From smoking scenarios to lung cancer mortality: Sequential Bayesian APC projections

dECON
Facultad de Ciencias Sociales

dECON
Facultad de Ciencias Sociales
UNIVERSIDAD DE LA REPÚBLICA

Ciencias Sociales
Universidad de la República
URUGUAY

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Documentos de Trabajo

Titulo

From smoking scenarios to lung cancer mortality: Sequential Bayesian APC projections

Manuel Flores

Documento No. 04/26
Julio 2026

ISSN 0797-7484

Resumen

Este documento desarrolla un marco bayesiano secuencial edad–período–cohorte (APC) para proyectar la mortalidad por cáncer bajo escenarios alternativos de tabaquismo usando datos agregados de población. El enfoque estima superficies de prevalencia de tabaquismo a partir de encuestas transversales repetidas, reconstruye la composición de estados de tabaquismo por cohorte, traduce la exposición efectiva en incidencia mediante riesgos relativos y trayectorias de reversión del riesgo, y proyecta la mortalidad a partir de la incidencia mediante calendarios externos de mortalidad posterior al diagnóstico. El método se evalúa mediante simulaciones con efectos de escenario conocidos y se ilustra con datos de cáncer de pulmón de Uruguay, incluyendo una extensión a nueve sitios tumorales relacionados con el tabaquismo.

Palabras clave: modelos edad–período–cohorte; métodos bayesianos; proyecciones de cáncer; tabaquismo; modelización epidemiológica.

Abstract

This working paper develops a sequential Bayesian age–period–cohort (APC) framework for projecting cancer mortality under alternative smoking scenarios using aggregate population data. The framework estimates smoking-prevalence surfaces, reconstructs cohort smoking states, maps effective exposure into incidence through external relative risks and risk-reversion schedules, and translates projected incidence into mortality using post-diagnosis mortality schedules. Simulation results evaluate recovery of known scenario effects, and the empirical illustration uses lung-cancer data from Uruguay with an extension to nine smoking-related cancer sites.

Keywords: age–period–cohort models; Bayesian methods; cancer projection; smoking; epidemiological modelling.

JEL classification: C11; C53; I10; I18.

1 Background

Changes in smoking prevalence affect cancer mortality through a long, staged epidemiologic pathway: smoking behaviour shapes incidence risk, incident cases translate into deaths through post-diagnosis mortality, and the resulting burden unfolds over cohorts and calendar time. For policy-oriented projection, the relevant estimand is therefore not only the future path of cancer mortality, but the contrast in future burden generated by alternative smoking-prevalence trajectories. A statistical framework for this task must combine smoothing on the Lexis surface with an explicit transmission mechanism from exposure scenarios to incidence and mortality.

This paper addresses this problem by proposing a sequential Bayesian age–period–cohort (APC) framework that integrates three layers of the Lexis surface—smoking prevalence, cancer incidence, and cancer mortality—within a sequential uncertainty-propagating architecture. The framework is designed for settings where long time series of outcomes are available but longitudinal data on the disease process are scarce. Bayesian APC models recover smooth latent structure from noisy, partially observed data, while externally supplied epidemiological parameters govern the transmission across layers. Statistical smoothing and epidemiologic transmission therefore play distinct roles: the former stabilises estimation on fragmented Lexis surfaces, and the latter defines how smoking scenarios move through incidence and mortality.

APC models provide a natural language for representing long-run variation on the Lexis surface, separating life-cycle patterns, period-specific disturbances, and generational differences. Bayesian APC formulations regularise these latent temporal components and provide structured uncertainty quantification, making them especially attractive for projection problems with sparse cells, long horizons, and partially observed surfaces (Clayton and Schifflers, 1987a,b; Carstensen, 2007; Knorr-Held, 2000; Blangiardo et al., 2013; Riebler and Held, 2017; Gascoigne et al., 2025). In cancer forecasting, APC-based approaches and related extrapolation tools remain central because they are interpretable, robust enough for routine registry use, and operationally feasible (Møller et al., 2002, 2003; Riebler and Held, 2017). For lung cancer, many projection studies model incidence or mortality directly from historical outcome series, while smoking is either absent from the statistical forecast or introduced through reduced-form exposure terms, lag structures, or cohort smoking summaries (Bashir and Estève, 2001; Brown and Kessler, 1988; Shibuya et al., 2005; Holford et al., 2014; Moolgavkar et al., 2012; Smith et al., 2021; Riebler and Held, 2017). A systematic review found that most lung-cancer projection studies did not incorporate smoking explicitly and that formal validation strategies were far from universal (Yu et al., 2019).

A related tradition uses hybrid epidemiologic models to connect observed outcomes with externally parameterised disease mechanisms. Microsimulation and comparative-

modelling efforts such as SimSmoke and CISNET represent smoking histories, natural-history pathways, and intervention scenarios, and have been influential in lung-cancer policy analysis (Levy et al., 2010; Holford et al., 2014; Meza et al., 2021). Incidence-to-mortality accounting tools such as MIAMOD similarly reflect a long-standing interest in linking epidemiological stages rather than projecting each outcome in isolation (De Angelis et al., 1994). Burden-of-disease exercises routinely combine observed or projected exposure with externally estimated relative risks to quantify future attributable mortality under alternative scenarios (Ezzati et al., 2004; GBD Risk Factors Collaborators, 2020; Luo et al., 2022). The framework proposed here adopts this hybrid orientation while keeping the statistical object aggregate: latent prevalence, incidence, and mortality surfaces are estimated from population data, and transmission across levels is parameterised through external epidemiological inputs.

The paper contributes a scenario-responsive projection framework for aggregate population data. The first contribution is methodological: smoking prevalence, cancer incidence, and cancer mortality are linked through a sequential Bayesian APC architecture in which statistical smoothing and epidemiological transmission are kept separate. The second contribution is evaluative: the framework is assessed through simulation with known scenario effects, making it possible to distinguish recovery of exposure-driven mortality contrasts from autonomous outcome forecasting. The third contribution is practical: the empirical illustration shows how the framework can be implemented with repeated risk-factor surveys, registry incidence data, mortality records, population projections, and externally supplied epidemiological inputs. Although the empirical illustration focuses on lung cancer, the framework is not site-specific. Extended to multiple smoking-related cancer sites, it can provide a structured basis for anticipating changes in the composition of cancer mortality under alternative exposure trajectories, with potential implications for health-system planning and resource allocation (Clements et al., 2005; Bray and Møller, 2006).

The paper proceeds as follows. Section 2 presents the modelling framework, external epidemiological inputs, simulation design, and empirical implementation. Section 3 reports the simulation validation and empirical illustration. Section 4 discusses contributions, limitations, and extensions, and Section 5 concludes.

2 Methods

This section formalises the sequential Bayesian age–period–cohort (SBAPC) framework, which links three Lexis-surface layers—smoking prevalence, lung-cancer incidence, and lung-cancer mortality—through external epidemiological inputs. Within each layer, Bayesian APC components recover smooth latent age, period, and cohort structure from noisy and partially observed data. Across layers, relative risks, risk reversion schedules, and post-diagnosis mortality schedules translate reconstructed smoking-state composition into incidence risk and

projected incidence into an expected-deaths surface. Bayesian APC models have previously been used to project lung-cancer mortality as a single surface (Knorr-Held and Rainer, 2001); here the APC structure is embedded in a sequential system that keeps the smoking channel explicit while allowing residual variation to be captured within each downstream layer. For notational clarity, the formal exposition is written for lung cancer, which is the main empirical illustration; the same structure applies to additional cancer sites once site- and sex-specific external inputs are supplied.

We index Lexis cells by age a, calendar year t, and birth cohort $c = t - a$, and estimate all models separately for males and females. The smoking-prevalence layer is fitted on ages 20–65, reflecting the support and quality of the pooled survey data. The incidence and mortality layers are fitted on ages 35–89, where lung-cancer counts are epidemiologically relevant and sufficiently informative for APC smoothing. Figure 1 provides a schematic overview of the architecture and of the external transmission components linking the three layers.

2.1 Smoking prevalence and smoking-state reconstruction

Let $Y^P_{a,t,g}$ denote the number of current smokers in age a, year t, and survey-instrument group g, and let $N^P_{a,t,g}$ denote the corresponding effective sample size. We assume

$$Y^P_{a,t,g} \mid \pi_{a,t,g} \sim \text{Binomial}\Big(N^P_{a,t,g}, \pi_{a,t,g}\Big), \tag{1}$$

where effective sample sizes are constructed at the cell level so that weighted survey information can be represented within a binomial mean–variance structure. Details of this construction are reported in Appendix A.1.

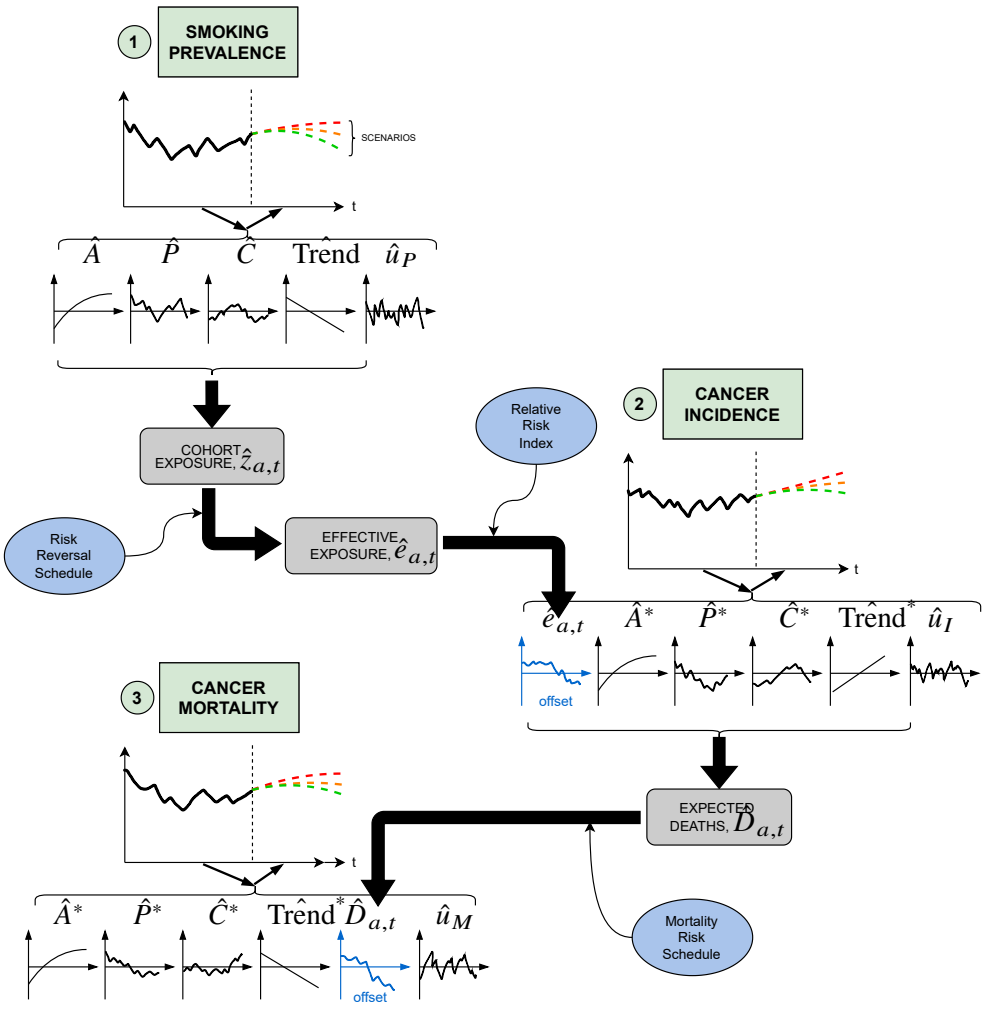
The smoking-prevalence surface is modelled as

$$\text{logit}\big(\pi_{a,t,g}\big) = \delta_g + \alpha^P_a + \beta^P_t + \gamma^P_c + \tau^P_t, \qquad c = t - a, \tag{2}$$

where δ_g captures systematic differences across survey instruments, and α^P_a, β^P_t, and γ^P_c denote smooth age, period, and cohort effects. The term τ^P_t is a low-frequency deterministic trend component, distinct from the smooth APC period effect. In the smoking layer, this trend is intended to absorb broad secular movement in smoking behaviour that may remain after accounting for survey harmonisation and structured age, period, and cohort variation.

The first layer produces an instrument-adjusted surface of current smoking prevalence. Over the historical period, this surface is learned from the pooled survey pseudo-counts. Over the projection period, it is extended forward under explicit smoking-prevalence scenarios. The four scenarios considered in the paper are prevalence frozen at its 2022 level, a one-percent annual increase, a one-percent annual decrease, and a quit scenario in which current

Figure 1: Sequential architecture of the proposed framework.



Notes: Rectangles denote APC model layers. Solid arrows denote sequential propagation from smoking prevalence to incidence and mortality through externally parameterised transmission inputs. Small embedded plots are schematic and do not correspond to actual data or model outputs. *Source:* Own elaboration.

smoking prevalence is set to zero from the start of the projection period.

For each projected scenario, the prevalence surface is converted into a cohort-age-period decomposition of current smokers, former smokers by years since cessation, and never smokers. This smoking-state reconstruction provides the epidemiological backbone through which future smoking scenarios affect cancer outcomes.

2.2 From smoking states to lung-cancer incidence

Lung-cancer incidence depends on accumulated exposure, so the projected prevalence surface is first converted into cohort-age-period smoking-state composition. A site-specific risk reversion schedule is applied to the former-smoker stock to construct an effective exposure term, denoted $e_{a,t}$, which measures the smoking-related excess risk remaining active in each age–period cell. Details are reported in Appendix A.2.

The corresponding incidence offset is

$$o^I_{a,t} = \log\left[1 + e_{a,t}(\mathrm{RR}_I - 1)\right],$$

where RR_I is an externally supplied relative-risk parameter. This offset translates reconstructed smoking-state composition into the incidence predictor.

Let $Y^I_{a,t}$ denote incident lung-cancer cases and let $E^I_{a,t}$ denote the corresponding population at risk. We assume

$$Y^I_{a,t} \mid \lambda^I_{a,t} \sim \mathrm{Poisson}\left(E^I_{a,t}\lambda^I_{a,t}\right). \tag{3}$$

The incidence surface is modelled as

$$\log \lambda^I_{a,t} = \mu^I + \alpha^I_a + \beta^I_t + \gamma^I_c + \tau^I_t + o^I_{a,t}, \qquad c = t - a, \tag{4}$$

where μ^I is an intercept, α^I_a, β^I_t, and γ^I_c are smooth APC effects, τ^I_t is a linear trend in centred calendar period, and $o^I_{a,t}$ enters as a fixed epidemiological offset. The absence of an estimated coefficient on $o^I_{a,t}$ reflects its role as an externally parameterised transmission input rather than as a covariate whose association is learned from the incidence data.

Over the historical period, the incidence model is fitted to registry counts conditional on the smoking-related offset implied by the estimated prevalence layer. The APC effects and trend therefore capture residual incidence structure after accounting for the smoking-transmission signal. This residual component may reflect other determinants of lung-cancer incidence, such as occupational exposures, air pollution, diagnostic intensity, registry ascertainment, unmodelled smoking-intensity patterns, or mismatch between the external relative-risk input and the target population. Over the projection period, each smoking-prevalence scenario produces a scenario-specific effective exposure surface and hence a scenario-specific

incidence offset; projected incidence combines that offset with extrapolated residual APC structure.

2.3 From projected incidence to expected-deaths surfaces and mortality

Projected incidence is translated into mortality through an external post-diagnosis mortality risk schedule. Let $\hat{I}_{a,t}$ denote projected incident cases from the incidence layer. We represent post-diagnosis mortality by a discrete schedule

$$\{w_0, w_1, \ldots, w_L\}, \qquad w_\ell \geq 0, \qquad \sum_{\ell=0}^{L} w_\ell \leq 1,$$

where w_ℓ is the share of diagnosis-related mortality assigned to year ℓ after diagnosis. In the empirical implementation, the interval mortality probabilities are annualised using a midyear convention: deaths occurring during the first post-diagnosis year are split across the diagnosis calendar year and the following calendar year. The exact annualisation and alignment rules are reported in Appendix A.3.

The expected-deaths surface is obtained by discrete convolution of projected incident cases with the annualised mortality schedule:

$$D_{a,t} = \sum_{\ell=0}^{L} w_\ell \hat{I}_{a-\ell,t-\ell}, \tag{5}$$

up to the alignment conventions imposed by the Lexis grid. The quantity $D_{a,t}$ is the mortality surface implied on epidemiological grounds by the projected incidence trajectory and the external post-diagnosis mortality schedule.

Let $Y^M_{a,t}$ denote observed lung-cancer deaths and let $E^M_{a,t}$ denote the corresponding population exposure. We assume

$$Y^M_{a,t} \mid \lambda^M_{a,t} \sim \text{Poisson}\left(E^M_{a,t}\lambda^M_{a,t}\right). \tag{6}$$

The mortality surface is modelled as

$$\log \lambda^M_{a,t} = \mu^M + \alpha^M_a + \beta^M_t + \gamma^M_c + \tau^M_t + \log D_{a,t}, \qquad c = t - a, \tag{7}$$

where μ^M is an intercept, α^M_a, β^M_t, and γ^M_c are smooth APC effects, τ^M_t is an explicit linear trend in centred calendar period, and $\log D_{a,t}$ enters as a fixed offset. The absence of an estimated coefficient on $\log D_{a,t}$ reflects its role as an externally parameterised incidence-to-mortality input.

Over the historical period, the mortality layer is fitted to observed death counts

conditional on the expected-deaths surface derived from the incidence layer. Over the projection period, each smoking scenario produces a projected incidence surface, which is mapped into a scenario-specific expected-deaths offset before mortality is projected. The residual APC structure captures variation around the incidence-to-death translation, including non-smoking determinants of mortality, cause-of-death coding practices, departures from the external post-diagnosis schedule, and long-run changes in lethality, treatment, or stage at diagnosis. The trend τ_t^M is retained as a parsimonious residual secular component once the expected-deaths offset has been imposed.

