	
		$\log RR_{F_i}(c) \mid \log RR_{F_i}(5) = a \sim \mathscr{N}_{trun}(\log 1.29, 0.03, 0, a)$, $c \geq 10$	
$\boldsymbol{RR}_{status} = (RR_C, RR_F)$	males	$\log RR_C \sim \mathscr{N}_{trun}(\log 2.43, 0.01, 0, \infty)$	Thun et al. [1]
		$\log RR_F \mid \log RR_C = a \sim \mathscr{N}_{trun}(\log 1.43, 0.01, 0, a)$	
	females	$\log RR_C \sim \mathscr{N}_{trun}(\log 2.08, 0.01, 0, \infty)$	Thun et al. [1]
		$\log RR_F \mid \log RR_C = a \sim \mathscr{N}_{trun}(\log 1.28, 0.01, 0, a)$	

[1] Average sample size of the AVQ survey in the period from 1993 to 2019.

A.4 Population Attributable Fraction computation

The Population Attributable Fraction for the class of age a at time t, $\text{PAF}(t;a)$, is calculated as the proportion of deaths that would be avoided if all current and former smokers of age a at time t in the population were never smokers [2]:

$$\text{PAF}(t;a) = \frac{\text{SAD}(t;a)}{D_N(t;a) - D_N(t-1;a) + \sum_i (D_{C_i}(t;a) - D_{C_i}(t-1;a)) + \sum_i \sum_c (D_{F_i}(t;a,c) - D_{F_i}(t;a,c))}.$$

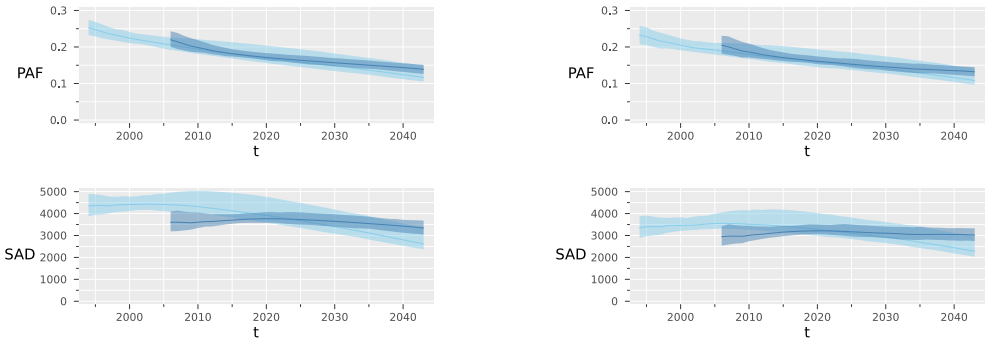
Analogously, the overall PAF at time t is:

$$\text{PAF}(t) = \frac{\text{SAD}(t)}{\sum_a ((D_N(t;a) - D_N(t-1;a)) + \sum_a \sum_i (D_{C_i}(t;a) - D_{C_i}(t-1;a)) + \sum_a \sum_i \sum_c (D_{F_i}(t;a,c) - D_{F_i}(t;a,c)))}.$$

A.5 Additional results

TABLE A.4: Estimated prevalence (%) of never, current, and former smokers in the population with 90% confidence intervals, evaluated every 10 years from 1993 to 2043 for males, by the period of calibration.

	Never		Current		Former	
Year	**1993 - 2004**	**2005 - 2019**	**1993 - 2004**	**2005 - 2019**	**1993 - 2004**	**2005 - 2019**
1993	35.7 (33.8 - 37.7)	-	33.7 (31.8 - 35.7)	-	30.6 (28.6 - 32.5)	-
2003	37.0 (35.4 - 38.5)	-	28.7 (27.9 - 30.3)	-	34.3 (32.7 - 35.3)	-
2013	40.1 (38.3 - 41.7)	39.1 (37.0 - 41.4)	24.1 (23.0 - 26.9)	25.8 (24.2 - 26.8)	35.8 (33.0 - 37.1)	35.1 (33.8 - 36.8)
2023	43.5 (41.2 - 45.4)	43.4 (41.7 - 45.09	21.4 (20.0 - 24.9)	23.2 (21.0 - 24.1)	35.1 (31.6 - 36.7)	33.4 (32.0 - 35.4)
2033	46.8 (44.0 - 49.2)	47.9 (46.3 - 50.9)	20.0 (18.5 - 23.9)	21.4 (18.4 - 22.6)	33.2 (29.4 - 35.1)	30.7 (28.8 - 33.1)
2043	49.5 (45.8 - 52.7)	52.6 (50.8 - 56.2)	19.8 (18.2 - 23.19)	19.6 (16.2 - 21.3)	30.7 (26.7 - 33.0)	27.8 (25.3 - 30.3)



a) b)

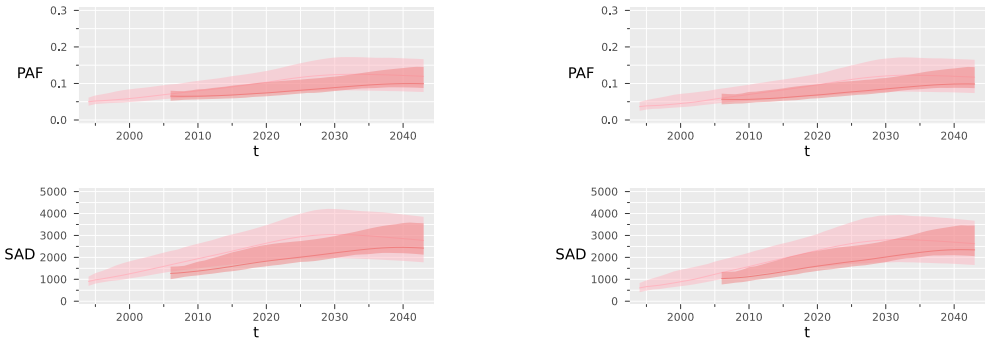
FIGURE A.8: Results of the two-step estimation procedure for males by the period of calibration (from 1993 to 2004 in a light color and from 2005 to 2019 in a dark color): Estimated Population Attributable Fraction (PAF) and the number of Smoking Attributable Deaths (SAD), with 90% confidence bands, for people over 35 years old (a) and over 65 years old (b).

TABLE A.5: Estimated Population Attributable Fraction (PAF) (%) and number of Smoking Attributable Deaths (SAD) in the years 1993, 2003, 2013, 2023, 2033, and 2043, with 90% confidence intervals, among males aged over 35 and 65, by period of calibration.

		PAF		SAD	
Age	**Year**	**1993 - 2004**	**2005 - 2019**	**1993 - 2004**	**2005 - 2019**
35+	**1994**	25.3 (23.2 - 27.4)	-	4,351 (3,900 - 4,858)	-
	2003	21.4 (20.5 - 23.0)	-	4,423 (4,214 - 4,811)	-
	2013	18.6 (17.4 - 20.9)	18.7 (17.6 - 20.2)	4,195 (3,907 - 4,793)	3,659 (3,423 - 3,992)
	2023	16.3 (15.0 - 19.3)	16.6 (15.7 - 18.0)	3,764 (3,476 - 4,507)	3,740 (3,521 - 4,087)
	2033	14.1 (13.1 - 17.2)	15.3 (14.1 - 16.7)	3,229 (2,965 - 3,956)	3,579 (3,291 - 3,922)
	2043	11.6 (10.6 - 14.8)	13.8 (12.5 - 15.3)	2,616 (2,396 - 3,329)	3,337 (2,983 - 3,701)
65+	**1994**	23.2 (20.8 - 25.9)	-	3,352 (2,901 - 3,873)	-
	2003	19.6 (18.5 - 21.3)	-	3,500 (3,299 - 3,855)	-
	2013	17.2 (16.0 - 19.7)	17.6 (16.4 - 19.3)	3,459 (3,200 - 4,000)	3,096 (2,859 - 3,427)
	2023	15.3 (14.0 - 18.3)	15.6 (14.6 - 17.2)	3,168 (2,895 - 3,823)	3,188 (2,969 - 3,526)
	2033	13.4 (12.2 - 16.5)	14.1 (12.9 - 15.6)	2,790 (2,557 - 3,436)	3,049 (2,783 - 3,383)
	2043	10.8 (9.8 - 13.9)	13.2 (11.8 - 14.7)	2,281 (2,071 - 2,921)	3,011 (2,684 - 3,374)

TABLE A.6: Estimated prevalence (%) of never, current, and former smokers in the population with 90% confidence intervals, evaluated every 10 years from 1993 to 2043 for females, by the period of calibration.

	Never		Current		Former	
Year	**1993 - 2004**	**2005 - 2019**	**1993 - 2004**	**2005 - 2019**	**1993 - 2004**	**2005 - 2019**
1993	66.9 (65.0 - 68.7)	-	20.3 (18.6 - 21.9)	-	12.8 (11.4 - 14.2)	-
2003	63.4 (61.8 - 64.9)	-	20.0 (18.9 - 21.0)	-	16.6 (15.7 - 17.7)	-
2013	60.5 (58.7 - 62.0)	63.5 (61.6 - 65.4)	20.2 (18.4 - 22.0)	17.2 (16.1 - 18.2)	19.3 (17.8 - 21.1)	19.3 (18.1 - 20.6)
2023	58.9 (56.5 - 60.7)	62.9 (61.1 - 65.0)	19.9 (17.5 - 22.4)	16.0 (15.0 - 17.8)	21.2 (19.2 - 23.9)	21.1 (19.1 - 22.1)
2033	59.0 (55.8 - 61.4)	62.9 (61.0 - 65.3)	19.3 (16.5 - 22.2)	14.6 (13.6 - 17.3)	21.7 (19.2 - 24.9)	22.4 (19.5 - 23.4)
2043	60.1 (55.9 - 63.4)	64.4 (62.1 - 67.2)	18.8 (15.8 - 22.0)	13.2 (12.2 - 16.6)	21.1 (18.4 - 24.5)	22.5 (18.8 - 23.3)

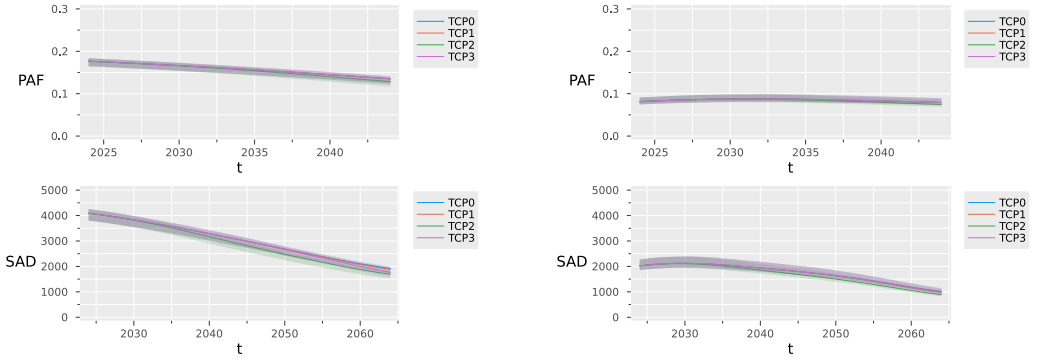


a) b)

FIGURE A.9: Results of the two-step estimation procedure for females by periods of calibration (from 1993 to 2004 in a light color and from 2005 to 2019 in a dark color): Estimated Population Attributable Fraction (PAF) and the number of Smoking Attributable Deaths (SAD), with 90% confidence bands, for people over 35 years old (a) and over 65 years old (b).

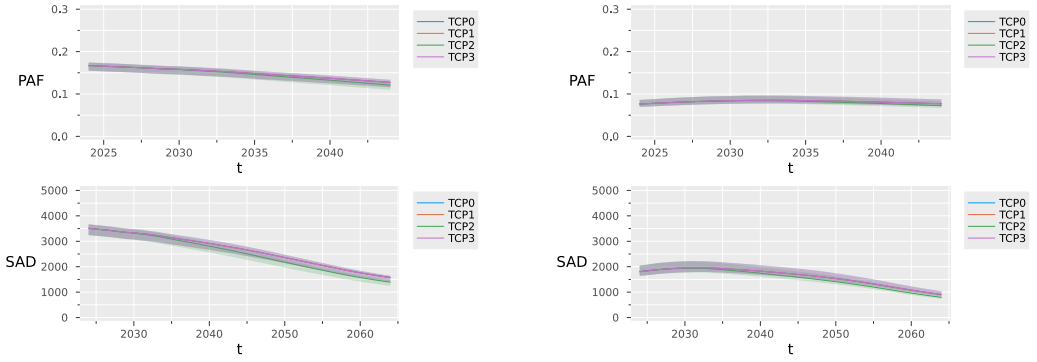
TABLE A.7: Estimated Population Attributable Fraction (PAF) (%) and number of Smoking Attributable Deaths (SAD) in the years 1993, 2003, 2013, 2023, 2033, and 2043, with 90% confidence intervals, among females aged over 35 and 65, by period of calibration.

		PAF		**SAD**	
Age	**Year**	**1993 - 2004**	**2005 - 2019**	**1993 - 2004**	**2005 - 2019**
35+	**1994**	5.0 (4.1 - 6.2)	-	902 (732 - 1,134)	-
	2003	6.5 (5.6 - 8.2)	-	1,455 (1,257 - 1,862)	-
	2013	8.6 (7.3 - 10.8)	6.6 (6.2 - 9.4)	2,147 (1,796 - 2,708)	1,487 (1,387 - 2,146)
	2023	11.2 (9.0 - 13.8)	7.7 (7.4 - 10.8)	2,844 (2,264 - 3,534)	1,914 (1,822 - 2,685)
	2033	12.6 (9.6 - 15.8)	9.2 (8.7 - 12.7)	3,029 (2,307 - 3,821)	2,292 (2,153 - 3,187)
	2043	12.0 (8.9 - 15.3)	9.8 (9.1 - 14.1)	2,786 (2,057 - 3,549)	2,389 (2,231 - 3,477)
65+	**1994**	3.7 (2.7 - 4.9)	-	606 (437 - 828)	-
	2003	5.2 (4.3 - 7.0)	-	1,079 (884 - 1,464)	-
	2013	7.8 (6.5 - 9.9)	5.8 (5.4 - 8.7)	1,812 (1,479 - 2,327)	1,232 (1,134 - 1,869)
	2023	10.6 (8.4 - 13.2)	7.2 (6.8 - 10.3)	2,525 (1,998 - 3,177)	1,696 (1,602 - 2,426)
	2033	12.3 (9.3 - 15.5)	8.9 (8.3 - 12.3)	2,824 (2,124 - 3,584)	2,112 (1,978 - 2,953)
	2043	11.8 (8.6 - 15.1)	9.7 (9.10 - 14.1)	2,635 (1,927 - 3,383)	2,302 (2,144 - 3,372)



a) b)

FIGURE A.10: Estimated Population Attributable Fraction (PAF) and number of Smoking Attributable Deaths (SAD) among people over 35 years of age, with 90% confidence bands, for males (a) and females (b) under different tobacco control policies (TCP).



a) b)

FIGURE A.11: Estimated Population Attributable Fraction (PAF) and number of Smoking Attributable Deaths (SAD) among people over 65 years of age, with 90% confidence bands, for males (a) and females (b) under different tobacco control policies (TCP).

TABLE A.8: Estimated Population Attributable Fraction (PAF) (%) under different tobacco control policies (TCP), in the years 2023, 2033, and 2043, with 90% confidence intervals, among males and females aged over 35 and over 65.