2.4 Bayesian estimation, uncertainty propagation, and horizon diagnostics

All three layers are estimated as latent Gaussian structured additive models within a Bayesian APC framework. Smooth age, period, cohort, and trend components are assigned regularising priors, and posterior inference is performed using integrated nested Laplace approximations (INLA), as implemented in the R-INLA framework (Rue et al., 2009; Martins et al., 2013). The approach provides stable smoothing on relatively high-dimensional Lexis surfaces, yields posterior approximations for latent components, and remains computationally tractable under repeated scenario evaluation and simulation-based validation. Additional implementation details on prior specification, posterior predictive simulation, and horizon diagnostics are reported in Appendices A.4 to A.6.

Estimation proceeds in three fitted stages. In the historical period, the prevalence layer is fitted to survey pseudo-counts, the incidence layer is fitted to registry counts conditional on the smoking-related incidence offset, and the mortality layer is fitted to death counts conditional on the expected-deaths offset derived from incidence. In the projection period, the fitted prevalence surface is extended under smoking scenarios, transformed into scenario-specific effective exposure and incidence offsets, and then propagated through projected incidence into scenario-specific expected-deaths offsets for mortality.

Future cohort trajectories are extrapolated using a common deterministic rule applied across the three layers. Specifically, cohort effects beyond the observed Lexis support are projected using a damped linear trend estimated over a fixed window of recent cohorts. Let γ_c denote the cohort effect; for cohorts beyond the observed range, the forecast is constructed as a continuation of the recent linear trend, attenuated by a damping factor to avoid explosive extrapolation. This rule is applied consistently in the prevalence, incidence, and mortality layers, although its contribution to the predictor differs across layers depending on the role of the cohort component within each APC specification. In the prevalence layer, it governs the continuation of cohort smoking trajectories; in the incidence layer, it affects the residual APC component around the epidemiological offset; and in the mortality layer, it operates on

the residual APC component around the expected-deaths surface. This design preserves a common extrapolation principle while allowing layer-specific interpretation of cohort effects.

Uncertainty is propagated across stages in a sequential manner. Each layer produces posterior summaries for its latent surface, which are then used to construct the inputs of the subsequent layer. In particular, uncertainty in the smoking-prevalence surface affects the reconstructed smoking states and the effective exposure term entering incidence; uncertainty in projected incidence affects the expected-deaths surface through the post-diagnosis mortality schedule; and the mortality layer is then estimated conditional on that constructed surface. As a result, predictive uncertainty in downstream quantities incorporates within-layer estimation uncertainty and the main upstream sources transmitted through the sequential transformations. We emphasise that this sequential architecture is formally designated as a structured projection decomposition rather than a jointly identified causal framework, a necessary and computationally tractable choice given the complex non-linear smoking reconstruction on aggregate data.

Long-horizon projections on the Lexis surface are increasingly shaped by boundary extrapolation and declining empirical support. The framework classifies future periods using endogenous horizon diagnostics derived from the geometry of the observed Lexis support. For each projected year, we evaluate how much of the underlying age–period–cohort structure remains anchored in historically observed information and how much depends on edge extrapolation. This yields a reproducible classification of projected periods into credible, caution, and risky regions. The maximum reported projection horizon is defined as the last projected year that remains within these endogenous regions; subsequent years are treated as beyond the maximum support horizon. These categories are reporting devices, not stochastic states of the model.

The framework is evaluated using both empirical fit diagnostics and simulation-based scenario recovery. Fit and historical predictive performance are assessed with standard model comparison tools and backtesting summaries. The main validation target, however, is the recovery of scenario effects in controlled simulations. This distinction is central: SBAPC is designed for coherent scenario-responsive projection, where upstream smoking scenarios propagate through epidemiological links into incidence and mortality, rather than for unconstrained outcome extrapolation alone.

2.5 External epidemiological inputs

The empirical illustration combines Uruguayan survey, registry, and mortality data with external epidemiological inputs that are not available from the target population at the same level of detail. This is a common difficulty in applied risk-factor projection: population data may support estimation of prevalence, incidence, and mortality surfaces, while the

epidemiological quantities needed to transmit exposure histories across disease stages must often be transported from external evidence. In the present framework, relative risks, risk reversion schedules, and post-diagnosis mortality schedules are therefore treated as explicit inputs to the projection exercise. Their provenance, scale, and interpretation are documented before the empirical application because they determine how smoking scenarios are translated into downstream incidence and mortality.

The incidence layer uses site- and sex-specific relative-risk parameters. In the present application, these quantities are treated as fixed external inputs and enter the predictor through the incidence offset in Section 2.2. Their role is to determine the magnitude of the smoking-related incidence contrast after the prevalence-derived exposure signal has been translated into effective exposure. The values used here coincide with the IECS/Pichon-Riviere relative-risk table for Mexico and with the corresponding supplementary table reported in the associated multicountry smoking-burden model family (Pichon-Riviere et al., 2020, 2013). Because related publications in that model family label analogous tables differently, we treat these quantities as external condition-specific relative-risk parameters by sex and smoking status, with never smokers as the reference category. The full table is reported in Appendix B.1.

Risk reversion after cessation is represented by site-specific schedules for the proportion of smoking-related excess risk remaining after a given number of years since quitting. The fitted prevalence surface is converted into smoking states for current, former, and never smokers, and the former-smoker stock is weighted by the relevant schedule to construct the effective exposure term. The schedules were assembled from the published literature and converted into a common excess-risk scale. Their evidentiary basis varies across cancer sites, with stronger support for some sites and more approximate or proxy constructions for others. Source-by-source provenance and construction details are reported in Appendix B.2.

The incidence-to-mortality link uses post-diagnosis mortality schedules derived from published net-survival information. For the broader multisite implementation, these schedules are constructed from REDECAN adult cancer survival estimates for Spain, 2013–2017, and converted into interval probabilities for death after diagnosis. Their homogeneous coverage by site and sex makes them useful for implementation across tumour sites, while their use in Uruguay should be interpreted as transport of external survival evidence rather than estimation from local follow-up data. The derivation, timing convention, and corresponding table are reported in Appendix B.3.

The main specification conditions on these external inputs as fixed quantities. This separates uncertainty in the estimated Lexis-surface components from uncertainty in transported epidemiological parameters. Consequently, posterior intervals reflect model-based uncertainty in the sequential APC projection, but do not represent the full uncertainty associated with relative-risk transport, risk-reversion functions, or post-diagnosis mortality schedules. To assess sensitivity to this specification, Appendix D.9 reports a deterministic

one-at-a-time sensitivity analysis for the Uruguayan lung-cancer application. The analysis perturbs the incidence relative risks, the risk reversion schedule, and the post-diagnosis mortality timing, and compares the resulting cumulative mortality projections with the baseline implementation.

2.6 Simulation design

The simulation study evaluates scenario-effect recovery under controlled conditions. The validation target is the mortality response induced by counterfactual smoking-prevalence trajectories and transmitted through the incidence and mortality links. The simulation also quantifies how much of the remaining discrepancy from the simulated truth is associated with the realistic observation window used by the estimator.

The simulation design follows the same epidemiological sequence as the proposed estimator. For each Monte Carlo seed, a latent smoking-prevalence surface is generated from a smooth APC structure. That surface is converted into cohort-age-period smoking states, transformed into an effective exposure term, linked to lung-cancer incidence through the smoking-related incidence offset, and mapped into lung-cancer mortality through the post-diagnosis mortality schedule. Prevalence, incidence, and mortality observations are then generated from the corresponding likelihoods on the Lexis grid.

The well-specified design aligns the data-generating mechanism with the transmission rule imposed by SBAPC. For evaluation, the true pathway is known on the full latent support used to generate smoking-state composition and downstream outcomes. The realistic estimator, however, is fitted using empirical-style observation windows: smoking prevalence is observed over ages 20–65, whereas incidence and mortality are observed over ages 35–89 up to 2022. Simulated projections are generated after 2022 and summarised by endogenous horizon regions rather than by a single fixed full-horizon aggregate. This separation between the full latent truth and the restricted observed window allows the simulation to evaluate both the sequential transmission logic and the support limitations faced in the empirical application.

The final simulation study uses 200 Monte Carlo seeds. The uncertainty bands reported in the results summarise empirical variation across independently generated simulation replications, not posterior uncertainty conditional on a single dataset. In each seed, four future smoking-prevalence paths are imposed after 2022: prevalence frozen at its 2022 level, a one-percent annual increase, a one-percent annual decrease, and a quit scenario that sets current smoking prevalence to zero from the start of the projection period. The frozen-prevalence path is the reference scenario. Let $B_s^Y(t)$ denote the aggregate burden in year t for outcome $Y \in \{I, M\}$ under scenario s, where I denotes incident cases and M denotes deaths.

The annual scenario effect is

$$\Delta_s^Y(t) = B_s^Y(t) - B_{\text{freeze}}^Y(t).$$

The main validation target is $\Delta_s^M(t)$, the annual mortality effect of each smoking scenario relative to frozen prevalence. Incidence effects are used as intermediate diagnostics and are reported in Appendix C.

The main estimator is SBAPC fitted under the realistic observed-window information set. The benchmark is an autonomous Bayesian APC (BAPC) projection fitted to the outcome series without the smoking-prevalence transmission channel. The benchmark is not expected to recover smoking-scenario effects, because it is deliberately scenario-blind. Its role is to clarify the estimand: autonomous outcome forecasting and scenario-responsive projection answer different methodological questions.

2.7 Empirical illustration data and implementation

The empirical application is used to illustrate implementation, interpretation, and diagnostic outputs under realistic data constraints; it is not intended as a standalone national burden assessment. We use Uruguayan data to report lung-cancer mortality projections by sex under the smoking-prevalence scenarios used throughout the paper, and then extend the implementation to nine smoking-related cancer sites to study the future composition of a wider cancer burden. Additional data documentation, input diagnostics, extended projection displays, benchmark comparisons, and fit summaries are reported in Appendix D.

The empirical application uses Uruguay as a case study because the country combines rich national data with a well-documented tobacco-control policy environment (see Appendix D.1). Smoking prevalence is estimated from harmonised repeated cross-sectional survey microdata covering 1998–2024. Lung-cancer incidence is obtained from National Cancer Registry tabulations for 1998–2022, reconstructed to the single-age, single-year Lexis grid. Lung-cancer mortality is obtained from official vital-statistics tabulations and smoothed to the same age-period structure. Population denominators are taken from official demographic estimates and projections, so projected death counts incorporate the effect of population ageing as well as changes in age-specific risks. The smoking-prevalence layer is estimated over ages 20–65, while incidence and mortality are estimated over ages 35–89, following the support restrictions described in Section 2 and Appendices D.2 to D.5.

Appendix D.6 provides descriptive input series for smoking prevalence, lung-cancer incidence, and lung-cancer mortality. These series motivate a sex-specific sequential model: smoking prevalence declined over the observed period, especially among males, while incidence and mortality show distinct profiles by sex. Because projections are reported as annual deaths, they combine age-specific risks, smoking-scenario effects, and population ageing.

The empirical projections begin after 2022. The frozen-prevalence scenario keeps smoking prevalence at its 2022 level; the one-percent increase and decrease scenarios apply proportional annual changes to that path; and the quit scenario sets current smoking prevalence to zero from the start of the projection period, treating all smokers present in 2022 as former smokers thereafter. For illustration, a male prevalence of about 22% in 2022 corresponds by 2048 to approximately 22% under freezing, 28.5% under a one-percent annual increase, and 17% under a one-percent annual decrease. Numerical summaries in the main text are restricted to the common endogenous horizon ending in 2048, with additional diagnostics in Appendices D.7 and D.8.

2.8 Use of AI-assisted tools

During manuscript preparation and code development, the author used ChatGPT and Codex to support structural drafting, language editing, programming assistance, and code review. All model specifications, code changes, numerical outputs, interpretations, and final text were reviewed and verified by the author, who takes full responsibility for the content of the manuscript.

3 Results

This section reports two components of the evaluation. First, the simulation study assesses whether the proposed framework recovers mortality contrasts generated by known smoking-scenario effects. Second, the Uruguayan application illustrates implementation and interpretation under realistic population-data constraints.

3.1 Simulation validation

The simulation results evaluate whether SBAPC recovers mortality effects generated by counterfactual smoking paths when the data-generating mechanism is known. The reporting focuses on two features: the role of endogenous support windows and the recovery of annual and cumulative mortality scenario effects. Extended simulation documentation, intermediate pathway diagnostics, sex-specific results, calibration checks, and misspecification analyses are reported in Appendix C.

Applying the endogenous horizon diagnostic to the simulation outputs yields common transition years for the aggregate mortality plots. The credible region runs through 2038, the caution region begins in 2039, and the risky region begins in 2048. These regions are used as reporting devices and as visual overlays in the main scenario-effect figure.

The simulation design also separates support-window effects from the rest of the estimation problem. The truth is generated on the full latent support required to construct

smoking histories and downstream outcomes. The realistic SBAPC estimator observes the empirical-style window, with smoking prevalence over ages 20–65 and incidence and mortality over ages 35–89. A full-support SBAPC diagnostic gives the estimator access to the broader support used to generate the truth, allowing us to separate discrepancies associated with support truncation from those associated with the sequential transmission structure.

A second source of divergence is intrinsic to forecasting. The simulated truth contains the future trajectory generated by the data-generating process, whereas the estimator conditions only on the historical window ending in 2022 and extrapolates forward from that information set. Differences between estimated and true scenario effects may therefore increase with horizon both because Lexis support weakens and because the estimator cannot condition on the realised future evolution of the latent APC components.

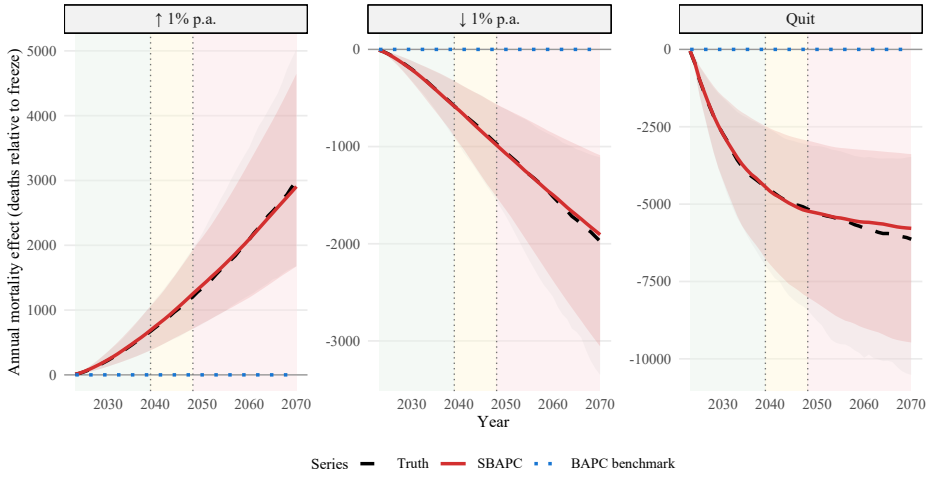
Pathway-level diagnostics support the same interpretation and are reported in Appendix C.5. In a representative simulation, the main observed-window discrepancy appears upstream, in the reconstruction of smoking-state composition and effective exposure beyond the directly observed prevalence window. The full-support diagnostic tracks the true exposure pathway more closely, while the observed-window estimator preserves the ordering and broad temporal shape of the scenario-specific exposure paths. Downstream incidence and mortality remain closely aligned across support variants because they are also anchored by simulated outcome data and by the epidemiological offsets.

Figure 2 reports the central simulation result. It shows annual mortality effects relative to frozen prevalence for the one-percent annual increase, one-percent annual decrease, and quit scenarios. Lines report the median across simulation seeds, and ribbons show the 10th–90th percentile range for Truth and SBAPC. The BAPC benchmark appears as a zero-effect reference because it has no smoking-scenario channel.

SBAPC recovers the sign, relative magnitude, and temporal profile of the mortality response under the well-specified design. The one-percent annual increase in smoking prevalence produces positive mortality effects relative to frozen prevalence, whereas the one-percent annual decrease and quit scenarios produce negative effects. The estimated trajectories are close to Truth in the credible region and remain well ordered in the caution and risky regions. Dispersion increases with the projection horizon, reflecting both weakening Lexis support and the growing distance between the estimator's historical information set and the future trajectory generated by the data-generating process. The scenario ranking nevertheless remains stable across the reported projection regions.

The quit scenario produces the largest absolute mortality reduction and the most nonlinear trajectory. This reflects the sequential timing of the mechanism: cessation changes the current-smoker stock first, effective exposure adjusts through the risk reversion schedule, incidence responds through the smoking offset, and mortality follows through the post-diagnosis mapping. The gradual increase and decrease scenarios produce smaller but clearly

Figure 2: Mortality scenario-effect recovery.



Notes: Scenario effects are annual deaths relative to the frozen-prevalence baseline, aggregated across sexes and simulation seeds. Lines show median effects across simulation replications; shaded bands summarise Monte Carlo variation across simulated datasets: they correspond to the empirical 10th and 90th percentiles of the seed-specific estimates and truths. The BAPC benchmark is scenario-blind and therefore has zero scenario response by construction. Background shading indicates the endogenous support regions. *Source:* Own elaboration.

signed effects, consistent with recovery of the downstream consequences of upstream smoking changes.

Table 1 summarises the same validation target in cumulative form. Within each horizon region, the cumulative recovery ratio is the estimated cumulative mortality effect divided by the true cumulative mortality effect. Annual mean absolute relative error is computed from the year-by-year differences between estimated and true scenario effects, scaled by the magnitude of the true annual effect.

The cumulative results reinforce the visual evidence from Figure 2. In the credible region, recovery ratios remain close to one across all three non-reference scenarios. Across the reported endogenous horizon regions, recovery ratios remain close to one for the one-percent increase, one-percent decrease, and quit scenarios. Annual relative errors rise as the projection enters weaker-support regions, but the cumulative magnitude and direction of the mortality effects remain well recovered.

The BAPC benchmark clarifies the role of the smoking channel. Its scenario effects are zero because future mortality is extrapolated autonomously from the outcome series. The comparison highlights the difference between baseline outcome forecasting and recovery of mortality contrasts generated by upstream smoking scenarios. The simulation results support interpreting SBAPC as a scenario-recovery framework centred on exposure-driven mortality contrasts.

Appendix C provides the full simulation documentation and extended diagnostics. Appendix C.5 reports the observed-window versus full-support comparison; Appendices

Table 1: Cumulative recovery of mortality scenario effects.

Scenario	Horizon	Truth	SBAPC	Recovery	MARE (%)
↑1%	Credible	4 371	4 509	1.02	7.1
	Caution	8 140	8 338	1.01	8.6
	Risky	46 807	46 878	0.99	12.4
↓1%	Credible	-3 894	-3 928	1.01	6.6
	Caution	-6 753	-6 874	0.99	8.3
	Risky	-33 781	-33 464	0.97	12.2
Quit	Credible	-42 489	-41 599	0.98	5.6
	Caution	-43 244	-43 531	0.97	8.2
	Risky	-129 743	-126 738	0.96	11.9

Notes: Values are cumulative deaths relative to the frozen-prevalence baseline, aggregated across sexes and simulation seeds. Recovery is the ratio of the estimated cumulative effect to the true cumulative effect within each endogenous horizon region. MARE is the annual mean absolute relative error. *Source:* Own elaboration.

C.6 to C.9 report sex-specific recovery, distributional diagnostics, misspecification results, and seed-level illustrations. Under misspecification, the estimated effects retain the correct sign and scenario ordering, while magnitude recovery deteriorates, especially for the gradual increase and decrease scenarios.

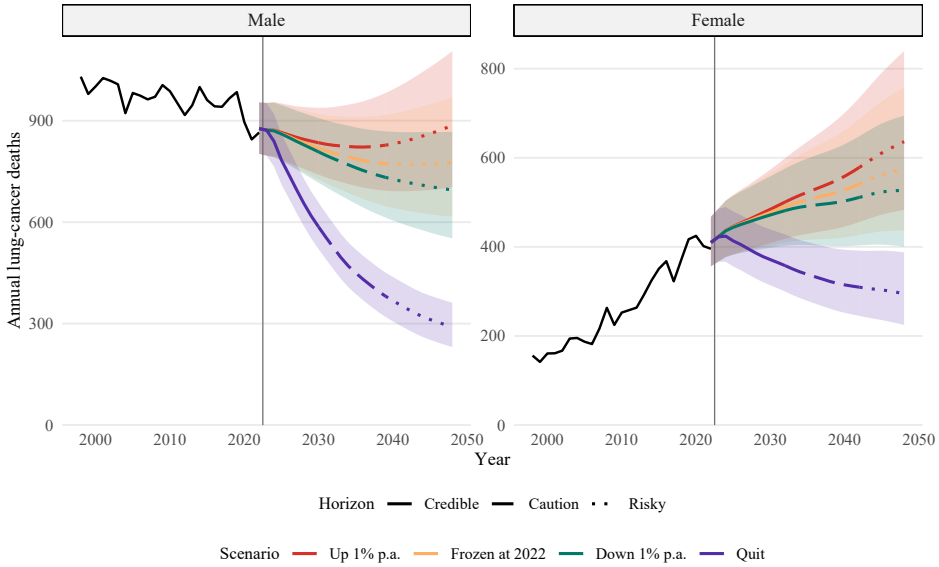
3.2 Uruguayan lung-cancer illustration

Figure 3 reports the main lung-cancer mortality projections. The figure shows observed annual deaths through 2022 and SBAPC projections under the four smoking-prevalence scenarios, separately for males and females. Projections are shown on the scale of annual deaths rather than rates because this is the scale most directly related to downstream burden and resource planning.

Because the projections are reported as annual deaths, the paths in Figure 3 should be read as the combined result of age-specific cancer risks, smoking-scenario effects, and population ageing. The projected paths differ sharply by sex. Among males, lung-cancer mortality declines under the frozen-prevalence scenario, from about 865 deaths in 2022 to 776 deaths in 2048. Continued increases in smoking prevalence reverse part of that decline, while the quit scenario produces a much steeper reduction, reaching roughly 290 annual deaths by 2048. Among females, the frozen-prevalence scenario projects rising lung-cancer mortality, from about 396 deaths in 2022 to 575 deaths in 2048. The quit scenario changes that trajectory substantially, leaving projected annual deaths close to 295 by 2048. These sex-specific patterns reflect differences in historical smoking trajectories, baseline mortality levels, relative-risk inputs, and the timing with which changes in smoking prevalence are transmitted through incidence and mortality.

Table 2 gives the corresponding annual death counts for selected years. By 2048, the

Figure 3: Projected lung-cancer mortality in Uruguay by sex.



Notes: Black lines show observed annual lung-cancer deaths through 2022. Colored lines and ribbons show SBAPC projected annual deaths and interval summaries under smoking-prevalence scenarios, separately for males and females. Line type indicates the endogenous horizon region; projections are clipped to the common endogenous horizon. Shaded bands show approximate 95% credible intervals for expected annual deaths, obtained from INLA marginal summaries and aggregated across age cells. *Source:* Own elaboration.

difference between the frozen-prevalence and quit scenarios is about 486 annual lung-cancer deaths for males and 280 for females. The gradual one-percent scenarios produce smaller changes over the same horizon: higher smoking prevalence increases projected deaths relative to the frozen-prevalence path, while lower prevalence reduces them. The empirical exercise therefore illustrates the main practical function of the framework: translating upstream smoking scenarios into sex-specific mortality trajectories on an interpretable burden scale, while retaining the demographic pressure embedded in the projected population denominators.

3.3 Extension to smoking-related cancer sites

To illustrate the extensibility of the framework, we also apply the same projection logic to nine cancer sites with established evidence of a causal relationship with tobacco smoking: lung, stomach, pancreas, bladder, oral cavity and pharynx, esophagus, kidney, larynx, and cervix (International Agency for Research on Cancer, 2012; National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health, 2014). The broader implementation contains site-specific external relative risks, risk reversion schedules, and post-diagnosis mortality inputs for each site and sex. The purpose of this exercise is methodological: it shows how the framework can be used to examine not only projected mortality levels, but also the composition of future burden across outcomes once the required external inputs are organised.

Table 2: Projected annual lung cancer deaths in selected years.

Sex	Scenario	2022	2030	2040	2048
Male	Up 1% p.a.	865	835	831	886
	Frozen at 2022	865	819	772	776
	Down 1% p.a.	865	808	727	695
	Quit	865	586	368	290
Female	Up 1% p.a.	396	484	559	636
	Frozen at 2022	396	477	528	575
	Down 1% p.a.	396	472	503	527
	Quit	396	371	315	295

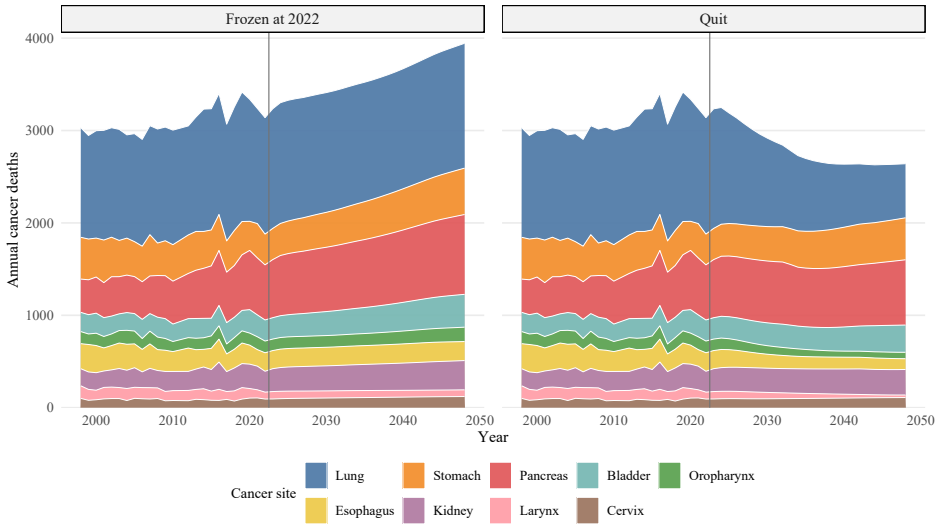
Notes: Values are annual deaths by sex and smoking-prevalence scenario. The 2022 column is the observed historical endpoint; future columns are SBAPC projections restricted to the common endogenous horizon.
Source: Own elaboration.

Figure 4 illustrates this extension. It reports projected annual mortality, summed over sexes, for nine smoking-related cancer sites under the frozen-prevalence and quit scenarios. The figure is clipped at 2048, the common endogenous projection year across the included sites. The purpose is to show how the framework can support composition-sensitive burden projections, not only single-site trajectories.

The multisite projection adds a policy-relevant composition margin that is not available from a lung-only display. Under frozen prevalence, total mortality across the nine sites rises from about 3 140 annual deaths in 2022 to 3 944 in 2048. Lung cancer remains the largest component in 2048, accounting for roughly one third of the total, followed by pancreas, stomach, bladder, and kidney. Under the quit scenario, total projected mortality in 2048 falls to about 2 640 deaths. The composition also changes: lung cancer falls to about 22% of the total, while pancreas becomes the largest component. Because the figure reports deaths from smoking-related sites rather than population-attributable fractions, the interpretation is compositional: scenario-responsive projections can change both the level and the site distribution of the future cancer burden.

This extension illustrates a broader use of the method. Once external epidemiological inputs are organised by site and sex, the framework can project how alternative smoking trajectories affect the distribution of future mortality across tumour sites. The relevant methodological point is that scenario responsiveness can be evaluated not only for a single outcome trajectory, but also for the composition of a multicomponent burden.

Figure 4: Projected mortality across smoking-related cancer sites.



Notes: Stacked areas report annual deaths summed over sexes for nine smoking-related cancer sites under the frozen-prevalence and quit scenarios. Site colors are constant across panels; projections are clipped to the common endogenous horizon across included sites. *Source:* Own elaboration.

4 Discussion

This paper develops a sequential Bayesian APC framework for scenario-responsive cancer projections using aggregate population data. The framework links three Lexis-surface layers—smoking prevalence, cancer incidence, and cancer mortality—through externally parameterised epidemiological transmission. Bayesian APC components recover smooth latent structure from fragmented survey, registry, and mortality data, while relative risks, risk reversion schedules, and post-diagnosis mortality schedules define how information moves across stages. The result is a modular projection architecture that remains feasible when individual longitudinal disease-process data are unavailable, while keeping the exposure channel explicit and auditable.

The simulation study provides the main methodological validation. Under the well-specified design, SBAPC recovers the sign, ordering, temporal profile, and cumulative magnitude of mortality scenario effects relative to frozen prevalence. Recovery is strongest in the credible horizon and becomes less precise as Lexis support weakens, but the qualitative ranking of scenarios remains stable. The BAPC benchmark clarifies the estimand: autonomous APC extrapolation is informative for baseline forecasting, whereas SBAPC targets mortality contrasts generated by counterfactual smoking trajectories. Support-window diagnostics show that the restricted observation window accounts for part of the remaining discrepancy from the simulated truth, and the misspecified-transmission exercise shows that qualitative recovery can remain stable even when quantitative magnitude recovery deteriorates under an incorrect smoking-to-incidence link.

The Uruguayan application illustrates implementation under realistic data constraints rather than serving as a standalone national burden assessment. The lung-cancer results show how the framework translates smoking scenarios into sex-specific mortality trajectories, while the multisite extension shows that the same architecture can be used to study changes in the composition of projected mortality across outcomes. These empirical results are therefore best read as demonstrations of feasibility, interpretation, and diagnostic output in a real population-data setting.

These results support a scenario-oriented interpretation of SBAPC. Autonomous APC projections remain useful for extrapolating historical outcome series, but policy-oriented questions often require projections that respond to explicit exposure trajectories. SBAPC addresses that task by combining APC smoothing with epidemiological transmission, so that scenario contrasts are generated upstream and propagated through incidence and mortality rather than inferred from outcome trends alone. The endogenous horizon diagnostic complements this interpretation by showing where projected values remain close to observed Lexis support and where they rely increasingly on edge extrapolation. This is particularly important in a sequential framework, where the support for smoking histories, incidence, and mortality may deteriorate at different rates.

Several limitations are important for interpretation. First, the scope of the paper is narrower than the full health burden of tobacco. The main empirical application focuses on lung cancer, and the multisite extension covers smoking-related cancer sites for which the broader implementation contains external inputs. Tobacco-related harm also includes other neoplasms and, importantly, cardiovascular and respiratory outcomes. The present framework is therefore a modular architecture for a defined part of the tobacco-related disease process.

Second, the epidemiological links are externally parameterised and treated as fixed in the main specification. This modelling choice separates uncertainty in the estimated population surfaces from uncertainty in the epidemiological parameters used to transmit exposure histories across disease stages. It also makes the relative risks, risk-reversion schedules, and post-diagnosis mortality schedules explicit components of the projection design. However, it shifts part of the inferential burden to the transportability of these inputs and does not fully propagate epidemiological-parameter uncertainty. The deterministic sensitivity analysis for Uruguay provides a first assessment of this dependence, although it does not replace a probabilistic treatment of uncertainty in the external inputs. This limitation is especially relevant for post-diagnosis mortality schedules, which are derived from external net-survival evidence and are not partitioned by smoking status at diagnosis. More generally, relative risks, risk reversal, and post-diagnosis mortality may vary with smoking intensity, product composition, screening, treatment, and cohort context.

Third, the framework is modular rather than jointly hierarchical. A joint hierarchical formulation would be a natural alternative and may be preferable in settings with

stronger individual-level or disease-process data. We chose a modular architecture because the available evidence differs sharply across layers and because modular estimation prevents downstream outcome irregularities from feeding back into the exposure surface. This improves transparency and robustness when the smoking-prevalence layer is based on repeated survey data, whereas the incidence and mortality layers are based on registry and vital-statistics counts. The cost is that not all cross-layer posterior dependence is represented in a single joint likelihood. The separation between smoking-related transmission and residual APC structure should therefore be read as a structured projection decomposition, not as point identification of a causal pathway.

A final empirical limitation concerns cause-of-death classification. In Uruguay, the residual category for symptoms, signs, and abnormal clinical and laboratory findings increased over the study period. To the extent that this residual category absorbs deaths that would otherwise have been assigned to specific causes, observed cause-specific mortality trends may partly reflect changes in coding completeness. This concern affects the interpretation of historical mortality trends and therefore of projections anchored to those trends.

These limitations point to methodological extensions. The most direct is to replace proxy transmission inputs with population-specific or regionally pooled evidence wherever possible, including Latin American estimates of site-specific relative risks, cessation-related risk reversal, and post-diagnosis mortality. A second extension is to treat external inputs probabilistically, by averaging over plausible schedules or assigning prior distributions to relative risks, risk reversion functions, and mortality kernels. A third extension is to apply the same architecture to additional smoking-related cancers and, where defensible transmission inputs exist, to cardiovascular or respiratory outcomes. The broader promise of the approach is as a reusable template for linking risk-factor dynamics to downstream health outcomes under explicit epidemiological assumptions.

5 Conclusions

The proposed sequential Bayesian age–period–cohort framework provides a reproducible way to link smoking-prevalence scenarios to downstream cancer incidence and mortality projections using aggregate population data. By separating statistical smoothing within each Lexis-surface layer from externally parameterised epidemiological transmission between layers, the framework makes scenario assumptions explicit and allows exposure-driven contrasts to be evaluated directly. Simulation results support its use for recovering the direction, timing, and cumulative magnitude of mortality scenario effects under correct specification, while support-window and misspecification diagnostics clarify where quantitative interpretation weakens. The framework is therefore best interpreted as a structured scenario-projection tool rather than a jointly identified causal model. Its practical value is greatest for researchers

and health agencies that need to evaluate exposure-scenario implications using repeated risk-factor surveys, registry incidence data, mortality records, and population projections, but do not have individual longitudinal data linking exposure histories to disease outcomes.

List of abbreviations

APC: age–period–cohort; BAPC: Bayesian age–period–cohort; INLA: integrated nested Laplace approximations; MARE: mean absolute relative error; RRI: relative-risk index; SBAPC: sequential Bayesian age–period–cohort.

Declarations

Ethics approval and consent to participate

This study used de-identified secondary survey microdata and aggregate cancer incidence, mortality, and population tabulations. No participants were recruited, no individual clinical records were linked, and no directly identifiable information was accessed. The Ethics Committee of the Faculty of Social Sciences, Universidad de la República, determined that the study did not require full ethics review and that individual informed consent was not applicable. Reference: [REFERENCE NUMBER], date [DATE].

Consent for publication

Not applicable.

Availability of data and materials

The replication materials supporting this article are available from GitHub at `https://github.com/manuelfloresg/sbapc-smoking-lung-mortality` and archived on Zenodo at `https://doi.org/10.5281/zenodo.21199445` (Flores, 2026). The archived release corresponds to version `v1.0.3-submission` and contains the code, article-specific configuration files, redistributable analysis-ready inputs, simulation outputs, and scripts required to reproduce the tables and figures. Project name: sequential Bayesian APC smoking-cancer projections. Project home page: `https://github.com/manuelfloresg/sbapc-smoking-lung-mortality`. Source-code release: `https://github.com/manuelfloresg/sbapc-smoking-lung-mortality/releases/tag/v1.0.3-submission`. Archived version: `https://doi.org/10.5281/zenodo.21199445`. Operating system: platform independent. Programming language: R.

Other requirements and package versions are documented in the repository README. License: MIT License for code; CC BY 4.0 for redistributable analysis-ready data. Restrictions for non-academic use: none. Raw survey microdata and institutional source files supplied by Uruguay's Ministry of Public Health and National Cancer Registry are not publicly redistributed because access is subject to institutional data-use conditions. The repository documents the provenance, redistribution status, and construction of all public analysis-ready inputs.

Competing interests

The author declares that he has no competing interests.

Funding

This work was supported by Bloomberg Philanthropies through the Universidad Adolfo Ibáñez, as part of the Bloomberg Initiative to Reduce Tobacco Use. The Department of Economics at the Universidad de la República received funding to conduct economic research on tobacco in Uruguay. The funder had no role in the conceptualisation, design, data analysis, interpretation of results, preparation of the manuscript, or decision to submit the manuscript for publication.

Authors' contributions

MF conceptualised the study, developed the methodology and software, conducted the formal analysis and simulation study, prepared the empirical application, produced the visualisations, drafted the manuscript, revised the manuscript, and approved the final version.

Acknowledgements

I thank Zuleika Ferre and Mariana Gerstenblüth for collaboration in the broader project on smoking and cancer burden in Uruguay and for discussions that helped contextualise the empirical application. I also thank Constanza Garín, Luca Pruzzo, Guillermo Paraje, and Catalina Torres for comments, collaboration, and support with data sources in the broader project. I thank participants at seminars where earlier versions of this work were presented for helpful feedback. The modelling framework, analyses, interpretation, and manuscript are my sole responsibility.

A Additional methodological and implementation details

This appendix expands on the statistical and computational details summarised in Section 2 of the main text. It documents implementation choices that are central to replication but would be too detailed for the main paper.