		Male				Female			
Age	**Year**	**TCP0**	**TCP1**	**TCP2**	**TCP3**	**TCP0**	**TCP1**	**TCP2**	**TCP3**
35+	**2023**	17.5 (16.5 - 18.2)	17.5 (16.5 - 18.2)	17.5 (16.5 - 18.2)	17.5 (16.5 - 18.2)	8.5 (7.7 - 9.1)	8.5 (7.7 - 9.1)	8.5 (7.7 - 9.1)	8.5 (7.7 - 9.1)
	2033	15.7 (14.6 - 16.3)	15.7 (14.6 - 16.3)	15.5 (14.4 - 16.2)	15.7 (14.6 - 16.3)	9.3 (8.3 - 9.9)	9.3 (8.3 - 9.9)	9.2 (8.2 - 9.8)	9.3 (8.3 - 9.9)
	2043	15.3 (12.2 - 14.0)	13.3 (12.2 - 14.0)	12.7 (11.6 - 13.3)	13.3 (12.2 - 14.0)	8.4 (7.6 - 9.0)	8.4 (7.6 - 9.0)	8.0 (7.1 - 8.6)	8.4 (7.6 - 9.0)
64+	**2023**	16.6 (15.5 - 17.3)	16.6 (15.5 - 17.3)	16.6 (15.5 - 17.3)	16.6 (15.5 - 17.3)	8.0 (7.2 - 8.6)	8.0 (7.2 - 8.6)	8.0 (7.2 - 8.6)	8.0 (7.2 - 8.6)
	2033	14.9 (13.8 - 15.6)	14.9 (13.8 - 15.6)	14.7 (13.6 - 15.4)	14.9 (13.8 - 15.6)	9.1 (8.1 - 9.7)	9.1 (8.1 - 9.7)	9.0 (7.9 - 9.6)	9.1 (8.1 - 9.7)
	2043	12.6 (11.5 - 13.2)	12.6 (11.5 - 13.2)	11.9 (10.9 - 12.6)	12.6 (11.5 - 13.2)	8.2 (7.3 - 8.8)	8.2 (7.3 - 8.8)	7.8 (6.9 - 8.4)	8.2 (7.3 - 8.8)

Bibliography

[1] M.J. Thun, B.D. Carter, D. Feskanich, N.D. Freedman, R. Prentice, A.D. Lopez, P. Hartge, and S.M. Gapstur. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine*, 368(4):351–364, 2013. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMsa1211127. URL http://www.nejm.org/doi/10.1056/NEJMsa1211127.

[2] GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in the 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*, 397(10292):2337–2360, 2021. ISSN 01406736. doi: 10.1016/S0140-6736(21)01169-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673621011697.

APPENDIX B

Supplemental Materials Chapter 2

B.1 DA-MCMC and detailed balance condition

Let us denote by $\pi(\theta, \mathbf{z} \mid \mathbf{x})$ the target distribution, and by $q(\theta, \mathbf{z} \mid \theta^{(s-1)}, \mathbf{z}^{(s-1)})$ the proposal distribution at the iteration s that, under the assumption of independence, factorizes as follows: $q(\theta, \mathbf{z} \mid \theta^{(s-1)}, \mathbf{z}^{(s-1)}) = q_\theta(\theta \mid \theta^{(s-1)}) q_\mathbf{z}(\mathbf{z} \mid \mathbf{z}^{(s-1)})$. Thus, for a given pair $(\theta^*, \mathbf{z}^*)$, the Metropolis-Hastings acceptance ratio becomes:

$$
\begin{aligned}
r\big((\theta^{(s-1)}, \mathbf{z}^{(s-1)}), (\theta^*, \mathbf{z}^*)\big) &= \frac{\pi(\theta^*, \mathbf{z}^* \mid \mathbf{x}) q(\theta^{(s-1)}, \mathbf{z}^{(s-1)} \mid \theta^*, \mathbf{z}^*)}{\pi(\theta^{(s-1)}, \mathbf{z}^{(s-1)} \mid \mathbf{x}) q(\theta^*, \mathbf{z}^* \mid \theta^{(s-1)}, \mathbf{z}^{(s-1)})} \\
&= \frac{\pi(\theta^*) p(\mathbf{z}^*, \mathbf{x} \mid \theta^*) p(\mathbf{x})}{p(\mathbf{x}) \pi(\theta^{(s-1)}) p(\mathbf{z}^{(s-1)}, \mathbf{x} \mid \theta^{(s-1)})} \frac{q_\theta(\theta^{(s-1)} \mid \theta^*) q_\mathbf{z}(\mathbf{z}^{(s-1)} \mid \mathbf{z}^*)}{q_\theta(\theta^* \mid \theta^{(s-1)}) q_\mathbf{z}(\mathbf{z}^* \mid \mathbf{z}^{(s-1)})},
\end{aligned}
$$

which requires the evaluation of the complete likelihood function and avoids the intractable probability of incomplete data. Furthermore, it is straightforward to show that the detailed balance condition is satisfied since

$$
\pi(\theta^{(s-1)}, \mathbf{z}^{(s-1)} \mid \mathbf{x}) q_\theta(\theta^* \mid \theta^{(s-1)}) q_\mathbf{z}(\mathbf{z}^* \mid \mathbf{z}^{(s-1)}) \min\left\{1, \frac{\pi(\theta^*, \mathbf{z}^* \mid \mathbf{x}) q_\theta(\theta^{(s-1)} \mid \theta^*) q_\mathbf{z}(\mathbf{z}^{(s-1)} \mid \mathbf{z}^*)}{\pi(\theta^{(s-1)}, \mathbf{z}^{(s-1)} \mid \mathbf{x}) q_\theta(\theta^* \mid \theta^{(s-1)}) q_\mathbf{z}(\mathbf{z}^* \mid \mathbf{z}^{(s-1)})}\right\}
$$

is equal to

$$
\pi(\theta^*, \mathbf{z}^* \mid \mathbf{x}) q_\theta(\theta^{(s-1)} \mid \theta^*) q_\mathbf{z}(\mathbf{z}^{(s-1)} \mid \mathbf{z}^*) \min\left\{1, \frac{\pi(\theta^{(s-1)}, \mathbf{z}^{(s-1)} \mid \mathbf{x}) q_\theta(\theta^* \mid \theta^{(s-1)}) q_\mathbf{z}(\mathbf{z}^* \mid \mathbf{z}^{(s-1)})}{\pi(\theta^*, \mathbf{z}^* \mid \mathbf{x}) q_\theta(\theta^{(s-1)} \mid \theta^*) q_\mathbf{z}(\mathbf{z}^{(s-1)} \mid \mathbf{z}^*)}\right\}.
$$

B.2 SIR model: sampling missing data and pseudo-data

In the implementation of the EM algorithm, at each iteration s, we impute missing data drawing samples from the distribution $p(r_{1:T} \mid i^*_{1:T}, \theta^{(s-1)})$, as follows:

- Set initial conditions $I(0), S(0), r(0), i(0)$ and consider observed data $i^*_{1:T}$
- For $t \in \{1, ..., T\}$
  1. Compute $I(t) = I(t-1) + i^*(t-1) - r(t-1)$;
  2. Draw $r(t) \sim \text{Binom}\big(I(t); 1 - \exp(-\gamma^{(s)})\big)$

In ABC algorithms pseudo-data are drawn from using a simulator that reproduces the stochastic data generative process, given $\theta^{(s)}$, as follows:

- Set initial conditions $I(0), S(0), r(0), i(0)$.
- For $t \in \{1, ..., T\}$
  1. Compute $S(t) = S(t-1) - i(t-1)$;
  2. Compute $I(t) = I(t-1) + i(t-1) - r(t-1)$;
  3. Compute $\pi^{(s)}_{\text{SI}}(t) = 1 - \exp\left(-\beta^{(s)} \frac{I(t)}{S(0)} \right)$;
  4. Draw $i(t) \sim \text{Binom}\big(S(t); \pi^{(s)}_{\text{SI}}(t)\big)$
  5. Draw $r(t) \sim \text{Binom}\big(I(t); 1 - \exp(-\gamma^{(s)})\big)$.