A.1 Weighted prevalence likelihood and effective sample size construction

The smoking-prevalence layer is estimated from pooled repeated cross-sectional surveys, all of which arise from complex sampling designs. In principle, a fully design-based treatment would require harmonising the original expansion factors and survey designs across sources before constructing cell-level likelihood contributions. In the present implementation, we instead adopt a pragmatic post-stratification scheme designed to recover the correct population scale within age–year–sex cells while retaining a tractable binomial representation for the APC model.

Let $n^{\text{raw}}_{a,t,s}$ denote the number of pooled survey records observed in age a, calendar year t, and sex s, and let $P_{a,t,s}$ denote the corresponding external population total obtained from official demographic sources. We define a cell-specific expansion factor

$$\omega_{a,t,s} = \frac{P_{a,t,s}}{n^{\text{raw}}_{a,t,s}},$$

and assign this common factor to all survey records belonging to that cell. After this rescaling, the expanded totals recover the population margins by construction. Let g index survey-instrument group. The binomial pseudo-counts used in the prevalence layer are then formed as

$$N^P_{a,t,g,s} = \sum_{i \in (a,t,g,s)} \omega_{a,t,s}, \qquad Y^P_{a,t,g,s} = \sum_{i \in (a,t,g,s)} \omega_{a,t,s}\, \mathbf{1}\{i \text{ is a current smoker}\},$$

so that the cell-level smoking proportion remains the weighted survey proportion, while the total count scale is aligned with the external population totals.

Because all prevalence models are estimated separately by sex, the sex subscript is omitted in the main text and we write $N^P_{a,t,g}$ and $Y^P_{a,t,g}$ for notational simplicity. We refer to $N^P_{a,t,g}$ as an effective sample size because it is the cell-level total entering the binomial likelihood, even though it is produced by post-stratification rather than by a full joint treatment of the original survey design. This choice should be interpreted as a practical approximation adopted to make rapid pooled estimation feasible across heterogeneous survey sources, not as a complete design-based solution to the weighting problem. Survey-instrument differences are instead handled explicitly through the fixed effects δ_g in the prevalence predictor.

A.2 Smoking-state reconstruction and effective exposure

The smoking-prevalence layer provides a smoothed and projected surface of current smoking prevalence, denoted $p^C_{a,t}$, which serves as the sole input for the reconstruction of smoking states. This surface is not used directly as a predictor of incidence. Instead, it is converted into a cohort-age-period decomposition of smoking states, distinguishing current smokers,

former smokers by time since cessation, and never smokers.

For each age–period cell (a, t), the population is decomposed as

$$p^{C}_{a,t} + \sum_{s=1}^{L_q} p^{F,s}_{a,t} + p^{N}_{a,t} = 1,$$

where $p^{F,s}_{a,t}$ denotes the proportion of former smokers with s years since cessation, and $p^{N}_{a,t}$ denotes the proportion of never smokers.

The smoking-state composition evolves along cohort lines. Moving from $(a-1, t-1)$ to (a, t), the cohort inherits its previous smoking-state composition, adjusted for changes in current smoking prevalence. Let

$$p^{C,-}_{a,t} = p^{C}_{a-1,t-1}.$$

Define the decrease in current smoking as

$$\Delta^{-}_{a,t} = \max\{0,\ p^{C,-}_{a,t} - p^{C}_{a,t}\},$$

and the increase as

$$\Delta^{+}_{a,t} = \max\{0,\ p^{C}_{a,t} - p^{C,-}_{a,t}\}.$$

A decrease $\Delta^{-}_{a,t}$ is interpreted as a flow into the former-smoker stock with one year since cessation. The existing stock of former smokers ages forward by one year. An increase $\Delta^{+}_{a,t}$ is accommodated by drawing proportionally from the pool of non-current individuals (former and never smokers), preserving the internal composition of that pool.

The resulting smoking-state composition provides a cohort-consistent accounting of smoking histories across the Lexis surface.

An effective exposure term is then constructed as

$$e_{a,t} = p^{C}_{a,t} + \sum_{s=1}^{L_q} w_q(s)\, p^{F,s}_{a,t},$$

where $w_q(s)$ is a site-specific risk-reversal schedule satisfying $0 \le w_q(s) \le 1$ and decreasing in s. Current smokers carry full excess risk, former smokers carry a fraction that declines with time since cessation, and never smokers carry no smoking-related excess risk.

The effective exposure term $e_{a,t}$ is dimensionless and represents the share of smoking-related excess risk that remains active in each cell. It enters the incidence model through the offset

$$o^{I}_{a,t} = \log\bigl(1 + e_{a,t}(\mathrm{RR}_I - 1)\bigr),$$

where RR_I is an externally supplied relative-risk parameter.

This construction provides an explicit representation of the persistence and gradual dissipation of smoking-related excess risk without relying on reduced-form lag specifications.

A.3 Expected-deaths surface and incidence-to-mortality mapping

The mortality layer is not driven directly by projected smoking histories, but by an expected-deaths surface constructed from projected incidence using an external post-diagnosis mortality schedule.

Let $\hat{I}_{a,t}$ denote projected incidence from the incidence layer, and let $\{d_0, d_1, \ldots, d_L\}$

denote the annualised mortality schedule used to distribute diagnosis-related deaths over the years following diagnosis, with $d_\ell \geq 0$ and $\sum_{\ell=0}^{L} d_\ell \leq 1$. The expected-deaths surface is then defined by the discrete convolution

$$D_{a,t} = \sum_{\ell=0}^{L} d_\ell \, \hat{I}_{a-\ell,t-\ell},$$

up to the alignment conventions imposed by the Lexis grid.

The quantity $D_{a,t}$ is not itself a forecast of observed mortality. It is the mortality surface implied on epidemiological grounds by the projected incidence trajectory and the externally supplied post-diagnosis mortality schedule. In the mortality APC layer, $\log D_{a,t}$ enters the predictor as an offset, so that the residual APC terms model departures of observed mortality from the expected-deaths surface implied by projected incidence and the external schedule.

In practice, the construction of $D_{a,t}$ requires additional implementation choices beyond the formal convolution in the equation above. These include the annualisation rule used for deaths occurring in the first post-diagnosis year, the alignment convention linking diagnosis year and death year on the Lexis grid, and the treatment of edge cells near the beginning and end of the observed or projected incidence surface. These conventions matter because they determine how the expected-deaths surface is phased relative to the incidence projection and therefore how the incidence-to-mortality link behaves around the historical/forecast junction.

In the current implementation, projected incidence rates are first converted into projected incident cases by multiplying the fitted rate surface by the corresponding population exposure in each age–period cell. Let $\hat{\lambda}^I_{a,t}$ denote the projected incidence rate and $N_{a,t}$ the population exposure; then projected incident cases are

$$\hat{I}_{a,t} = \hat{\lambda}^I_{a,t} N_{a,t}.$$

The post-diagnosis mortality schedule is annualised from the three external probabilities reported for death within 0–1, 1–3, and 3–5 years after diagnosis, denoted $p_{0,1}$, $p_{1,3}$, and $p_{3,5}$. Under the operative `midyear_uniform` rule used in the code, these are converted into six annual lag weights:

$$w_0 = \tfrac{1}{2} p_{0,1}, \qquad w_1 = \tfrac{1}{2} p_{0,1} + \tfrac{1}{4} p_{1,3}, \qquad w_2 = \tfrac{1}{2} p_{1,3},$$

$$w_3 = \tfrac{1}{4} p_{1,3} + \tfrac{1}{4} p_{3,5}, \qquad w_4 = \tfrac{1}{2} p_{3,5}, \qquad w_5 = \tfrac{1}{4} p_{3,5}.$$

The expected-deaths surface is then constructed by discrete convolution of projected incident cases with this annualised kernel:

$$D_{a,t} = \sum_{\ell=0}^{5} w_\ell \, \hat{I}_{a-\ell,t-\ell}.$$

The alignment is cohort-consistent in the sense that both age and calendar period are shifted backward by ℓ years when tracing deaths at (a, t) back to diagnoses at $(a-\ell, t-\ell)$. In the operative implementation, if these shifted diagnosis cells fall below the minimum age or minimum calendar year available in the projected incidence surface, they are clipped to the

minimum supported age and year before the corresponding incident cases are retrieved. The resulting quantity $D_{a,t}$ is the expected number of externally implied deaths, and the mortality offset entering the predictor is $\log D_{a,t}$.

A.4 Prior specification and estimation strategy

All three model layers are estimated as latent Gaussian structured additive models within a Bayesian APC framework. The prevalence layer is specified with a binomial likelihood and logit link, while the incidence and mortality layers use Poisson likelihoods with population exposures entering as offsets. Estimation is carried out separately by sex.

Age, period, and cohort effects are modelled as smooth stochastic processes using Gaussian random walks. In the current implementation, second-order random walks (RW2) are used for age effects in all layers, while period and cohort effects are modelled using either first-order (RW1) or second-order random walks depending on the layer and the desired degree of smoothness. In the incidence and mortality layers, cohort effects are additionally regularised using coverage-weighted penalties to stabilise estimation near the boundaries of the observed Lexis surface.

The smoothing parameters governing these latent components are assigned penalised-complexity priors, parameterised in terms of the marginal standard deviation of the corresponding effect. These priors shrink the model toward a simpler baseline with no temporal variation while allowing sufficient flexibility to capture persistent structure in the data. All structured components are constrained to have zero mean, ensuring identifiability of the APC decomposition. The specific hyperparameter values used in the empirical implementation are reported in Table A1. These values define the degree of smoothing applied to each APC component and were chosen to provide stable estimation across layers while preserving interpretable temporal structure.

Table A1: APC hyperparameters.

Layer	Component	Model	Prior / value
Prevalence	Age	RW2	PC-prec (0.35, 0.01)
	Period	RW1	PC-prec (1.00, 0.01)
	Cohort	RW1	PC-prec (1.00, 0.01)
Incidence	Age	RW2	PC-prec (0.35, 0.01)
	Period	RW1	PC-prec (1.00, 0.01)
	Cohort	RW1	PC-prec (1.00, 0.01)
Mortality	Age	RW2	PC-prec (0.35, 0.01)
	Period	RW1	PC-prec (1.00, 0.01)
	Cohort	RW1	PC-prec (1.00, 0.01)

Notes: Values correspond to penalised-complexity priors on the marginal standard deviations of the latent effects. RW1 and RW2 denote first- and second-order Gaussian random walks. Future cohort effects are extrapolated using the common damped-trend rule described in the text. *Source:* Own elaboration.

In all layers, an explicit linear trend in centred calendar period is included as a separate component from the smooth period effect. This trend provides a parsimonious representation of low-frequency secular movement and helps stabilise long-horizon extrapolation. The corresponding slope parameter is assigned a weakly informative Gaussian prior centred at zero, with scale chosen to penalise unrealistically large long-run growth or decline while

allowing the data to determine the overall direction and magnitude of the trend. This specification separates long-run drift from higher-frequency variation captured by the smooth period effect and avoids confounding between these components.

Although the three layers share a common APC backbone, their roles differ across the sequential architecture. The prevalence layer recovers a smoothed representation of smoking behaviour from survey data, including adjustments for survey-instrument differences. The incidence layer combines APC structure with an externally constructed incidence offset, and the mortality layer combines APC structure with an expected-deaths offset derived from projected incidence. Accordingly, the APC components in the latter two layers should be interpreted as residual smooth deviations around these structured epidemiological inputs rather than as fully unconstrained APC components.

Cohort effects beyond the observed Lexis support are extrapolated using a deterministic damped-trend rule applied consistently across the three layers. Let γ_c denote the estimated cohort effect. For cohorts outside the observed range, a linear trend is estimated over the most recent observed cohorts using a fixed window, and future values are obtained by extending this trend with exponential damping. This prevents unrealistic divergence of cohort effects in long-horizon projections while preserving recent directional information.

Although the extrapolation rule is common across layers, its role in the predictor differs. In the prevalence layer, it governs the continuation of cohort smoking trajectories. In the incidence and mortality layers, it operates on the residual APC structure around the epidemiological offsets. This separation ensures that long-run extrapolation remains stable without overriding the transmission mechanisms encoded by the external epidemiological inputs.

In the empirical implementation, the damped-trend extrapolation is constructed using a fixed window of the last $K = 5$ observed cohorts and a damping factor $\rho = 0.8$. A linear trend is fitted over the last K cohorts of the estimated cohort trajectory, and future cohort effects are obtained by extending this trend with exponential damping governed by ρ. These parameter values are held fixed across all layers of the model and are chosen to ensure stable long-horizon behaviour while preserving recent cohort dynamics.

Posterior inference is performed using integrated nested Laplace approximations (INLA), which provide accurate approximations to the posterior distributions of latent Gaussian models while remaining computationally tractable for high-dimensional APC structures on the Lexis surface.

A.5 Posterior predictive simulation and uncertainty propagation

Uncertainty is propagated sequentially across epidemiological layers. The prevalence layer produces posterior summaries for the smoking surface and for the reconstructed smoking-state composition. These summaries define the effective exposure term $e_{a,t}$ and the incidence offset $o^I_{a,t}$. Projected incidence is then passed through the post-diagnosis mortality schedule to construct the expected-deaths surface $D_{a,t}$, which anchors the mortality layer.

Downstream predictive uncertainty therefore combines within-layer APC uncertainty with uncertainty inherited from upstream components. Variation in the estimated smoking surface affects effective exposure and the incidence offset; variation in projected incidence affects the expected-deaths surface through the convolution step; and the mortality layer is estimated conditional on that constructed surface.

Operationally, the model remains modular. Each layer is estimated separately within

the latent Gaussian APC framework, and posterior means and uncertainty intervals from one layer are transformed deterministically into inputs for the next. This avoids joint sampling from a single integrated posterior, which would be computationally demanding given the nonlinear exposure reconstruction, convolution-based incidence-to-mortality mapping, and high-dimensional APC surfaces. The resulting approximation keeps the transmission mechanism transparent while propagating the main upstream sources of uncertainty into downstream projections.

A.6 Border treatment and endogenous horizon diagnostics

Long-horizon projections on the Lexis surface become progressively less supported as the forecast moves away from the observed region. This is especially important in APC settings, because a future value depends not only on calendar-time extrapolation but also on the extent to which its age, period, and cohort coordinates remain anchored in historically observed information. The framework therefore supplements posterior uncertainty summaries with endogenous horizon diagnostics that describe the geometry of support.

Let $\mathcal{O}$ denote the observed region of the Lexis surface and let $\mathcal{P}$ denote the projected region. For each projected cell (a, t), cohort and period indices are mapped back to the observed support by clamping them to the historical range. Cell-level diagnostics record whether the mapped cohort lies on the boundary of the observed cohort support and how far the projected cohort must be shifted to match observed cohorts. When historical cohort information is available, a cohort-support measure is computed as the relative number of observed cells associated with the mapped cohort, normalised by the maximum support observed across cohorts.

These cell-level diagnostics are aggregated to the year level, separately by sex, using exposure weights. The main annual metrics are the exposure-weighted mean cohort-support fraction and the exposure-weighted share of cells whose cohort lies on the boundary of the observed support. Projected years are then classified into endogenous horizon categories. A year remains in the *credible* region while the mean support fraction is at least 0.50 and the boundary share is below 0.30. It moves to *caution* once either the mean support fraction falls below 0.50 or the boundary share reaches 0.30, and to *risky* once the mean support fraction falls below 0.33 or the boundary share reaches 0.50. Years with mean support below 0.20 or boundary share at least 0.80 lie beyond the maximum endogenous horizon supported by the data.

These categories are reporting devices rather than stochastic states of the model. They indicate where projected values remain close to observed Lexis support and where they rely increasingly on edge extrapolation. This distinction is useful in the sequential framework because support for smoking histories, incidence, and mortality may deteriorate at different rates. Projected trajectories remain continuous at the historical/forecast junction; horizon labels are applied ex post as diagnostic overlays.

B External epidemiological inputs: sources, constructions, and caveats

This appendix documents the external epidemiological inputs summarised in Section 2.5 of the main text. These inputs are not estimated within the APC framework itself. Instead, they encode the epidemiological quantities used to link smoking histories to incidence risk and projected incidence to an expected-deaths surface. Their role in the framework is therefore structural rather than descriptive: they determine how upstream smoking scenarios are transmitted across stages of the disease process once the prevalence, incidence, and mortality surfaces have been estimated from the data.

B.1 Site- and sex-specific external relative-risk parameters

The incidence layer requires site- and sex-specific smoker relative risks relative to the never-smoker baseline. In the present implementation, these are treated as fixed external parameters and enter the incidence predictor through the incidence offset defined in Section 2.2. Their substantive role is to determine the magnitude of the smoking-related incidence contrast once the reconstructed smoking-state composition has been translated into the effective exposure term.

In the final implementation, the hardcoded values coincide exactly with the IECS/Pichon-Riviere relative-risk table reported for Mexico in the 2013 technical report and with the corresponding supplementary table reported in the associated multicountry smoking-burden article (Pichon-Riviere et al., 2020, 2013). These tables report relative-risk parameters by condition, sex, and smoking status, with never smokers as the reference category. Because closely related publications within the same model family label analogous tables somewhat differently, we avoid imposing a narrower interpretation here and instead treat these quantities as external condition-specific epidemiological parameters used to map smoking exposure into excess cancer risk. In the lung-cancer application developed in the main text, the operative values are those reported for lung cancer by sex for current smokers relative to never smokers. Table B2 reports the full set of site- and sex-specific values used in the broader implementation.

Table B2: External relative-risk parameters.

	Relative risk	
Site	Men	Women
Lung	23.26	12.69
Larynx	14.60	13.02
Oral/pharynx	10.89	5.08
Esophagus	6.76	7.75
Stomach	1.96	1.36
Pancreas	2.31	2.25
Cervix	–	1.59
Kidney	2.72	1.29
Bladder	3.27	2.22

Notes: Values are external condition-specific relative-risk parameters by sex and smoking status, with never smokers as the reference category. They coincide with the IECS/Pichon-Riviere relative-risk table reported for Mexico and with the corresponding supplementary table in the associated multicountry smoking-burden article. *Source:* Own elaboration.

These relative-risk parameters are interpreted as applying to the smoking-state composition within each age–period cell, where current smokers carry the full excess risk, former smokers carry a time-dependent fraction determined by the risk-reversal schedule, and never smokers carry no smoking-related excess risk.