B.3 SIR model: further details on the algorithm implementations

B.3.1 Frequentist algorithms: MLE, EM, Calibration

To obtain the ML estimates, we optimized the likelihood function based on complete data. At each iteration i of the optimization, the algorithm checks the convergence. In particular, if $\sqrt{\sum_{j=1}^{i} \frac{(\ell_j - \bar{\ell})^2}{i}} \leq 1.0e^{-08}$, where ℓ_j is the the log-likelihood value reached at iteration j and $\bar{\ell}$ is the mean of the likelihood values until iteration i, the algorithm stops. This optimization outputted a minimum log-likelihood value equal to -507.44. The bootstrap sample size is equal to $1,000$ and the whole procedure (optimization and bootstrap) took 6 seconds.

The EM algorithm has been implemented as displayed in Algorithm 1 with $S = 15,430$ iterations touching 1 hour and a half. The E-step uses an MC estimate based on $m = 1,000$ simulations

at each iteration. The optimization algorithm at each M-step has been stopped as for the ML estimation procedure. The threshold e for the convergence of the EM algorithm has been set at 10^{-10}.

The minimization of the distance function has been performed using the same algorithm as in the MLE procedure and the same assessment of the convergence. The procedure got a minimum distance equal to 0.35. The bootstrap sample size is equal to $1,000$ and the whole procedure (calibration and bootstrap) took 9 seconds.

B.3.2 Bayesian algorithms: IS, DA-MCMC, ABC

The IS algorithm has been implemented with $S = 2,000,000$ iterations. The procedure has been run in parallel on 7 cores and took 3 minutes. The final Effective Sample Size (ESS) is greater than $15,000$, thus ensuring stable estimates.

The DA-MCMC has been run with $S = 10,000,000$ with a running time of 11 minutes. After a burn-in of $6,000,000$ and a thinning of 100 iterations we tested the convergence of the chains using the R-hat statistic [1] (1.023 for τ and 1.005 for R_0) and the Geweke test [2] (p-value equal to 0.372 for τ and 0.263 for R_0). The final ESS is 101, probably due to the strong autocorrelation of the chain. The MH acceptance ratio is 0.48.

The ABC algorithm has been run following the scheme proposed by Lenormand et al. [3]. The algorithm has been run in parallel on 7 cores with a running time of 9 minutes. We ran $S = 149$ iterations by sampling $M = 5,000$ particles per iteration. Furthermore, we computed the Kullback-Leibler [4] and Hellinger [5] divergences between the posterior distributions approximated via DA-MCMC and those approximated at each iteration s of the PMC-ABC. It turned out that these divergences are stable after 80 iterations, suggesting that this number of iterations is enough to get results comparable with those of the DA-MCMC algorithm. The Kullback-Leibler divergence is equal to 0.2 for τ and 0.3 for R_0. The final ESS is 419.

B.4 SHC model

B.4.1 Transition probabilities

Considering that the age a takes values from 0 to 100. The probability of starting $\gamma(a)$ and quitting $\varepsilon(a)$ smoking are defined as follow:

$$\gamma(a) = \begin{cases} 0 & 0 \leq a \leq 13 \cup a \geq 35 \\ \frac{\exp(s(a;\boldsymbol{\psi}))}{1+\exp(s(a;\boldsymbol{\psi}))} & 14 \leq a \leq 34 \end{cases} \qquad \varepsilon(a) = \begin{cases} 0 & 0 \leq a \leq 19 \\ \frac{\exp(s(a;\boldsymbol{\phi}))}{1+\exp(s(a;\boldsymbol{\phi}))} & a \geq 20, \end{cases}$$

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 probabilities of starting and quitting smoking, respectively.

Instead, the relapsing probabilities, $\eta(c)$, were modeled as a negative exponential function of the time since cessation, with parameters $\boldsymbol{\omega} = (\omega_0, \omega_1)$:

$$\eta^*(c) = \begin{cases} 0 & c = 0 \\ 1 - exp(-\omega_0 \omega_1 \exp(-\omega_1 c)) & 1 \leq c \leq 15 \\ 1 - exp(-\omega_0 \omega_1 \exp(-\omega_1 15)) & c \geq 16, \end{cases}$$

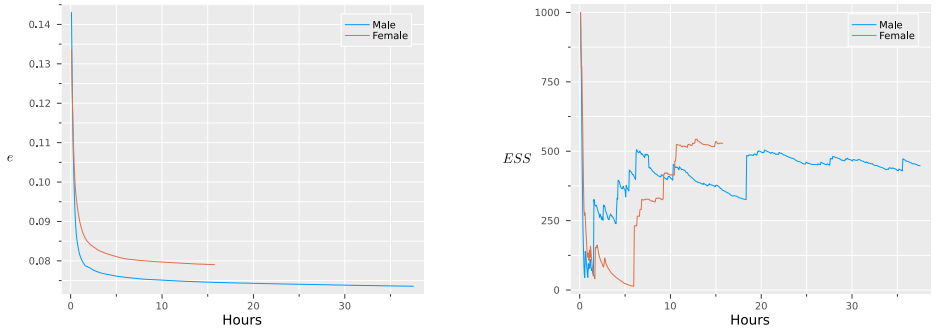
where ω_0 governs the lifetime probability of no relapse and ω_1 tunes how fast the probabilities of smoking relapse declines with the time from cessation [6–9]. Both ω_0 and ω_1 are assumed to be positive, so $\eta^*(c)$ is a positive and decreasing function of c.

B.4.2 Further details of the algorithm implementations

The calibration procedure, along with the bootstrap, took 2 hours. The bootstrap sample has a size equal to $1,000$. The check of the convergence of the optimization algorithm has been implemented as described for the SIR model. The final minimum values of the Hellinger distance are 0.070 and 0.077 in the model for males.

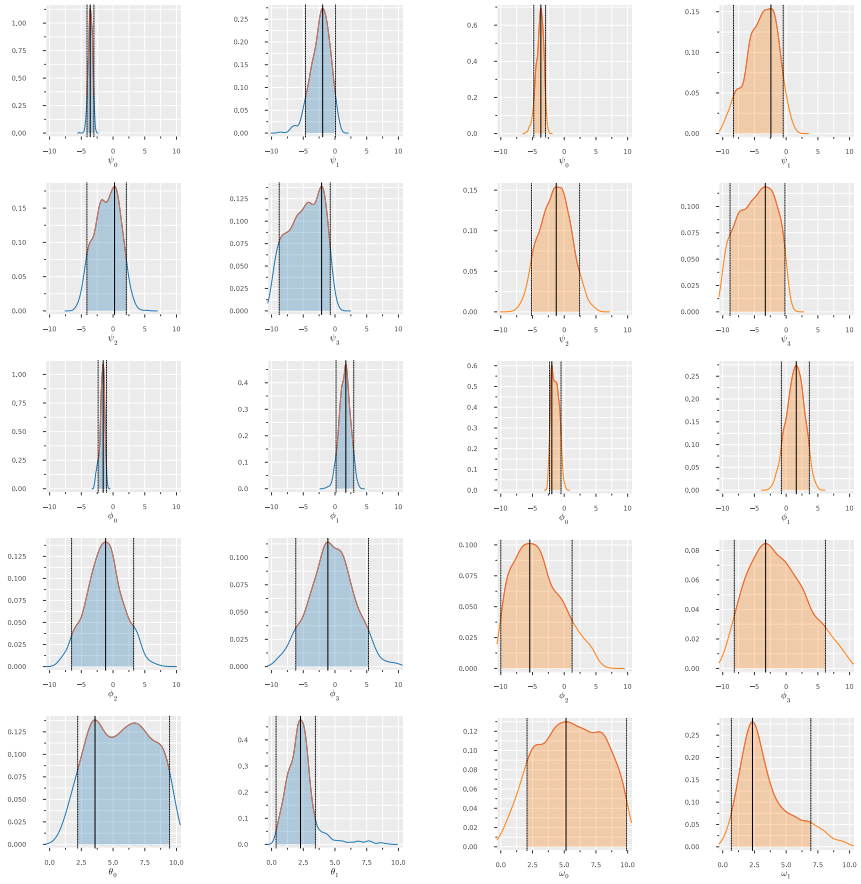
The ABC algorithm uses the stopping rule described by Lenormand et al. [3]. In particular we set $p_{\text{acc}} = 0$, meaning that the algorithm stops when a further decrease of the tolerance threshold would lead to reject the whole sample. Figure B.1 shows the history of the tolerance threshold e (a) and the ESS (b) over the time of the simulation. Looking at (a) we can see that after about 3 hours of simulations the e values are stable and we noted that the approximation of the posterior quantities does not change in a substantial way, suggesting that this is an adequate running time for ABC in this application. Thus, the results reported in the paper, as long as the posterior distributions in Figure B.2 are those obtained after 3 hours.

Both the algorithms have been run in parallel on 50 cores.



a) b)

FIGURE B.1: Evolution of e (a) and ESS (b) in ABC over the time.



a) b)

FIGURE B.2: Posterior density of $\boldsymbol{\theta}$ in SHC for male (a) and female (b).

Bibliography

[1] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457 – 472, 1992. doi: 10.1214/ss/1177011136.

[2] J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis, 1991.

[3] M. Lenormand, F. Jabot, and G. Deffuant. Adaptive approximate bayesian computation for complex models. *Computational Statistics*, 28(6):2777–2796, 2013. doi: doi.org/10.1007/s00180-013-0428-3.

[4] S. Kullback and Leibler R.A. On information and sufficiency. *The Annals of Mathematical Statistics,*, 22(1):79–86, 1951. doi: 10.1214/aoms/1177729694.

[5] E. Hellinger. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136):210–271, 1909. ISSN 1435-5345, 0075-4102. doi: 10.1515/crll.1909.136.210. URL `https://www.degruyter.com/document/doi/10.1515/crll.1909.136.210/html`.

[6] R.T. Hoogenveen, P.Hm. Van Baal, H.C. Boshuizen, and T.L. Feenstra. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Effectiveness and Resource Allocation*, 6(1):1, 2008. ISSN 1478-7547. doi: 10.1186/1478-7547-6-1. URL `http://resource-allocation.biomedcentral.com/articles/10.1186/1478-7547-6-1`.

[7] G. Carreras, S. Gallus, L. Iannucci, and G. Gorini. Estimating the probabilities of making a smoking quit attempt in Italy: stall in smoking cessation levels, 1986-2009. *BMC Public Health*, 12(1):183, 2012. ISSN 1471-2458. doi: 10.1186/1471-2458-12-183. URL `http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-183`.

[8] G. Carreras, G. Gorini, and E. 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*, 5(6):874–882, 2012. ISSN 1940-6207, 1940-6215. doi: 10.1158/1940-6207.CAPR-12-0019. URL

https://aacrjournals.org/cancerpreventionresearch/article/5/6/874/49965/Can-a-National-Lung-Cancer-Screening-Program-in.

[9] G. Carreras, G. Gorini, S. Gallus, L. Iannucci, and D.T. Levy. Predicting the future prevalence of cigarette smoking in Italy over the next three decades. *European Journal of Public Health*, 22(5):699–704, 2012. ISSN 1464-360X, 1101-1262. doi: 10.1093/eurpub/ckr108. URL https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckr108.

APPENDIX C

Supplemental Materials Chapter 3

C.1 MPOWER

TABLE C.1: Italy's MPOWER status in 2021.

MPOWER measure	**Status in Italy (2021)**	**Level of the measure**
Monitor tobacco use and prevention policies	Recent, representative, and periodic data for both adults and youth	Complete
Protect people from tobacco smoke	Complete absence of ban, or up to two public places completely smoke-free	Weak
Offer help to quit tobacco smoke	Nicotine replacement therapy and/or some cessation services (at least one of which is cost-covered)	Moderate
Warn about the dangers of tobacco	Large warnings with all appropriate characteristics; No national campaign conducted between July 2018, and June 2020 with a duration of at least three weeks	Complete Weak
Enforce bans on tobacco advertising, promotion, and sponsorship	Ban on national TV, radio, and print media as well as on some but not all other forms of direct and/or indirect advertising	Moderate
Raise taxes on tobacco	76.6% of retail price is taxes; Reduced affordability between 2010 and 2020[1]	Complete Yes

[1]Per capita GDP needed to buy cigarettes increased on average.

C.2 Total index

The total indexes are calculated according to the matrix-based trick proposed by Saltelli et al. [1].

TABLE C.2: Total variance indices quantifying the contribution of each input on the model outputs in GSA function following *procedure* 1, for males and females.

	ν		RR		π		τ^{TCP}		τ^{window}	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Hellinger distance	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.94	0.94
Prevalence of never-smokers in 2013	0.09	0.06	0.1	0.05	0.09	0.05	0.1	0.05	0.4	0.82
Prevalence of never-smokers in 2023	0.26	0.08	0.26	0.07	0.26	0.07	0.27	0.07	0.58	0.9
Prevalence of never-smokers in 2033	0.3	0.13	0.3	0.13	0.3	0.12	0.34	0.14	0.79	0.89
Prevalence of never-smokers in 2043	0.23	0.15	0.22	0.15	0.22	0.14	0.32	0.2	0.85	0.88
Prevalence of never-smokers in 2053	0.17	0.13	0.16	0.13	0.16	0.13	0.27	0.21	0.87	0.89
Prevalence of never-smokers in 2063	0.13	0.11	0.12	0.11	0.12	0.11	0.24	0.21	0.86	0.88
Prevalence of current-smokers in 2013	0.39	0.08	0.36	0.09	0.36	0.09	0.34	0.09	0.78	0.98
Prevalence of current-smokers in 2023	0.54	0.09	0.51	0.09	0.51	0.1	0.49	0.1	0.84	1.0
Prevalence of current-smokers in 2033	0.5	0.09	0.48	0.09	0.48	0.1	0.58	0.12	0.71	0.99
Prevalence of current-smokers in 2043	0.34	0.08	0.32	0.08	0.33	0.08	0.57	0.13	0.63	0.96
Prevalence of current-smokers in 2053	0.22	0.06	0.21	0.06	0.21	0.07	0.47	0.14	0.65	0.95
Prevalence of current-smokers in 2063	0.18	0.05	0.17	0.05	0.17	0.06	0.47	0.14	0.63	0.93
Prevalence of former-smokers in 2013	0.32	0.32	0.29	0.32	0.3	0.29	0.27	0.31	0.66	0.68
Prevalence of former-smokers in 2023	0.32	0.42	0.29	0.43	0.3	0.38	0.27	0.41	0.77	0.69
Prevalence of former-smokers in 2033	0.26	0.39	0.25	0.39	0.25	0.36	0.25	0.41	0.77	0.71
Prevalence of former-smokers in 2043	0.25	0.43	0.24	0.42	0.24	0.38	0.25	0.44	0.77	0.72
Prevalence of former-smokers in 2053	0.22	0.5	0.21	0.49	0.21	0.44	0.23	0.52	0.84	0.69
Prevalence of former-smokers in 2063	0.15	0.47	0.14	0.47	0.15	0.42	0.16	0.53	0.91	0.68
SAD over 35 years old in 2013	0.09	0.07	0.49	0.13	0.08	0.07	0.08	0.07	0.5	0.88
SAD over 35 years old in 2023	0.21	0.08	0.85	0.12	0.19	0.09	0.18	0.09	0.27	0.96
SAD over 35 years old in 2033	0.18	0.15	0.71	0.21	0.16	0.16	0.15	0.16	0.37	0.94
SAD over 35 years old in 2043	0.12	0.29	0.42	0.39	0.1	0.31	0.11	0.3	0.61	0.8
SAD over 35 years old in 2053	0.09	0.35	0.28	0.46	0.08	0.37	0.09	0.37	0.71	0.72
SAD over 35 years old in 2063	0.13	0.39	0.38	0.51	0.12	0.41	0.14	0.41	0.61	0.64
SAD over 64 years old in 2013	0.1	0.07	0.6	0.14	0.1	0.07	0.09	0.07	0.37	0.86
SAD over 64 years old in 2023	0.19	0.09	0.84	0.13	0.18	0.1	0.16	0.09	0.25	0.95
SAD over 64 years old in 2033	0.17	0.16	0.76	0.21	0.15	0.16	0.14	0.16	0.31	0.94
SAD over 64 years old in 2043	0.1	0.31	0.39	0.42	0.09	0.33	0.09	0.33	0.63	0.78
SAD over 64 years old in 2053	0.07	0.38	0.23	0.51	0.06	0.41	0.07	0.4	0.75	0.69
SAD over 64 years old in 2063	0.09	0.39	0.29	0.52	0.08	0.42	0.09	0.41	0.69	0.63
Minimum prevalence of current-smokers	0.23	0.05	0.22	0.05	0.22	0.05	0.52	0.14	0.62	0.94
Maximum prevalence of never-smokers	0.13	0.17	0.12	0.19	0.12	0.16	0.24	0.36	0.86	0.84
Maximum probability of starting smoking	0.51	0.23	0.49	0.21	0.49	0.2	0.54	0.21	0.82	0.93
Maximum probability of relapsing smoking	0.64	0.63	0.61	0.66	0.59	0.62	0.59	0.58	0.82	0.89
Age at maximum probability of starting smoking	0.06	0.4	0.06	0.4	0.06	0.37	0.06	0.4	0.99	0.76
Age at maximum probability of quitting smoking	0.53	0.01	0.66	0.01	0.66	0.01	0.47	0.01	0.69	1.0
Mean probability of starting smoking	0.08	0.08	0.08	0.07	0.08	0.07	0.08	0.08	0.96	1.0
Mean probability of quitting smoking	0.62	0.04	0.58	0.03	0.58	0.03	0.61	0.03	0.76	0.99
Mean probability of relapsing smoking	0.58	0.14	0.52	0.15	0.53	0.14	0.53	0.14	0.88	0.98
Mean age of starting smoking	0.03	0.04	0.03	0.03	0.03	0.03	0.03	0.04	0.99	0.98
Mean age of quitting smoking	0.12	0.0	0.14	0.0	0.13	0.0	0.11	0.0	0.87	0.99
Mean time of relapsing smoking	0.27	0.31	0.25	0.32	0.25	0.32	0.25	0.31	0.96	0.9