B.2 Risk-reversion schedules

The translation from smoking histories to current incidence risk is governed by a risk-reversion schedule that specifies how smoking-related excess risk persists or declines after cessation. The target quantity is the proportion of reducible excess risk remaining after u years since quitting, denoted

$$w(u) = \frac{RR(u) - 1}{RR(0) - 1},$$

where $RR(0)$ denotes the current-smoker relative risk and $RR(u)$ the relative risk after u years since cessation. By construction, all schedules satisfy $w(0) = 1$.

To preserve traceability, the project does not smooth the source values before defining the reported knots. When the literature reports exact years since cessation, those points are used directly. When the literature reports intervals rather than exact years, the implementation adopts representative thresholds or endpoints so that the resulting schedule remains directly linked to the published categories. Between reported knots, the operational implementation uses linear interpolation with endpoint clamping. This avoids introducing hidden interpolation assumptions at the schedule-construction stage. Likewise, when no credible sex-specific schedule is available, the same empirical schedule is duplicated across sexes rather than replaced by an ad hoc sex-specific approximation.

This choice implies that the evidentiary basis is heterogeneous across sites. Some schedules are closely aligned with the target quantity and come from high-quality pooled or meta-analytic evidence; others are based on indirect reconstructions or histology-specific proxies used to approximate the site definitions required by the project. The schedules are

therefore best understood as transparent external inputs with varying confidence rather than as quantities estimated internally from a common data-generating process.

The current project classifies the site-specific schedules into broad confidence tiers. Higher-confidence schedules include lung, bladder, and oral/pharynx (Reitsma et al., 2020; Freedman et al., 2011; La Vecchia et al., 1999). Moderate-confidence schedules include pancreas, stomach, kidney, and cervix (Iodice et al., 2008; Chao et al., 2002; Yuan et al., 1998; Appleby et al., 2006; HHS-OSG and NCCDPHP, 2020). Lower-confidence schedules include larynx and esophagus, where the implemented schedules rely more heavily on mixed comparators or proxy evidence (Altieri et al., 2002; Wang et al., 2017).

Table B3: Risk-reversion construction rules.

Rule	Description
Target quantity	$w(u) = (RR(u) - 1)/(RR(0) - 1)$
Current smokers	All rows with $u = 0$ set to $w(0) = 1$ by definition
Interpolation	None imposed
Category mapping	Representative thresholds used when sources report intervals rather than exact years
Sex handling	Empirical schedule duplicated across sexes when no credible sex-specific evidence is available
Negative values	Retained when published estimates cross the never-smoker baseline

Source: Own elaboration.

The general logic of the schedules follows the epidemiological literature on smoking cessation and excess-risk decline reviewed in sources such as (HHS-OSG and NCCDPHP, 2020), but the schedules used here are assembled site by site from the studies listed in Table B4.

Table B4: Risk-reversion schedule sources.

Site	Sources	Sex-specific	Confidence tier	Notes
Lung	Reitsma et al. (2020)	No	High	Direct meta-analytic estimates of reducible excess risk remaining after cessation.
Bladder	Freedman et al. (2011)	Yes	High	NIH-AARP; exact transformation within sex.
Oral/pharynx	La Vecchia et al. (1999)	No	High	Total oral/pharyngeal cancer; duplicated across sexes.
Larynx	Altieri et al. (2002)	No	Low	Approximate reconstruction from mixed comparators.
Esophagus	Wang et al. (2017)	No	Low	ESCC-based proxy for the aggregated esophagus site.
Pancreas	Iodice et al. (2008); HHS-OSG and NCCDPHP (2020)	No	Moderate	Slightly negative transformed value at 20 years retained as published.
Stomach	Chao et al. (2002); HHS-OSG and NCCDPHP (2020)	Yes	Moderate	Female schedule is non-monotone and should be treated as noisy evidence.
Kidney	Yuan et al. (1998); HHS-OSG and NCCDPHP (2020)	No	Moderate	Duplicated across sexes.
Cervix	Appleby et al. (2006); HHS-OSG and NCCDPHP (2020)	Female only	Moderate	Slightly negative transformed value at 10 years retained as published.

Source: Own elaboration.

The final implementation uses a site- and sex-specific excess-risk reversal schedule after cessation, expressed as the remaining fraction of smoking-related excess risk as a function of years since quitting; the exact schedule used in the code is reported in Table B5.

Table B5: Excess-risk reversal schedule after smoking cessation.

Site	Years since quit	Men	Women
Bladder	0	1.000000	1.000000
	4	0.803419	0.813699
	9	0.639667	0.682192
	10	0.321799	0.295890
Cervix	0	–	1.000000
	4	–	0.108696
	9	–	0.108696
	10	–	0.000000
Esophagus	0	1.000000	1.000000
	5	0.460650	0.460650
	10	0.237899	0.237899
	20	0.132354	0.132354
Kidney	0	1.000000	1.000000
	9	0.622642	0.622642
	19	0.471698	0.471698
	20	0.283019	0.283019
Larynx	0	1.000000	1.000000
	2	1.000000	1.000000
	9	0.578947	0.578947
	19	0.210526	0.210526
	20	0.126316	0.126316
Lung	0	1.000000	1.000000
	1	0.814000	0.814000
	5	0.572000	0.572000
	10	0.369000	0.369000
	15	0.267000	0.267000
	20	0.197000	0.197000
Oral cavity and pharynx	0	1.000000	1.000000
	2	0.709016	0.709016
	5	0.479564	0.479564
	9	0.332879	0.332879
	14	0.084469	0.084469
	15	0.057221	0.057221
Pancreas	0	1.000000	1.000000
	9	0.648649	0.648649
	10	0.202703	0.202703
	20	0.000000	0.000000
Stomach	0	1.000000	1.000000
	10	0.793103	0.632653
	19	0.551724	0.632653
	20	0.198276	0.632653

Notes: The schedule is implemented as the remaining fraction of smoking-related excess risk after cessation. Values between listed knots are obtained by linear interpolation, with endpoint clamping outside the observed range. *Source:* Own elaboration.

B.3 Post-diagnosis mortality schedules

The expected-deaths surface is parameterised by post-diagnosis mortality schedules derived from published net-survival information. For the broader multistate implementation developed in the project, the working source is the REDECAN report on adult cancer survival in Spain, period 2013–2017, using the age-standardised 15–99 summary (REDECAN, 2025). The attraction of this source is its homogeneous coverage by site and sex, which allows the project to build a consistent set of post-diagnosis schedules across all nine smoking-related cancer sites considered in the broader workflow.

The implemented quantities are not survival probabilities themselves, but probabilities of death by post-diagnosis interval derived from the published net-survival values at 1, 3, and 5 years. Let $SN(k)$ denote net survival at year k. The working transformation is:

$$p_{0\text{–}1} = 1 - SN(1), \qquad p_{1\text{–}3} = SN(1) - SN(3), \qquad p_{3\text{–}5} = SN(3) - SN(5),$$

with cumulative five-year mortality given by $1 - SN(5)$. In the current implementation, these three interval probabilities are annualised under the operative `midyear_uniform` rule into a six-lag kernel with weights

$$w_0 = \tfrac{1}{2}p_{0\text{–}1}, \qquad w_1 = \tfrac{1}{2}p_{0\text{–}1} + \tfrac{1}{4}p_{1\text{–}3}, \qquad w_2 = \tfrac{1}{2}p_{1\text{–}3},$$

$$w_3 = \tfrac{1}{4}p_{1\text{–}3} + \tfrac{1}{4}p_{3\text{–}5}, \qquad w_4 = \tfrac{1}{2}p_{3\text{–}5}, \qquad w_5 = \tfrac{1}{4}p_{3\text{–}5},$$

which are then used to construct the expected-deaths surface described in Section 2.3.

These inputs should be interpreted as pragmatic proxies rather than as direct estimates of Uruguayan post-diagnosis mortality. First, they are based on Spanish population net survival rather than on local follow-up data. Second, they are not crude cause-specific competing-risk probabilities, and therefore do not answer exactly the same question as the probability of dying from the cancer in the presence of competing causes. Third, the available temporal partition is limited to 0–1, 1–3, and 3–5 years because the source reports only 1-, 3-, and 5-year horizons. Finer temporal schedules would require annual survival tables or microdata. The use of these schedules therefore makes explicit a broader data limitation in applied burden projections: the epidemiological quantities needed to connect incidence and mortality are often unavailable from the target population and must be approximated from the best available external evidence.

Table B6: Post-diagnosis mortality probabilities.

	Men				Women			
Site	0–1	1–3	3–5	≤5	0–1	1–3	3–5	≤5
Oral/pharynx	29.8	19.4	6.7	55.9	20.2	13.8	6.6	40.6
Esophagus	56.5	24.2	4.9	85.6	56.6	22.0	3.9	82.5
Stomach	46.4	20.8	6.0	73.2	42.6	17.2	3.6	63.4
Pancreas	67.2	18.3	4.2	89.7	63.3	20.5	4.1	87.9
Larynx	14.3	16.0	7.1	37.4	15.1	13.0	7.1	35.2
Lung	58.8	19.6	5.3	83.7	50.0	20.2	6.3	76.5
Cervix	–	–	–	–	14.1	13.5	5.2	32.8
Kidney	17.4	9.1	5.6	32.1	19.6	7.7	4.1	31.4
Bladder	11.3	8.5	4.9	24.7	13.7	7.5	1.3	22.5

Notes: Values are percentages derived from REDECAN adult cancer net-survival estimates for Spain, 2013–2017. The intervals correspond to death probabilities within 0–1, 1–3, and 3–5 years after diagnosis; ≤ 5 reports cumulative five-year mortality. *Source:* Own elaboration.

C Additional simulation details and results

This appendix complements Section 3.1 of the main text. It documents the simulation design, the estimators compared in the validation exercises, the performance metrics used in the reported summaries, and the extended diagnostics that are not included in the main paper. The main text emphasises mortality scenario-effect recovery under the well-specified design. Here we provide the underlying data-generating details, intermediate pathway diagnostics, support-window comparisons, sex-specific results, and a misspecified transmission exercise.

C.1 Data-generating process

The simulation design follows the three-stage structure of the proposed framework. For each Monte Carlo seed, a latent smoking-prevalence surface is generated first, then translated into a smoking-related incidence offset, and finally mapped into mortality through the post-diagnosis mortality schedule. The simulation therefore provides a controlled setting in which the true prevalence, effective exposure, incidence, expected-deaths, and mortality surfaces are known.

The latent Lexis grid is indexed by sex, age, calendar year, and birth cohort. The outcome grid covers ages 35–89 and calendar years 1990–2070, with 2022 as the last historical year and 2023–2070 as the projection window. The prevalence process is generated on the broader age support required to reconstruct smoking histories for the outcome ages. In the realistic observed-window estimator, however, smoking prevalence is observed only over ages 20–65, matching the support restriction used in the empirical implementation. This distinction between the full latent support and the observed estimation window is central to the support-window diagnostics reported below.

For sex r, age a, period t, and cohort $c = t - a$, the latent current-smoking prevalence is generated from a smooth APC structure,

$$\operatorname{logit}\left(p^{C}_{a,t,r}\right) = \eta^{P}_{0} + \xi^{P}_{r} + A^{P}_{a} + B^{P}_{t} + C^{P}_{c},$$

where ξ^{P}_{r} is a sex-specific shift and A^{P}_{a}, B^{P}_{t}, and C^{P}_{c} are smooth stochastic age, period, and cohort components. The implementation uses random-walk-type components to generate smooth variation over the Lexis surface. Survey-like prevalence observations are then generated as binomial microdata within observed age–period–sex cells, with a fixed effective cell size used to mimic repeated cross-sectional prevalence information.

Future smoking-prevalence scenarios are imposed after 2022. Four trajectories are considered. In the frozen-prevalence scenario, each age–sex-specific prevalence path is held fixed at its 2022 value. In the one-percent annual increase and decrease scenarios, post-2022 prevalence is scaled from the 2022 age–sex reference level by $(1.01)^{t-2022}$ and $(0.99)^{t-2022}$, respectively. In the quit scenario, post-2022 current smoking prevalence is set to zero. The resulting scenario-specific prevalence surfaces are bounded to remain within the unit interval.

For each scenario, the current-smoking prevalence surface is converted into smoking-state histories for current, former, and never smokers using the same accounting logic as in the estimator. Former-smoker stocks are weighted by the risk-reversion schedule, producing an effective exposure term $e_{a,t,r}$. The smoking-related incidence offset is then

$$z_{a,t,r} = \log\{1 + e_{a,t,r}(\mathrm{RR}^{I}_{r} - 1)\},$$

where RR_r^I is the sex-specific external relative-risk parameter used in the incidence link.

The well-specified incidence data-generating process uses the same functional form as the estimator. The latent incidence log-rate is

$$\log \lambda_{a,t,r}^I = \eta_0^I + \xi_r^I + A_a^I + B_t^I + C_c^I + z_{a,t,r}.$$

Historical incidence counts are generated as

$$Y_{a,t,r}^I \sim \text{Poisson}\left(E_{a,t,r}^I \lambda_{a,t,r}^I\right),$$

where the simulation uses a common exposure scale across Lexis cells. The true future incidence surface is retained for all scenarios and is used to evaluate scenario-effect recovery.

Mortality is generated from incidence through the same post-diagnosis mortality mapping used by the estimator. Let $w_{\ell,r}$ denote the sex-specific annualised post-diagnosis mortality kernel. The true expected mortality burden is constructed as

$$D_{a,t,r} = \sum_{\ell=0}^{L} w_{\ell,r} \hat{I}_{a-\ell,t-\ell,r},$$

with the same Lexis alignment convention used in the estimation pipeline. Historical mortality counts are generated as Poisson draws around the resulting expected deaths. This construction makes mortality a downstream consequence of the simulated incidence path, rather than an independently generated outcome surface.

The robustness exercise uses a misspecified transmission design. In that design, prevalence still affects incidence monotonically and scenario responses remain well defined, but the incidence link no longer follows the log-linear offset imposed by SBAPC. Instead, the incidence contribution of smoking exposure is transformed through a saturating function,

$$g(z) = 2\tanh(z/1.5),$$

while the estimator continues to impose the well-specified log-linear offset. This design tests whether scenario-effect recovery deteriorates in an interpretable way when the smoking-to-incidence transmission rule is approximately but not exactly represented by the estimator.

C.2 Estimators compared in the simulation

The main estimator is SBAPC fitted under the realistic observed-window information set. It estimates the prevalence layer from the restricted survey-like prevalence window, reconstructs smoking states outside the directly observed support, constructs the incidence offset from effective exposure, and fits the mortality layer around the incidence-linked expected-deaths surface.

The autonomous BAPC benchmark is fitted as an outcome extrapolation model without the smoking-prevalence transmission channel. In the scenario-effect exercises, it receives no scenario-specific smoking information. Its estimated scenario effects relative to the frozen-prevalence path are therefore zero by construction. This property makes it a useful scenario-blind reference for the validation target considered in the paper.

The full-support SBAPC variant is used as a support-window diagnostic. It gives the estimator access to the broader latent support used to generate the truth. It is not used as an

empirical estimator. Its role is to quantify how much of the difference between observed-window SBAPC and Truth is associated with the restricted prevalence observation window.

The incidence-anchored SBAPC variant is used only in decomposition diagnostics. It preserves the incidence-to-mortality anchoring while disabling the prevalence-channel scenario contribution. This separates the mortality anchoring role of projected incidence from the full prevalence-to-incidence-to-mortality scenario pathway.

C.3 Evaluation metrics

The main scenario estimand is the annual mortality effect of scenario s relative to frozen prevalence,

$$\Delta_s^M(t) = B_s^M(t) - B_{\text{freeze}}^M(t),$$

where $B_s^M(t)$ is the aggregate number of deaths in year t under scenario s. The same definition is used for incidence when intermediate incidence diagnostics are reported.

For a horizon region $\mathcal{T}$, the cumulative scenario effect is

$$\Delta_s^M(\mathcal{T}) = \sum_{t \in \mathcal{T}} \Delta_s^M(t).$$

The cumulative recovery ratio is

$$R_s^M(\mathcal{T}) = \frac{\sum_{t \in \mathcal{T}} \hat{\Delta}_s^M(t)}{\sum_{t \in \mathcal{T}} \Delta_s^M(t)}.$$

A value close to one indicates that the estimator recovers the cumulative magnitude of the true mortality effect in that horizon region.

Annual mean absolute relative error is computed as

$$\text{MARE}_s^M(\mathcal{T}) = \frac{100}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left| \frac{\hat{\Delta}_s^M(t) - \Delta_s^M(t)}{\Delta_s^M(t)} \right|,$$

with implementation safeguards for years in which the true annual effect is numerically close to zero. Sign agreement is the share of projected years in which $\hat{\Delta}_s^M(t)$ and $\Delta_s^M(t)$ have the same sign.

For baseline-chain diagnostics under frozen prevalence, the same logic is applied to levels rather than scenario contrasts. For each object $Q_{a,t,r}$ in the sequential chain, annual aggregates are compared with their true counterparts. The objects reported below are effective exposure, incident cases, incidence-linked expected deaths, and mortality deaths. Effective exposure is aggregated using the exposure-weighted annual mean, while incidence and mortality objects are aggregated as annual counts.

Projection years are also grouped by endogenous horizon region. The classification follows the Lexis-support diagnostic described in Section 2 of the main text. In the simulation output used for the paper, the aggregate mortality displays share common transition years: the credible region runs through 2038, the caution region begins in 2039, and the risky region begins in 2048.

C.4 Baseline pathway recovery under frozen prevalence

The frozen-prevalence baseline provides a pathway check before scenario contrasts are evaluated. Table C7 compares observed-window SBAPC with Truth for the main sequential objects under the frozen-prevalence path. The purpose is to show where discrepancies arise along the chain and whether upstream deviations propagate strongly into downstream incidence and mortality.

Table C7: Baseline pathway recovery under frozen prevalence.

Object	Horizon	Recovery	MARE (%)	Bias (%)
Effective exposure	Credible	0.92 [0.89, 0.95]	7.8	-7.8
	Caution	0.92 [0.90, 0.95]	7.5	-7.4
	Risky	0.93 [0.90, 0.96]	7.5	-7.4
Incident cases	Credible	0.99 [0.96, 1.02]	4.2	-0.9
	Caution	0.98 [0.91, 1.06]	8.0	-1.6
	Risky	0.98 [0.87, 1.08]	11.9	-1.8
Incidence-linked expected deaths	Credible	0.99 [0.97, 1.02]	3.7	-0.7
	Caution	0.98 [0.92, 1.06]	7.7	-1.5
	Risky	0.98 [0.88, 1.08]	11.7	-1.7
Mortality deaths	Credible	0.99 [0.97, 1.02]	3.7	-0.6
	Caution	0.99 [0.92, 1.06]	7.7	-1.0
	Risky	0.98 [0.89, 1.09]	11.6	-1.0

Notes: Recovery is the estimated aggregate value divided by the true aggregate value within each endogenous horizon region. Effective exposure is aggregated as an exposure-weighted annual mean; incidence, expected deaths, and mortality are aggregated as annual counts. Bracketed intervals report the interquartile range across simulation seeds. *Source:* Own elaboration.