TABLE C.3: Total variance indices quantifying the contribution of each input on the model outputs in GSA function following *procedure* 2, for males and females.

	v		RR		π		χ^{TCP1}		χ^{TCP2}		χ^{TCP3}		χ^{TCP4}		τ^{window}	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Prevalence of never smokers in 2053	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Prevalence of never smokers in 2063	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Prevalence of current smokers in 2053	0.15	0.03	0.14	0.03	0.14	0.04	0.15	0.04	0.14	0.04	0.74	0.8	0.63	0.41	0.14	0.22
Prevalence of current smokers in 2063	0.22	0.1	0.2	0.08	0.18	0.06	0.23	0.08	0.27	0.06	0.76	0.69	0.66	0.56	0.36	0.29
SAD above 35+ in 2053	0.08	0.15	0.14	0.17	0.08	0.14	0.11	0.19	0.09	0.18	0.11	0.2	0.31	0.93	0.83	0.7
SAD above 35+ in 2063	0.09	0.07	0.15	0.09	0.1	0.07	0.14	0.07	0.2	0.07	0.32	0.1	0.16	0.35	0.76	0.79
SAD above 65+ in 2053	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAD above 65+ in 2063	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Bibliography

[1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, Ltd, 2008. ISBN 978-0-470-05997-5.

APPENDIX D

Supplemental Materials Chapter 4

D.1 Global Sensitivity Analysis

While sensitivity analysis theory and techniques rapidly developed, a new community of practitioners also embarked on an epistemological journey, delving into the concept of Global Sensitivity Analysis. Traditionally, sensitivity analysis in various disciplines had been dominated by “local" approaches, involving small perturbations around a reference value while keeping other input parameters fixed. Although this method provided computationally efficient results, it proved inadequate for non-linear and non-additive models or when interactions between inputs played a significant role [1].

In contrast, Global Sensitivity Analysis offered a more comprehensive perspective by simultaneously varying model inputs across a wide range of uncertainty, capturing possible nonlinearities and interactions among parameters. This approach resulted in a broader understanding of input-output dependencies and mitigated the risk of type II errors (nonidentification of influential factors) associated with the traditional one-at-a-time or derivative-based SA.

From the early 1990s, the focus shifted towards exploring and applying new GSA methods across diverse domains, quickly gaining prominence in modeling. Among the early global approaches were variance-based methods [2], screening methods [3], non-parametric or regression-based approaches [4, 5], and density-based analysis [6]. Over time, Global Sensitivity Analysis continued to evolve and expand, with numerous avenues explored by researchers and practitioners alike [7–11].

Despite the extensive activity in the field, Global Sensitivity Analysis remains relatively underrecognized as a crucial component of the modeling process, and its applications are not widespread across many fields. The journey of sensitivity-analysis practitioners and the development of Global Sensitivity Analysis techniques have been lengthy, aiming to achieve more comprehensive

insights and enhanced reliability in scientific research and decision-making processes. There is still work to be done to fully integrate Global Sensitivity Analysis into various domains and realize its potential as a powerful tool for improving model evaluation and understanding complex systems.

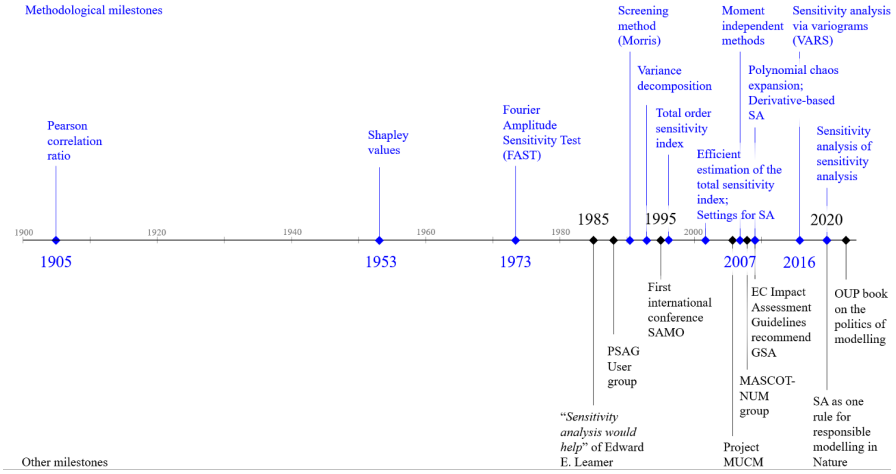


FIGURE D.1: Milestones of sensitivity analysis: publications, projects and scientific meetings.

D.1.1 Using of GSA in compartmental model

Compartmental models are a class of models used to understand and describe the dynamic evolution of a phenomenon of interest in a population. Due to their simple mechanistic nature, they are widely used for modeling infectious diseases. An important issue in compartmental models concerns parameter identifiability [12]. Complex models with many compartments have many parameters governing the admitted transitions, but unfortunately observed data are often insufficient to estimate all of them. To overcome this problem we fixed some of the parameters to values from the literature or external data.

If we consider a forward perspective, GSA allows us to better understand how the parameters affect the total variance of such model output. We considered as model parameters all the model inputs, known or unknown, that predict epidemic dynamics. With GSA, different parameters are combined to obtain, for instance, a prediction of the number of infections or the date of an epidemic peak. In this way, GSA propagates uncertainty from the inputs to the outputs and returns a probability distribution for each uncertain input. If we consider as model output the discrepancy measure used for the estimation procedure of unknown model parameters, GSA allows us to better understand who are the parameters affected by identifiable problems.