The table shows that incidence, incidence-linked expected deaths, and mortality deaths are recovered closely under the well-specified design. The largest discrepancy appears in effective exposure, which is the object most directly affected by the restricted prevalence observation window. Downstream recovery remains closer to Truth because incidence and mortality are also anchored by their own outcome histories and by the external transmission inputs. This pattern motivates the support-window comparison in the next subsection.

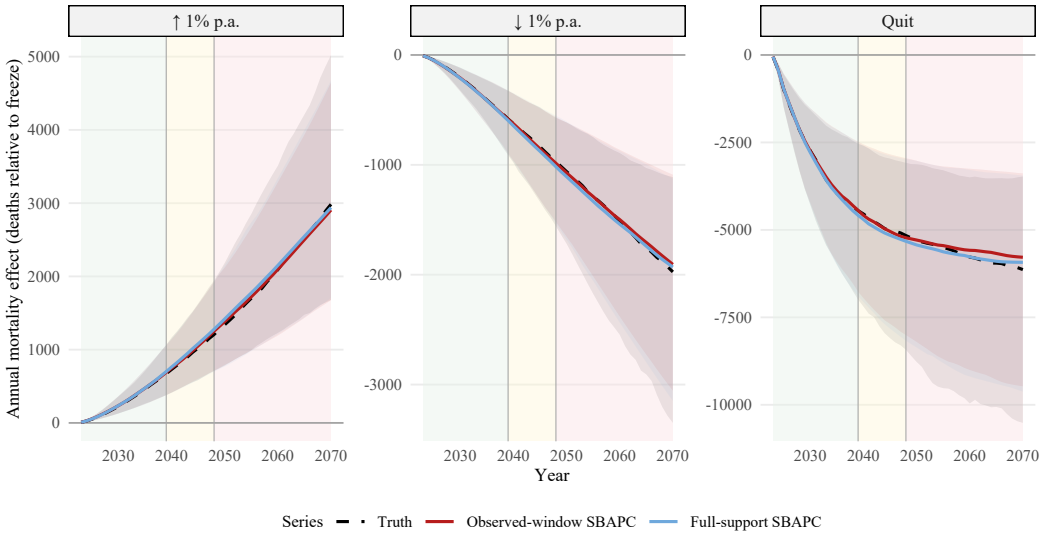
C.5 Observed-window and full-support diagnostics

Figure C1 compares annual mortality scenario effects for Truth, observed-window SBAPC, and full-support SBAPC. The figure aggregates across seeds and focuses on the same scenario-effect estimand used in the main text. The full-support curve shows how much recovery improves when the estimator is given access to the broader latent support used to generate Truth.

Table C8 summarises the same comparison by scenario and horizon region. The difference columns report the observed-window value minus the full-support value, so positive MARE differences indicate lower annual error under full support.

The full-support diagnostic reduces annual error modestly relative to the realistic observed-window estimator, while cumulative recovery ratios remain very similar. This supports the interpretation developed in Section 3.1: the observation window explains part of

Figure C1: Observed-window and full-support scenario-effect recovery.



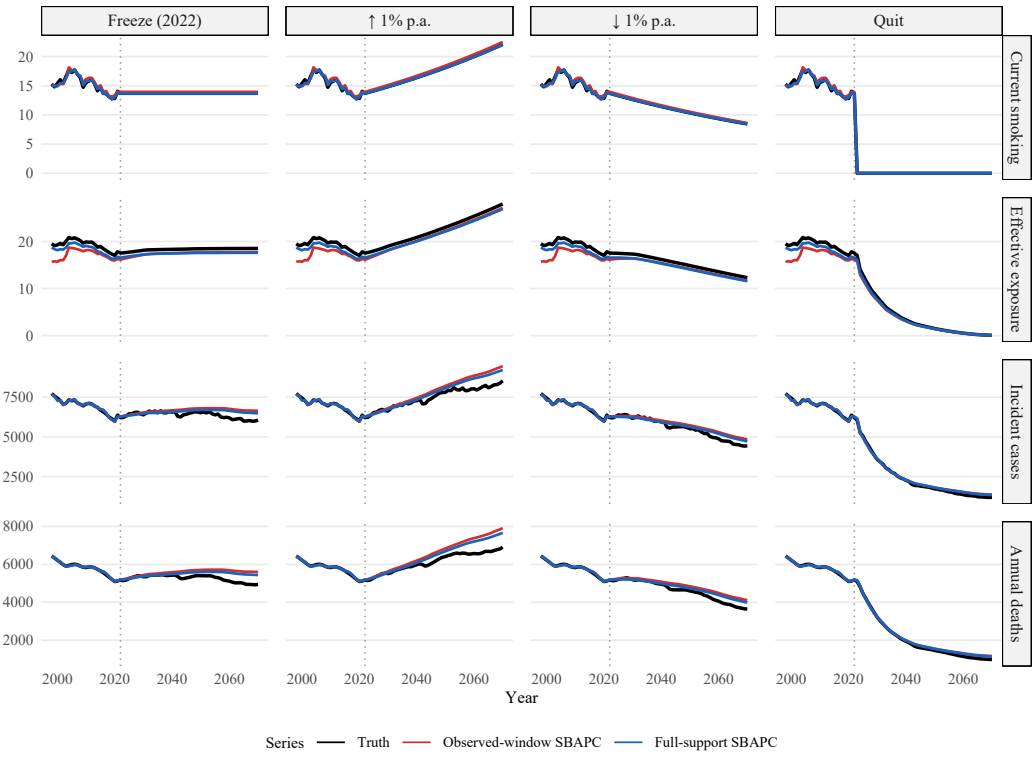
Notes: The figure compares Truth, observed-window SBAPC, and full-support SBAPC for annual mortality scenario effects relative to the frozen-prevalence baseline. Full-support SBAPC is an oracle-style diagnostic used to assess the role of support truncation. *Source:* Own elaboration.

Table C8: Observed-window penalty in scenario-effect recovery.

Scenario	Horizon	Obs. MARE	Full MARE	Obs. rec.	Full rec.
↑1%	Credible	7.1	5.9	1.02	1.02
	Caution	8.6	7.5	1.01	1.01
	Risky	12.4	11.5	0.99	1.00
↓1%	Credible	6.6	5.5	1.01	1.01
	Caution	8.3	7.4	0.99	1.00
	Risky	12.2	11.3	0.97	0.99
Quit	Credible	5.6	4.6	0.98	0.99
	Caution	8.2	7.3	0.97	0.99
	Risky	11.9	11.0	0.96	0.98

Notes: The table compares the realistic observed-window SBAPC estimator with the full-support diagnostic. MARE is reported in percentage points. Recovery is the cumulative estimated mortality scenario effect divided by the cumulative true effect within each endogenous horizon region. Difference columns are computed as observed-window minus full-support values. *Source:* Own elaboration.

Figure C2: Representative pathway recovery under restricted and full support.



Notes: The figure reports a representative male simulation and follows the sequential pathway from smoking prevalence and effective exposure to incidence and mortality. Truth is compared with observed-window SBAPC and full-support SBAPC under the four smoking scenarios. The full-support variant is an oracle-style diagnostic used to assess the role of support truncation; the observed-window estimator corresponds to the realistic information set used in the main simulation. *Source:* Own elaboration.

the remaining discrepancy from Truth, especially in upstream smoking-state reconstruction, but it does not drive the main scenario-effect recovery result.

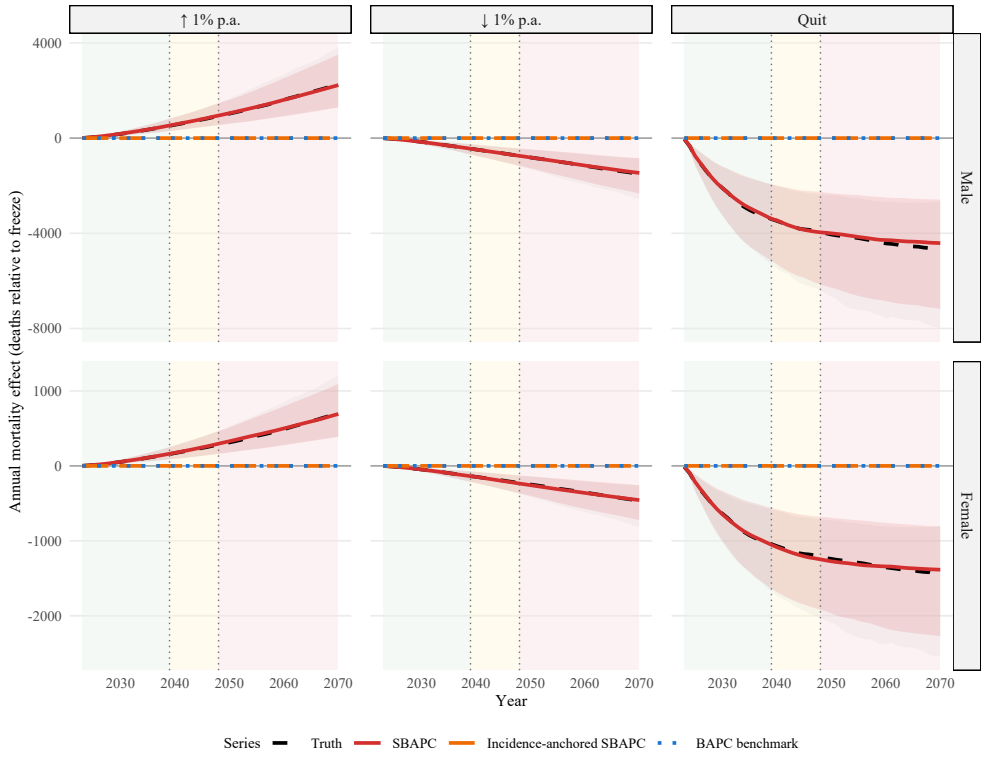
Figure C2 provides a representative pathway display in levels for one male simulation. The display is illustrative rather than a basis for aggregate performance assessment. Its purpose is to show where the observation-window restriction enters the sequential architecture and how the resulting discrepancy propagates from smoking prevalence and effective exposure to incidence and mortality.

The pathway display shows that the main support-window discrepancy appears upstream, in the reconstruction of smoking-state composition and effective exposure beyond the directly observed prevalence window. The full-support diagnostic tracks the true exposure pathway more closely because it is given access to the broader support used by the data-generating process. The observed-window estimator nevertheless preserves the ordering and broad temporal shape of the scenario-specific exposure paths. Downstream, incidence and mortality remain closely aligned across support variants because they are also anchored by simulated outcome data and by the epidemiological offsets.

C.6 Extended scenario-effect recovery

Figure C3 reports the scenario-effect recovery exercise separately by sex. It extends the main-text mortality figure by showing whether aggregate recovery masks systematic differences between males and females. The figure also includes the incidence-anchored diagnostic variant, which helps distinguish the contribution of the incidence-to-mortality anchor from the full prevalence-channel response.

Figure C3: Mortality scenario-effect recovery by sex.



Notes: Scenario effects are annual deaths relative to the frozen-prevalence baseline. The figure reports sex-specific recovery for Truth, SBAPC, the BAPC benchmark, and the incidence-anchored SBAPC diagnostic variant. *Source:* Own elaboration.

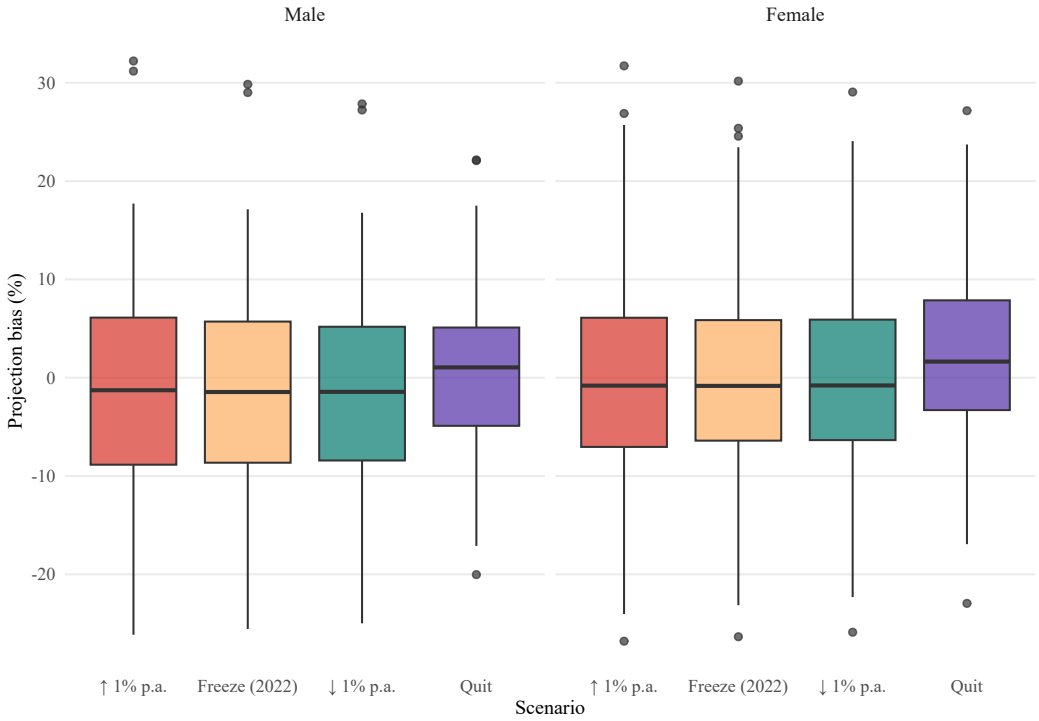
The sex-specific results show the same qualitative pattern as the aggregate main-text figure. SBAPC recovers the direction and temporal ordering of scenario effects for both males and females. Differences in absolute scale reflect the sex-specific relative risks, baseline burden, and mortality schedules used in the simulation. The incidence-anchored diagnostic helps show that mortality anchoring alone does not reproduce the full scenario response when the prevalence-to-incidence pathway is disabled.

C.7 Distributional diagnostics and calibration

Figure C4 summarises projection bias across simulation seeds. The purpose of this display is to show the dispersion of errors around the aggregate summaries reported in the main text and in the tables above.

The bias distributions are centred close to zero for the main well-specified scenario-effect contrasts, with wider dispersion in later projection years and in scenarios that generate

Figure C4: Distribution of projection bias across simulation seeds.



Notes: Boxplots summarise projection bias by scenario and sex across simulation seeds. *Source:* Own elaboration.

larger absolute mortality effects. This pattern is consistent with the horizon-based summaries: uncertainty and error increase as projections move away from the observed Lexis support.

Figure C5 reports reliability calibration by support horizon. This diagnostic evaluates the behaviour of uncertainty summaries and is secondary to the main validation target, which is recovery of scenario effects.

Calibration deteriorates as support weakens, which is expected in long-horizon APC projection. The diagnostic reinforces the role of the endogenous horizon classification as an interpretive device for reading projection results.

C.8 Misspecified transmission design

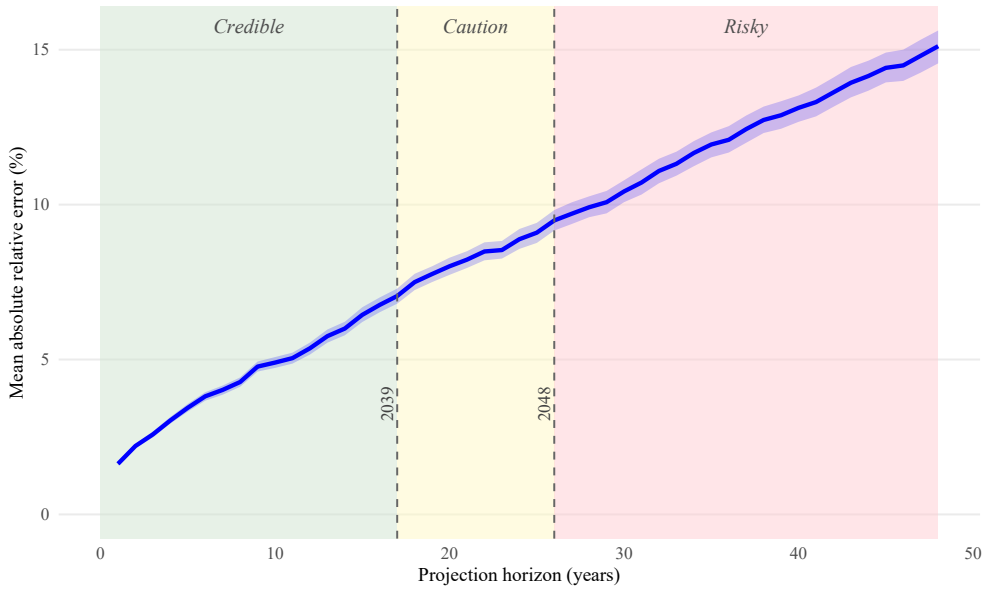
The misspecified transmission exercise perturbs the smoking-to-incidence link while preserving a monotone and scenario-responsive smoking channel. The estimator continues to impose the well-specified log-linear incidence offset, while the simulated truth uses the saturating function described in Subsection C.1. This exercise evaluates how scenario-effect recovery changes when the imposed transmission rule is only approximately correct.

Figure C6 compares scenario-effect recovery under the well-specified design and the misspecified transmission design.

Table C9 provides the corresponding numerical summary by scenario and horizon region.

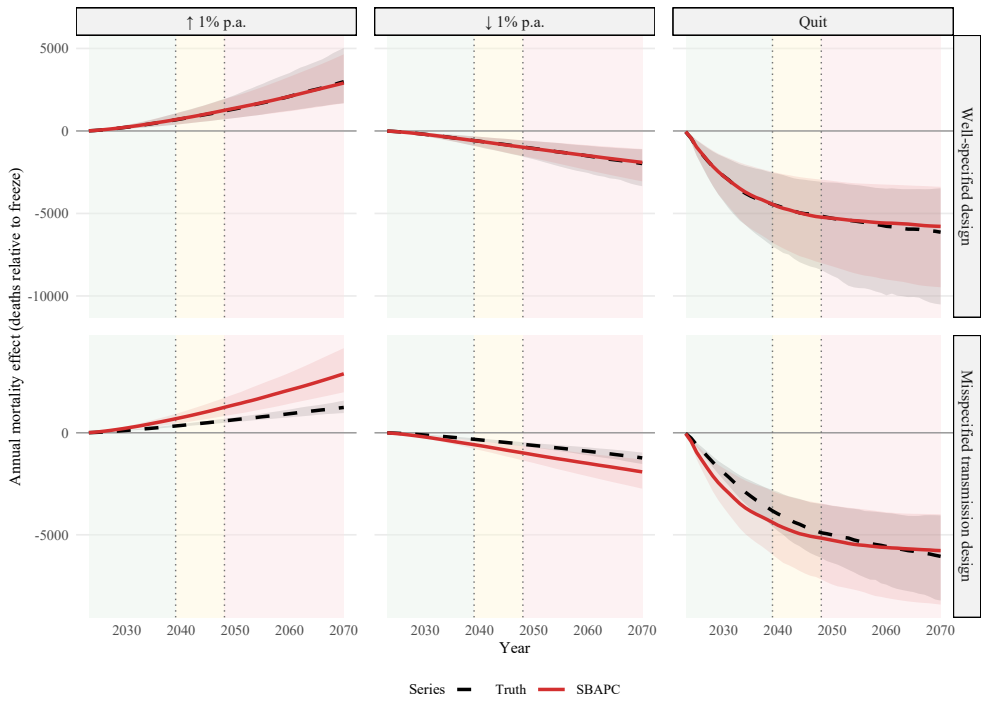
The misspecified design produces substantially larger magnitude errors than the well-specified design, especially for the gradual increase and decrease scenarios. Sign agreement remains high because the perturbed transmission rule preserves the monotone direction of the

Figure C5: Projection reliability calibration by support horizon.



Notes: Calibration is treated as an uncertainty-summary diagnostic. The central validation target is recovery of mortality scenario effects. *Source:* Own elaboration.

Figure C6: Scenario-effect recovery under transmission-rule misspecification.



Notes: Panels compare the well-specified design with a misspecified monotone transmission design. Scenario effects are annual mortality deaths relative to the frozen-prevalence baseline. *Source:* Own elaboration.

Table C9: Scenario recovery under transmission-rule misspecification.

Design	Scenario	Horizon	MARE (%)	Recovery	Sign (%)
Well-specified design	↑1%	Credible	7.1	1.02	100.0
		Caution	8.6	1.01	100.0
		Risky	12.4	0.99	100.0
	↓1%	Credible	6.6	1.01	100.0
		Caution	8.3	0.99	100.0
		Risky	12.2	0.97	100.0
	Quit	Credible	5.6	0.98	100.0
		Caution	8.2	0.97	100.0
		Risky	11.9	0.96	100.0
Misspecified transmission design	↑1%	Credible	101.7	1.94	100.0
		Caution	109.8	2.03	100.0
		Risky	130.6	2.21	100.0
	↓1%	Credible	84.9	1.78	100.0
		Caution	77.9	1.75	100.0
		Risky	66.2	1.61	100.0
	Quit	Credible	30.6	1.27	100.0
		Caution	14.0	1.10	100.0
		Risky	12.7	1.02	100.0

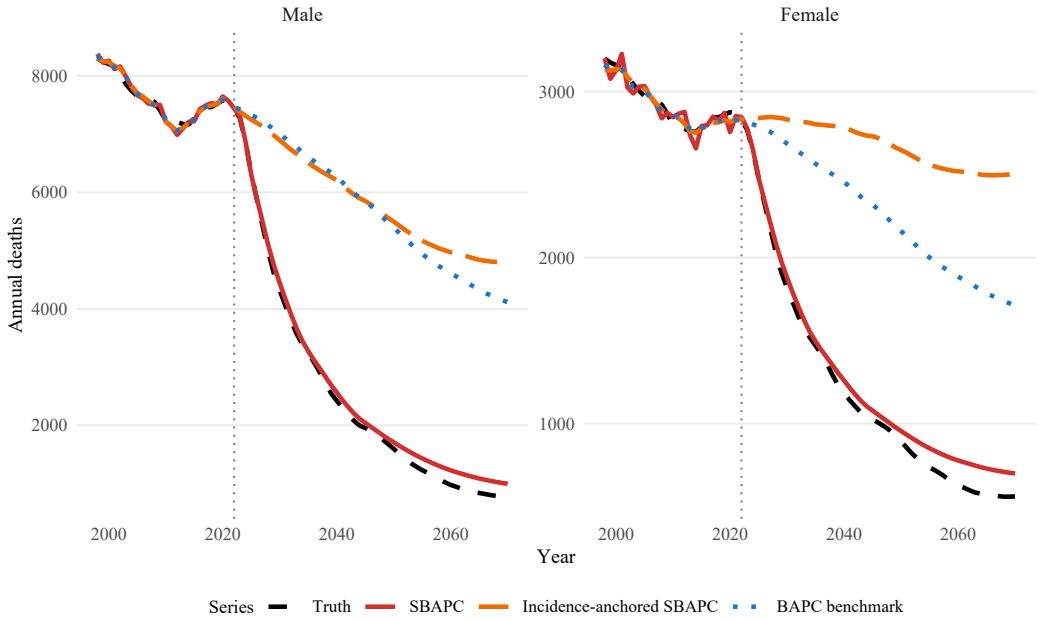
Notes: MARE is the annual mean absolute relative error for mortality scenario effects relative to frozen prevalence. Recovery is the cumulative estimated scenario effect divided by the cumulative true effect within each endogenous horizon region. The misspecified design uses a monotone but saturating smoking-to-incidence transmission rule, while SBAPC is fitted with the maintained log-linear offset. *Source:* Own elaboration.

smoking channel. The exercise therefore indicates that SBAPC is sensitive to the functional form of the transmission input when the objective is quantitative effect recovery, while still preserving the qualitative ordering and direction of scenario responses in this robustness design.

C.9 Seed-level illustrations

The aggregate results above summarise performance across simulation seeds. Figure C7 provides a representative seed-level trajectory display for the quit scenario. Its role is illustrative: it helps visualise how the pathway behaves in one simulation close to the median recovery pattern, while the aggregate figures and tables remain the basis for performance assessment.

Figure C7: Representative seed-level trajectory diagnostic.



Notes: The figure reports a median-recovery simulation for the quit scenario and compares Truth with the fitted projection along the simulated trajectory. *Source:* Own elaboration.

This single-seed display should be read together with the aggregate diagnostics. It is included to make the simulated trajectories visually interpretable, not to substitute for the seed-level summaries reported elsewhere in this appendix.

D Uruguay application: data, preprocessing, and empirical diagnostics

This appendix documents the Uruguayan empirical implementation used in Section 3.2. It records the data sources, harmonisation decisions, reconstruction procedures, and Lexis support underlying the application, and then reports extended diagnostics kept outside the main text: descriptive input series, mortality effects relative to frozen prevalence, benchmark comparisons, cumulative scenario effects, and fit summaries.

D.1 Tobacco-control policy context

The empirical application is interpreted against Uruguay's tobacco-control policy environment. The estimation window overlaps with a sustained sequence of regulatory and fiscal interventions that are relevant for understanding observed changes in smoking prevalence and, with a lag, downstream cancer incidence and mortality. The purpose of this subsection is not to provide a full institutional history, but to summarise the policy milestones most directly relevant for the application.

Table D10 summarises selected milestones. The sequence begins with ratification of the WHO Framework Convention on Tobacco Control, followed by implementation of smoke-free environments from 2006 onward, progressively stronger package warnings, reinforcement of restrictions on advertising, promotion, and sponsorship, ratification of the Protocol to Eliminate Illicit Trade in Tobacco Products, and the adoption of plain packaging. Tobacco taxation also became markedly stricter over this period.

Table D10: Selected tobacco-control milestones in Uruguay.

Year	Policy milestone
2004	Ratification of the WHO Framework Convention on Tobacco Control.
2005	Entry into force of the WHO Framework Convention on Tobacco Control; IMESI tax set at 70% as part of stricter tobacco taxation.
2006	100% smoke-free environments implemented in enclosed public spaces and in public and private workplaces.
2006	Rotating graphic health warnings introduced on cigarette packages, covering 50% of the front and back faces.
2007	VAT set at 22%, reinforcing the fiscal component of tobacco control.
2014	Law 19.244 reinforces the ban on tobacco advertising, promotion, and sponsorship, including points of sale previously exempted.
2014	Uruguay ratifies the Protocol to Eliminate Illicit Trade in Tobacco Products.
2019	New package-warning rotation established, with health warnings covering 80% of both principal faces of the pack.
2020	Plain packaging for cigarettes enters into force.

Notes: The table provides concise policy context for the empirical Uruguayan application rather than a full legal chronology. It highlights milestones most directly relevant for interpreting long-run smoking dynamics.
Source: Own elaboration.

D.2 Smoking-prevalence survey sources and harmonisation

Smoking prevalence is constructed from a harmonised repository of repeated cross-sectional surveys carried out in Uruguay between 1998 and 2024 (Ferre et al., 2025). The pooled file includes sources with probability-based sampling, sufficient methodological documentation, a smoking module, and variables comparable in age, sex, and geographic location. The resulting harmonised database contains 536,263 individual records after harmonisation and geographic restrictions.

The surveys incorporated in the pooled file include the Encuesta Continua de Hogares (ECH), Encuesta Nacional de Adolescencia y Juventud (ENAJ), Encuesta Nacional de Salud (ENS), Encuesta Nacional en Hogares sobre Consumo de Drogas (JND-H), Global Adult Tobacco Survey (GATS), Stepwise approach to NCD risk factors (STEPS), the International Social Survey Programme (ISSP), Encuesta Nacional sobre Consumo de Drogas en Estudiantes de Enseñanza Media (JND-EM), Global Youth Tobacco Survey (GYTS), and Global school-based student health survey (GSHS). For compactness, the bibliography reports a single institutional reference for each survey series, corresponding to the most recent wave used in the harmonised dataset.

Table D11 reports the survey microdata retained after harmonisation and geographic restrictions, by calendar year and source. Youth-only surveys are included only when a broad-coverage household survey is also available in the same calendar year, preventing structural zeros at adult ages that would otherwise compromise APC estimation on the age–period grid.

Table D11: Pooled smoking-prevalence microdata.

Year	ECH	ENAJ	ENS	JND-H	GATS	STEPS	ISSP	JND-EM	GYTS	GSHS	Total
1998				3 474							3 474
2001				2 256				2 108			4 364
2006	27 538			6 442	1 994			1 208			37 182
2008				4 032			805				4 837
2009				2 559				2 349			4 908
2011	72 402			4 817		4 023					81 242
2013				3 245				2 462			5 707
2014	72 767		2 532	4 229		6 705			1 289		87 522
2017	62 580			3 342	3 184						69 106
2018	57 342			1 861				4 616		3 416	67 235
2019	56 828	333		1 169				2 079			60 409
2021	14 378							3 605			17 983
2022	28 914										28 914
2023	28 862										28 862
2024	29 316							5 202			34 518
Total	450 927	12 813	2 532	31 036	5 743	4 456	805	22 206	2 458	3 287	536 263

Notes: Cell entries report the number of individual records retained after harmonisation and geographic restrictions. Blank cells indicate that the corresponding source does not contribute records in that calendar year. Survey acronyms and design features are summarised in Table D12. *Source:* Own elaboration.

Table D12 summarises the main design features of the survey series contributing to the pooled smoking-prevalence dataset. Survey years are reported only in Table D11 to avoid duplicating temporal information.

Smoking is harmonised as a binary indicator of current smoking. When a survey

Table D12: Survey series used for prevalence harmonisation.

Acronym	Survey	Institution	Target population	Coverage	Smoking item
ECH	Encuesta Continua de Hogares	INE	Ages 12+	National coverage, including rural and urban areas and localities with fewer than 5 000 inhabitants	Current smoking
ENAJ	Encuesta Nacional de Adolescencia y Juventud	MIDES / INJU	Adolescents and youth	National coverage, including urban localities with 5 000+ inhabitants	Habitual smoking
ENS	Encuesta Nacional de Salud	MSP	Ages 14+	Urban population in localities with 5 000+ inhabitants	Current smoking
JND-H	Encuesta Nacional en Hogares sobre Consumo de Drogas	JND / OUD	Household population	Urban population in localities with more than 10 000 inhabitants	Smoked in the last 30 days
GATS	Encuesta Mundial de Tabaquismo en Adultos	WHO / MSP / INE	Ages 15–64	National coverage, including rural and urban areas and localities with fewer than 5 000 inhabitants	Current smoking
STEPS	Encuestas Nacionales de Factores de Riesgo de Enfermedades Crónicas No Transmisibles	MSP	Adults	Urban population in localities with more than 10 000 inhabitants	Current smoking
ISSP	International Social Survey Programme	dECON-FCS	Ages 18+	Urban population in localities with 5 000+ inhabitants	Current smoking
JND-EM	Encuesta Nacional sobre Consumo de Drogas en Estudiantes de Enseñanza Media	JND / OUD	Secondary-school students aged 13–17	Secondary-school population in localities with more than 10 000 inhabitants	Smoked in the last 30 days
GYTS	Encuesta Mundial de Tabaquismo en Jóvenes	WHO / MSP	Students aged 13–15	Secondary-school population with national coverage	Smoked in the last 30 days
GSHS	Encuesta Mundial de Salud Escolar	WHO / MSP	School-based adolescents	Secondary-school population in urban localities with 5 000+ inhabitants	Smoked in the last 30 days

Notes: Coverage statements and target populations summarise the original survey documentation and may vary across waves. Rural observations are excluded in all waves for the analysis in this paper. *Source:* Own elaboration.

records current smoking directly, that variable is used; when it instead records cigarette use in the last 30 days, that measure is used as the operational proxy. Because questionnaire wording and target populations differ across surveys, the prevalence APC model includes survey-instrument effects in the predictor, as described in Section 2 of the main text.

Geographic comparability is imposed by restricting the analysis to Montevideo and the rest of urban Uruguay, excluding rural observations throughout. This restriction reflects the minimum common support available across the pooled surveys. The empirical implementation further restricts the prevalence layer to ages 20–65 to maintain comparability across survey designs and to avoid unstable estimation in sparsely observed edge cells.

For weighting, the pooled survey file does not attempt to reconstruct a unified design-based system from the original survey expansion factors. Instead, post-stratification is used to recover known population totals within year–age–sex–region cells, as detailed in Supplementary Appendix A. This choice should be interpreted as a pragmatic harmonisation device adopted to make pooled APC estimation feasible across heterogeneous surveys.

D.3 Lung-cancer incidence data and reconstruction to the Lexis grid

Lung-cancer incidence data are drawn from Uruguay's National Cancer Registry. The analytical file used in this paper consists of tabulations by sex, age group, and period provided directly by the Registry and not publicly distributed in this form. The source files cover the period 1998–2022, and incidence is observed by grouped ages and multi-year periods rather than on a single-year Lexis grid.

The reconstruction begins by expressing each observed cell as a case density over the corresponding age-by-period rectangle. Let $Y_{g,p}$ denote the observed number of cases in age group g and period block p. For each cell, the script extracts the initial and final years of the

block, computes its width in years, and defines the midpoint year. Age-group lower bounds are parsed from the reported labels, age-group widths are assigned, and midpoint ages are computed. Observed grouped counts are then converted into densities by dividing by the product of the period width and age width.

These midpoint densities are interpolated over a fine grid of single years and single ages using bivariate interpolation implemented with `akima::interp`. The working grid spans calendar years 1991–2022 and ages 0–85. The default interpolation is nonlinear, with automatic fallback to linear interpolation if the nonlinear routine fails for a given cause–sex combination. Negative fitted values are truncated at zero. This yields a preliminary annualised surface of nonnegative incidence counts on the fine Lexis grid.

Because the interpolated surface is only an approximation to the grouped observations, it is reconciled back to the original published totals through a multiplicative raking step. Each single-year observation is reassigned to its original grouped age band and period block using a dynamically constructed year-to-period mapping. The interpolated surface is re-aggregated to the original grouping, and for each grouped cell a correction factor is computed as the ratio of observed to re-aggregated cases. These cell-specific factors are merged back onto the annualised surface and applied multiplicatively, so that the final reconstructed surface remains smooth at the single-age, single-year level while reproducing the original grouped counts by construction when re-aggregated.

A final adjustment is applied to the oldest ages. Because the working grid is truncated at age 85 while the original grouped data include an open-ended category, the script sets all counts at ages 85 and above equal to the fitted value at age 84 within each cause–sex–year cell. For the empirical application studied here, lung-cancer incidence is defined using the registry's ICD-10 site classification for malignant neoplasm of bronchus and lung (C34). Since the analytical incidence series begins in 1998, no crosswalk across earlier ICD revisions is required within the study window.

D.4 Lung-cancer mortality data and coding

Lung-cancer mortality is obtained from official Uruguayan vital statistics compiled by the Ministry of Public Health. The analytical file used in this paper consists of disaggregated tabulations by cause, age, and sex provided directly by the Ministry and not publicly distributed in this form. For the empirical implementation, grouped death counts are transformed into a single-age annual mortality surface using penalised composite link modelling (PCLM), applied separately by cause, year, and sex.

Let $D_{g,t,s}$ denote observed deaths in grouped age interval g, year t, and sex s. The smoothing script nests the mortality data by cause, year, and sex and applies a PCLM fit to each grouped profile. Before fitting, exact zeros are replaced by a small positive value to avoid numerical issues. The PCLM routine is then called with grouped ages as inputs and deaths as responses, using the control settings $\lambda = 1$, deg = 2, kr = 2, and nlast = 5. The fitted output is a vector of single-age death counts running from age 0 up to the last age implied by the grouped profile.

This smoothing step serves a different purpose from the incidence reconstruction. Mortality is already observed annually, so no interpolation over period is needed. The objective is to disaggregate grouped age counts into a smooth single-age profile that can be aligned with the APC mortality layer and with the population denominators defined on the same Lexis grid. Because the procedure is applied separately by cause, year, and sex, it

preserves the annual timing of the mortality series while regularising the age profile within each observed year.

For the empirical application studied here, lung-cancer mortality is defined using ICD-10 code C34. Since the analytical mortality series used in the model begins in 1998, the operative cause definition is already observed under ICD-10 throughout the estimation window.