A different perspective of the use of GSA considers the calibration procedure as the model function. The parameters that are not being estimated and the observed data represent the inputs of the model. The estimates of the model parameters assumed as unknown and other quantities of

interest represent the outputs of the model. The goal is to assess the robustness of the estimates to changes in the inputs. This procedure takes the name of modeling of the modeling process, introduced by Piano et al. [13].

D.2 Variance based decomposition

D.2.1 Sobol decomposition

Let us defined a model $Y = f(X)$, where X represent the vector of k factor $X_1, ..., X_k$. Thanks to the Sobol decomposition [2], we can write $Y = f_0 + \sum_{i=1}^{k} f_i + \sum_{i<j}^{k} f_{ij} + ... + f_{123...k}$ where $f_0 = E(Y)$ is the expected value of Y and $f_{S_1S_2...S_j} = f_{S_1S_2...S_j}(X_{S_1}, X_{S_2}, ..., X_{S_j})$ is a general function of j factor.

For the reason of simplicity, let us consider the case of $X_i \sim U[0,1]$ where $P(X_j) = 1$. In this way, we can write

$$\int_0^1 \int_0^1 ... \int_0^1 f_{S_1S_2...S_j} dX_{S_1}, dX_{S_2}, ..., dX_{S_j}$$

instead of

$$\int \int ... \int f_{S_1S_2...S_j} p_{S_1}(X_{S_1}) p_{S_2}(X_{S_2}) ... p_{S_j}(X_{S_j}) dX_{S_1}, dX_{S_2}, ..., dX_{S_j}.$$

A condition for the unicity of the Sobol variance decomposition is that

$$\int_0^1 \int_0^1 ... \int_0^1 f_{S_1S_2...S_j} dX_{S_1}, dX_{S_2}, ..., dX_{S_j} = 0.$$

In this case, all the terms of the functional decomposition are independent and this leads to definitions of the terms of the functional decomposition in terms of conditional expected values.

Let us defining $f_i(X_i) = E_{X_{\sim i}}(Y|X_i) - f_0$ as the effect of varying X_i, $f_{ij}(X_i, X_j) = E_{X_{\sim ij}}(Y|X_i, X_j) - f_i - f_j - f_0$ as the effect of varying X_i and X_j simultaneously, and so on. If we assume that $f(X)$ is square-integrable, the functional decomposition may be squared and integrated to give

$$V(Y) = E[Y^2] - E[Y]^2 = \underbrace{\int f^2(X) dX}_{E(Y^2)} - \underbrace{f_0^2}_{E(Y)^2} = \sum_{i=1}^{k} \sum_{S_1<...<S_j}^{k} f_{S_1,...,S_j} dX_{S_j}.$$

This finally leads to the decomposition of variance expression $V(Y) = \sum_{i=1}^{k} V_i + \sum_{i<j}^{k} V_{ij} + \ldots + V_{123\ldots k}$ where $V_i = V_{X_i}(E_{X\sim i}(Y|X_i))$ is the first order effect or top marginal variance, that represents the expected reduction of the variance that would be achieved if the factor X_i could be fixed, $V_{X_{ij}}(E_{X\sim ij}(Y|X_{ij})) = V_i + V_j + V_{ij}$ is the second order effect, that represents the expected reduction of the variance that would be achieved if the factors X_i and X_j could be fixed, and so on. Note that, when the factors are independent the total variance can be decomposed into main effects and interaction effects up to the order k.

If we consider the first-order effect, we can decompose $V(Y)$ as:

$$V(Y) = V_{X_i}(E_{X\sim i}(Y|X_i)) + E_{X_i}(V_{X_{\sim i}}(Y|X_i)),$$

where $V_{X_i}(E_{X\sim i}(Y|X_i))$, the first order effect, is the main effect and $E_{X_i}(V_{X_{\sim i}}(Y|X_i))$ is the residuals effect.

Instead, if we consider $V_{X_{\sim i}}(E_{X_i}(Y|X_{\sim i}))$, that represents the effect not due to X_i, we can decompose $V(Y)$ as:

$$V(Y) = V_{X_{\sim i}}(E_{X_i}(Y|X_{\sim i})) + E_{X\sim i}(V_{X_i}(Y|X_{\sim i})),$$

where $V_{X_{\sim i}}(E_{X_i}(Y|X_{\sim i}))$ is the main effect, and the residual is given by$E_{X\sim i}(V_{X_i}(Y|X_{\sim i}))$ that represent the total order effect, or bottom marginal variance, that is the expected variance that would be left if all the factors excluded X_i could be fixed.

Finally, if we normalize the sensitivities effect we obtain the sensitivities index:

- the first order index is given by $S_i = \frac{V_{X_i}(E_{X\sim i}(Y|X_i))}{V(Y)}$;
- the second order index is given by $S_{ij} = \frac{V_{X_{ij}}(E_{X\sim ij}(Y|X_{ij}))}{V(Y)}$;
- the second order total order index $S_{T_i} = \frac{E_{X\sim i}(V_{X_i}(Y|X_{\sim i}))}{V(Y)} = 1 - \frac{V_{X_{\sim i}}(E_{X_i}(Y|X_{\sim i}))}{V(Y)}$.

In general:

- $S_{T_i} = 0$ means that X_i is a non-influential factor;
- $S_{T_i} \approx S_i$ means that the interaction between X_i and the other factors does not affect the variability of the output;
- $\sum_i S_i = 1$ if the model is additive and in general $\sum_i S_{T_i} \geq 1$.

Proof. of $V(Y) = E(Y^2) - E(Y)^2$

$$\begin{aligned} V(Y) =& E((Y - E(Y))^2) = E(Y^2 - 2YE(Y) + E(Y)^2) = \\ =& E(Y^2) - 2E(Y)E(Y) + E(Y)^2 = E(Y^2) - 2E(Y)^2 + E(Y)^2 = E(Y^2) - E(Y)^2 \end{aligned}$$

□

Proof. of $E(E(Y|X)) = E(Y)$ (more in general $E(E(Y^k|X)) = E(Y^k)$)

$$\begin{aligned} E(E(Y|X)) =& \int_{-\infty}^{\infty} E(Y|X = x) f_x(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x) dy dx = \\ =& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y,X}(y,x) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y,X}(y,x) dx dy = \int_{-\infty}^{\infty} y f_Y(y) dy = E(Y) \end{aligned}$$

□

Proof. of $E(Y^2|X) = V(Y|X) + E(Y|X)^2$

$$\text{Define } V(Y|X) = E(Y^2|X) - E(Y|X)^2$$

$$\text{Adding to either member } E(Y|X)^2 \text{ and simplify}$$

$$V(Y|X) + E(Y|X)^2 = E(Y^2|X) \cancel{- E(Y|X)^2} \cancel{+ E(Y|X)^2}$$

□

Proof. of $V(Y) = E(V(Y|X)) + V(E(Y|X))$

$$\text{Define } E(Y^2) = E(E(Y^2|X)) = E(V(Y|X) + E(Y|X)^2)$$

$$\text{Now } E(Y^2) - E(Y)^2 = E(V(Y|X) + E(Y|X)^2) - E(E(Y|X))^2$$

Since the expectation of a sum is the sum of expectations, the terms can be regrouped

$$E(V(Y|X)) + (E(E(Y|X)^2) - E(E(Y|X))^2)$$

The second term is $V(E(Y|X))$ and finally we obtain

$$V(E(Y|X)) + E(V(Y|X))$$

□

D.2.2 Computation of sensitivities index

Let us apply the relation $V(Y) = E(Y^2) - E(Y)^2$ to $V_{X_i}(E_{X\sim i}(Y|X_i))$. We obtain that:

$$\underbrace{\int E^2_{X_{\sim i}}(Y|X_i)dX_i}_{E(Y^2)} - \underbrace{(\int E_{X_{\sim i}}(Y|X_i)dX_i)^2}_{E(Y)^2}.$$

We note that the second term, $E(Y)^2$, is equal to f_0^2. Instead, $E^2_{X_{\sim i}}(Y|X_i)dX_i$ can be expressed as the product between $E_{X_{\sim i}}(Y|X_i)$ and $E_{X'_{\sim i}}(Y'|X_i)$:

$$\begin{aligned} E^2_{X_{\sim i}}(Y|X_i)dX_i =& E_{X_{\sim i}}(Y|X_i)E_{X'_{\sim i}}(Y'|X_i) = \\ =& \int\int ... \int f(X_1,...,X_k)f(X'_1,...,X'_{i-1},X_i,X'_{i+1},...,X'_k)dX_{\sim i},dX'_{\sim i}. \end{aligned}$$

Now, integrates over X_i the previous expected value $E^2_{X_{\sim i}}(Y|X_i)dX_i$ considering the previous integral:

$$\begin{aligned} \int E^2_{X_{\sim i}}(Y|X_i)dX_i =& \int\int ... \int f(X_1,...,X_k)f(X'_1,...,X'_{i-1},X_i,X'_{i+1},...,X'_k)dX_{\sim i},dX'_{\sim i},dX_i = \\ =& \int\int ... \int f(X_1,...,X_k)f(X'_1,...,X'_{i-1},X_i,X'_{i+1},...,X'_k)dX,dX'_{\sim i}. \end{aligned}$$

This integral represents the expectation in the $k+k-1$ dimension of two function values, wherein in the second one all is re-sampled excluding the factor X_i:

$$\underbrace{f(X_1,...,X_k)}_{f}\underbrace{f(X'_1,...,X'_{i-1},X_i,X'_{i+1},...,X'_k)}_{f'}.$$

Wrapping all together we obtain that $V_{X_i}(E_{X\sim i}(Y|X_i)) = E_{XX'_{\sim i}}(ff') - f_0^2$.

We can reproduce this scheme via Monte Carlo simulation [2, 14, 15]:

1. Sample AB= $\begin{bmatrix} X_{1,1} & X_{1,2} & ... & X_{1,2k} \\ X_{2,1} & X_{2,2} & ... & X_{2,2k} \\ ... & ... & ... & ... \\ X_{N,1} & X_{N,2} & ... & X_{N,2k} \end{bmatrix}$

2. Split in A= $\begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,k} \\ X_{2,1} & X_{2,2} & \dots & X_{2,k} \\ \dots & \dots & \dots & \dots \\ X_{N,1} & X_{N,2} & \dots & X_{N,k} \end{bmatrix}$ and B= $\begin{bmatrix} X_{1,k+1} & X_{1,k+2} & \dots & X_{1,2k} \\ X_{2,k+1} & X_{2,k+2} & \dots & X_{2,2k} \\ \dots & \dots & \dots & \dots \\ X_{N,k+1} & X_{N,k+2} & \dots & X_{N,2k} \end{bmatrix}$

3. Define $A_i^B = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,k+i} & \dots & X_{1,k} \\ X_{2,1} & X_{2,2} & \dots & X_{2,k+i} & \dots & X_{2,k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N,1} & X_{N,2} & \dots & X_{N,k+i} & \dots & X_{N,k} \end{bmatrix}$

Finally, as resume in Saltelli et al. [8], calculate First and Total sensitivities indexes:

- $S_i = \frac{\frac{1}{N}\sum_n^N f(B)_n (f(A_i^B)_n - f(A)_n)}{V(Y)}$;
- $S_{T_i} = \frac{\frac{1}{N}\sum_n^N f(A)_n (f(A)_n - f(A_i^B)_n)}{V(Y)}$,

Where $V(Y) = \frac{1}{N}\sum_n^N (f(A)_n - f_0)^2$ and $f_0 = \frac{1}{N}\sum_n^N f(A)_n$. Note that in S_i, $f(B)_n \times f(A_i^B)_n = E_{XX'_{\sim i}}$ and $f(B)_n \times f(A)_n = f_0^2$.

Example of Monte Carlo matrix sampling scheme:

- AB= $\begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.25 & 0.75 & 0.25 & 0.75 & 0.25 & 0.75 \\ 0.75 & 0.25 & 0.75 & 0.25 & 0.75 & 0.25 \\ 0.125 & 0.625 & 0.875 & 0.875 & 0.625 & 0.125 \end{bmatrix}$
- A= $\begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.25 & 0.75 & 0.25 \\ 0.75 & 0.25 & 0.75 \\ 0.125 & 0.625 & 0.875 \end{bmatrix}$ B= $\begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0.75 \\ 0.25 & 0.75 & 0.25 \\ 0.875 & 0.625 & 0.125 \end{bmatrix}$
- $A_1^B = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.75 & 0.75 & 0.25 \\ 0.25 & 0.25 & 0.75 \\ 0.875 & 0.625 & 0.875 \end{bmatrix}$ $A_2^B = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.25 \\ 0.75 & 0.75 & 0.75 \\ 0.125 & 0.625 & 0.875 \end{bmatrix}$ $A_3^B = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.25 & 0.75 & 0.75 \\ 0.75 & 0.25 & 0.25 \\ 0.125 & 0.625 & 0.125 \end{bmatrix}$

Bibliography

[1] A. Saltelli and P. Annoni. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517, 2010. ISSN 13648152. doi: 10.1016/j.envsoft.2010.04.012. URL https://linkinghub.elsevier.com/retrieve/pii/S1364815210001180.

[2] I.M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.*, 1:407, 1993.

[3] M.D. Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174, 1991. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.1991.10484804. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.1991.10484804.

[4] A. Saltelli and J. Marivoet. Non-parametric statistics in sensitivity analysis for model output: A comparison of selected techniques. *Reliability Engineering & System Safety*, 28 (2):229–253, 1990. ISSN 09518320. doi: 10.1016/0951-8320(90)90065-U. URL https://linkinghub.elsevier.com/retrieve/pii/095183209090065U.

[5] J.C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, 42(2-3):327–367, 1993. ISSN 09518320. doi: 10.1016/0951-8320(93)90097-I. URL https://linkinghub.elsevier.com/retrieve/pii/095183209390097I.

[6] C.K. Park and K. Ahn. A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliability Engineering & System Safety*, 46(3):253–261, 1994. ISSN 09518320. doi: 10.1016/0951-8320(94)90119-8. URL https://linkinghub.elsevier.com/retrieve/pii/0951832094901198.

[7] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007. ISSN 09518320. doi: 10.1016/j.ress.2006.04.015. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832006000883.

[8] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018. URL https://linkinghub.elsevier.com/retrieve/pii/S0010465509003087.

[9] T.A. Mara and S. Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121, 2012. ISSN 09518320. doi: 10.1016/j.ress.2011.08.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832011001724.

[10] S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937–946, 2012. ISSN 00104655. doi: 10.1016/j.cpc.2011.12.020. URL https://linkinghub.elsevier.com/retrieve/pii/S0010465511004085.

[11] E. Plischke, E. Borgonovo, and C.L. Smith. Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550, 2013. ISSN 03772217. doi: 10.1016/j.ejor.2012.11.047. URL https://linkinghub.elsevier.com/retrieve/pii/S0377221712008995.

[12] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3):379–398, 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042717300234.

[13] S. Lo Piano, R. Sheikholeslami, A. Puy, and A. Saltelli. Unpacking the modelling process via sensitivity auditing. *Futures*, 144:103041, 2022. ISSN 00163287. doi: 10.1016/j.futures.2022.103041. URL https://linkinghub.elsevier.com/retrieve/pii/S0016328722001410.

[14] I.M. Sobol. *A primer for the Monte Carlo method*. CRC Press, Boca Raton, 1994. ISBN 978-0-8493-8673-2.

[15] A. Saltelli, S. Tarantola, and K.P.S. Chan. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1):39–56, 1999. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.1999.10485594. URL http://www.tandfonline.com/doi/abs/10.1080/00401706.1999.10485594.