D.5 Population denominators and combined Lexis support

Population denominators by age, sex, and region are constructed for 1950–2045 from two official INE revisions (INE, 2025; INE and CELADE, 1998). For 1996–2045, the empirical application uses the 2025 revision; for 1950–1995, it uses the previous national revision prepared jointly with CELADE. The two series are spliced in 1996 by sex and age using a multiplicative adjustment factor based on the overlap year.

Because both revisions are provided primarily in five-year age groups, they are disaggregated to single-year ages before alignment with the incidence and mortality inputs. The open-ended group at ages 90 and above also requires harmonisation, since the projection series extends to older ages with more internal detail than the historical estimates. To make the full period comparable, the sex structure of the projected 90+ group is used retrospectively to distribute the corresponding historical totals.

The Uruguayan empirical illustration combines information observed on different parts of the Lexis surface. Smoking prevalence is learned from pooled repeated cross-sectional surveys observed over a relatively recent period and a younger age range, whereas the incidence and mortality layers rely on registry and vital-statistics data observed over older ages. Figure D8 summarises this support geometry. The prevalence window corresponds to the survey-based smoking layer, estimated over ages 20–65. The outcome window corresponds to the incidence and mortality layers, estimated over ages 35–89. Their overlap identifies the cohort region over which upstream smoking histories and downstream cancer outcomes can be linked most directly in the empirical illustration.

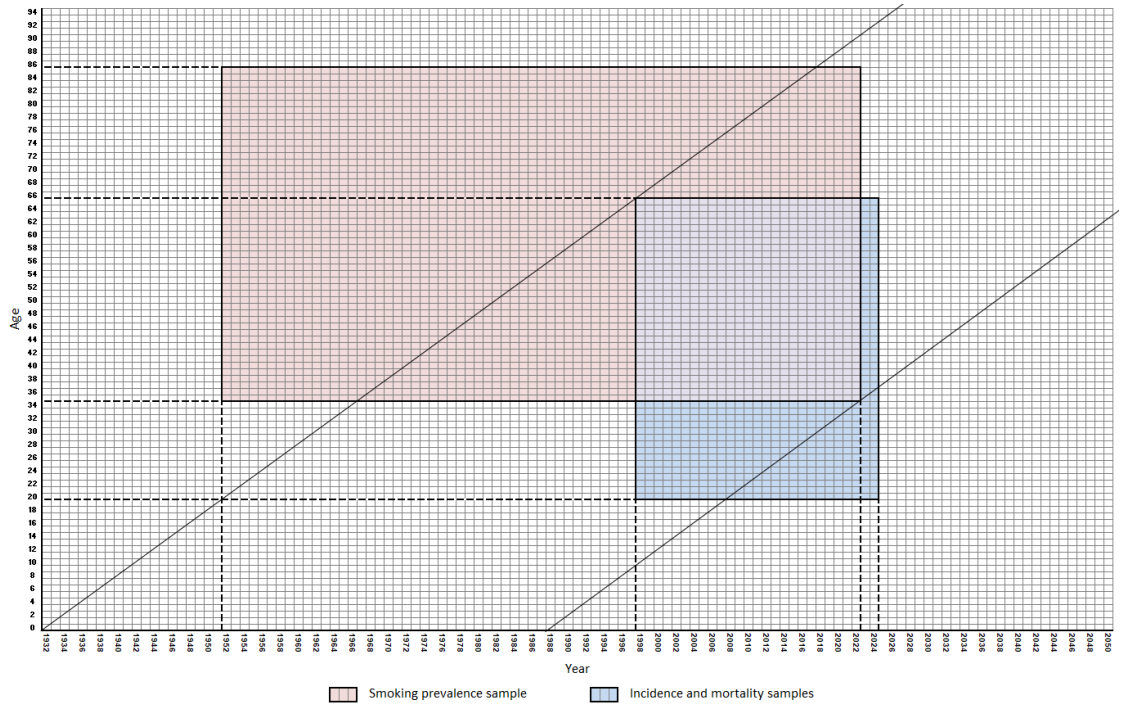
This geometry matters for interpretation. The availability of outcome data at older ages does not imply equally rich support for the full smoking-to-incidence-to-mortality sequence, because the upstream smoking information is only available for a narrower part of the Lexis surface. Likewise, recent prevalence surveys do not provide direct information at the oldest ages where cancer outcomes are concentrated. This motivates the use of endogenous horizon diagnostics in the main framework and the restriction of the main empirical summaries to the common endogenous horizon used in Section 3.2.

D.6 Descriptive overview of empirical inputs

Figure D9 brings together the three empirical inputs used in the lung-cancer application. The figure is descriptive: it does not report fitted APC surfaces, model-based projections, or scenario effects. Smoking prevalence is computed from the weighted survey microdata underlying the prevalence layer and aggregated over ages 20–65. Lung-cancer incidence and mortality are computed from the preprocessed historical inputs used in the corresponding outcome layers, aggregated over ages 35–89 and expressed as crude annual rates per 100,000 population.

The descriptive series help situate the empirical application before the model-based

Figure D8: Combined Lexis support for the Uruguayan application.



Notes: The pink rectangle indicates the approximate support of the smoking-prevalence sample, and the blue rectangle indicates the support used for the incidence and mortality layers. Their overlap identifies the cohort region over which smoking histories and downstream cancer outcomes can be linked most directly. Diagonal lines indicate birth cohorts on the Lexis surface. *Source:* Own elaboration.

results. They show a marked decline in smoking prevalence among males and a more gradual evolution among females, alongside sex-specific differences in lung-cancer incidence and mortality. These patterns are the empirical context within which the scenario-responsive projections in Section 3.2 should be interpreted.

D.7 Extended lung-cancer projection diagnostics

The main text reports projected lung-cancer mortality levels by sex. Figure D10 expresses the same projections as annual effects relative to the frozen-prevalence scenario, summed over sexes.

Table D13 summarises cumulative incidence and mortality effects by endogenous horizon region.

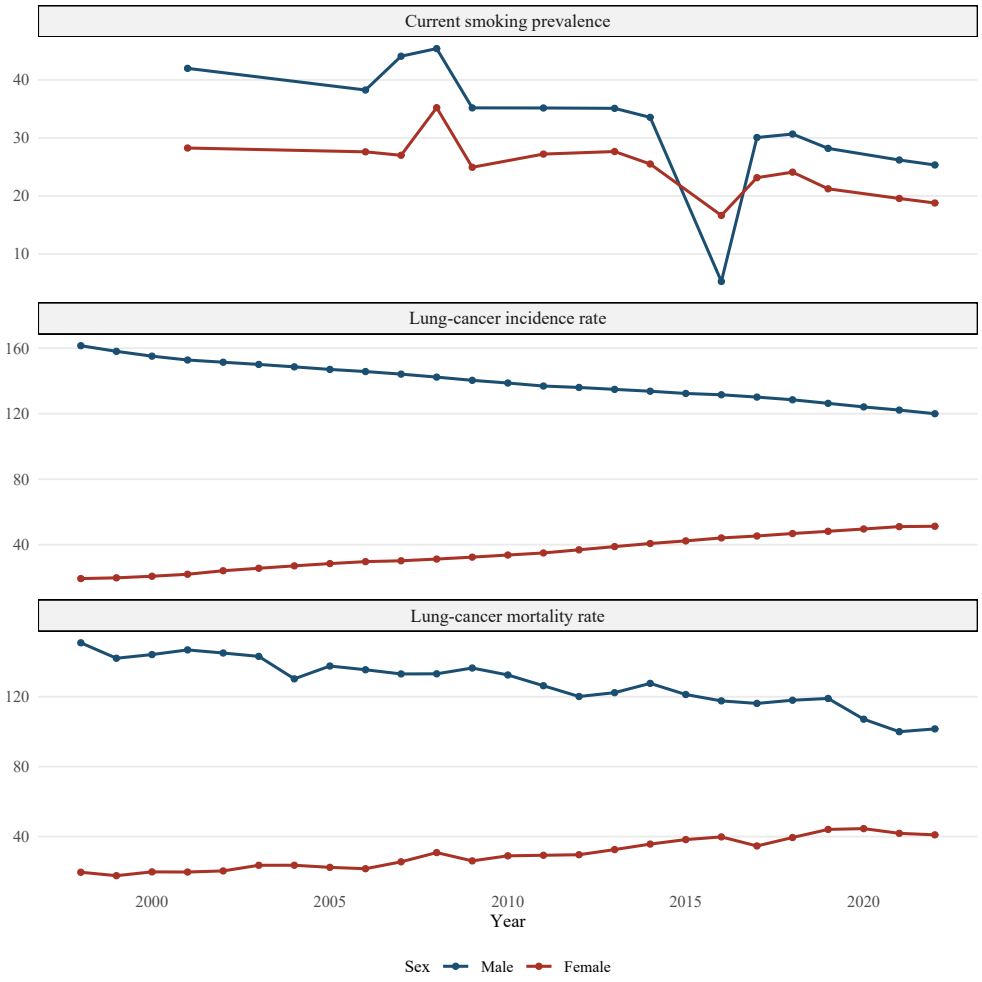
These results are reported as diagnostics because the main empirical section emphasises projected mortality levels rather than differences from the frozen-prevalence path.

D.8 Benchmark comparison and fit diagnostics

Figure D11 compares SBAPC projections with an autonomous BAPC benchmark. The benchmark extrapolates outcome histories without scenario-specific smoking information, while SBAPC transmits smoking scenarios through the sequential prevalence–incidence–mortality architecture. The comparison is therefore a diagnostic contrast between autonomous outcome extrapolation and scenario-responsive projection.

Table D14 reports compact fit diagnostics for the empirical models. These summaries

Figure D9: Historical empirical inputs for the Uruguayan application.



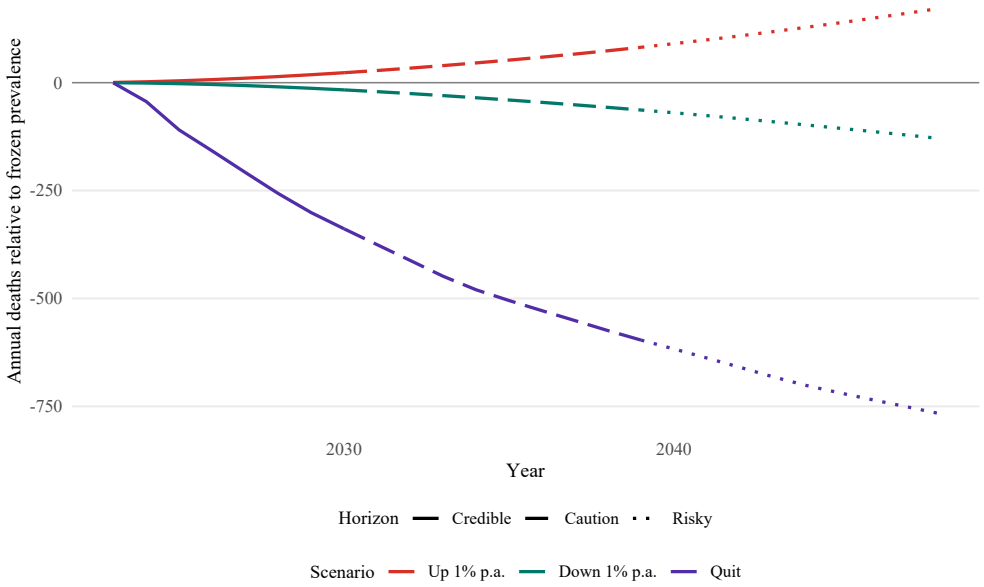
Notes: Smoking prevalence is computed from the weighted harmonised survey microdata used in the prevalence layer and aggregated over ages 20–65. Incidence and mortality are computed from the preprocessed historical inputs used in the empirical implementation, aggregated over ages 35–89, and expressed as crude annual rates per 100,000 population. *Source:* Own elaboration.

Table D13: Cumulative scenario effects in the Uruguay application.

Scenario	Horizon	Inc. effect	Mort. effect	Inc. %	Mort. %
↑1%	Credible	125	80	1.0	0.8
	Caution	657	481	4.8	4.1
	Risky	1 566	1 164	10.9	9.8
↓1%	Credible	-88	-54	-0.7	-0.5
	Caution	-513	-369	-3.7	-3.2
	Risky	-1 198	-886	-8.3	-7.4
Quit	Credible	-2 126	-1 416	-17.8	-13.6
	Caution	-5 734	-4 471	-41.8	-38.4
	Risky	-7 985	-6 267	-55.3	-52.5

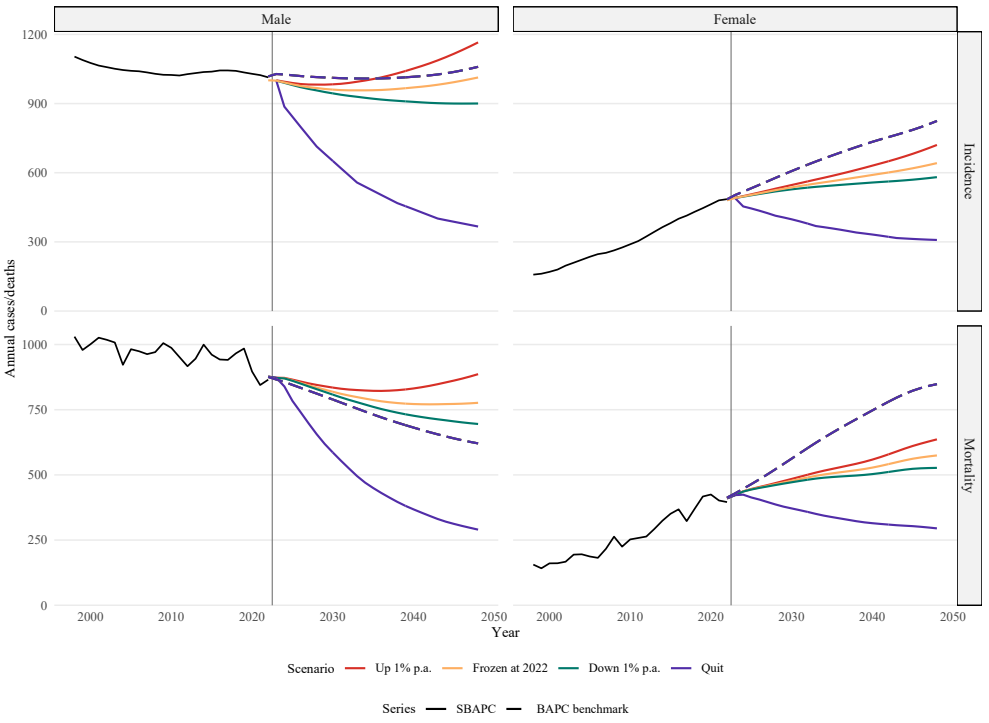
Notes: Values are cumulative effects relative to the frozen-prevalence scenario within each endogenous horizon region. Incidence effects are cumulative incident cases, and mortality effects are cumulative deaths. *Source:* Own elaboration.

Figure D10: Mortality effects relative to frozen prevalence.



Notes: Effects are annual lung-cancer deaths relative to the frozen-prevalence scenario, summed over sexes. Only non-reference smoking-prevalence scenarios are shown. Line type indicates the endogenous horizon region. *Source:* Own elaboration.

Figure D11: SBAPC and BAPC benchmark projections.



Notes: The figure compares scenario-responsive SBAPC projections with scenario-blind BAPC benchmark projections for the Uruguay application. The benchmark is included as a diagnostic reference for autonomous outcome extrapolation. *Source:* Own elaboration.

help assess whether the sequential specification behaves plausibly on the observed data, but they should not be read as the primary validation of the framework. The methodological validation is provided by the simulation study in Section 3.1; the empirical diagnostics reported here are used to detect gross misalignment between the fitted model, the benchmark projection, and the historical outcome series.

Table D14: Empirical fit diagnostics.

Sex	Layer	Estimator	ΔWAIC	ΔDIC	dLCPO	RMSE
Male	Incidence	BAPC benchmark	0.0	0.0	0.0	22.0
	Incidence	SBAPC incidence layer	-3.8	-7.9	1.9	22.1
	Mortality	BAPC benchmark	0.0	0.0	0.0	18.4
	Mortality	SBAPC mortality layer	0.8	-4.8	-0.4	18.4
	Mortality	Incidence-anchored mortality layer	3.6	-0.0	-1.8	18.4
Female	Incidence	BAPC benchmark	0.0	0.0	0.0	8.3
	Incidence	SBAPC incidence layer	-0.1	-1.8	0.0	8.3
	Mortality	BAPC benchmark	0.0	0.0	0.0	6.8
	Mortality	SBAPC mortality layer	-14.8	-18.7	7.4	6.6
	Mortality	Incidence-anchored mortality layer	-11.7	-12.4	5.8	6.6

Notes: The table reports compact diagnostic summaries for the empirical models. ΔWAIC and ΔDIC are differences in the widely applicable information criterion and deviance information criterion, computed relative to the relevant BAPC benchmark when such a comparison is meaningful; negative values favour the row estimator. dLCPO is the difference in log conditional predictive ordinate relative to the benchmark; positive values favour the row estimator. RMSE is the backtesting root mean squared error; lower values favour the row estimator. These diagnostics are descriptive and are not the main validation target of the paper.
Source: Own elaboration.

The diagnostics do not indicate a gross empirical penalty from introducing the sequential structure. In the incidence layer, SBAPC performs very similarly to the autonomous BAPC benchmark: the male incidence fit improves modestly in information criteria with virtually unchanged backtest RMSE, while the female incidence fit is nearly indistinguishable from the benchmark. In the mortality layer, the male results remain essentially comparable across estimators, whereas the female SBAPC mortality layer improves substantially in WAIC, DIC, dLCPO, and RMSE. The incidence-anchored mortality variant generally lies close to the full SBAPC mortality layer, which is consistent with the role of projected incidence as an important mortality anchor.

These diagnostics support the interpretation used in the main text. SBAPC is evaluated primarily as a scenario-responsive projection framework, so empirical fit diagnostics are used to detect gross misalignment with observed data rather than to define the main validation target.

D.9 Sensitivity to fixed external inputs

The empirical implementation treats relative risks, risk-reversion schedules, and post-diagnosis mortality schedules as fixed external inputs. This subsection reports a deterministic sensitivity analysis for the Uruguayan lung-cancer application. The purpose is to assess how much the main empirical projections depend on the particular values used for these externally supplied transmission components.

The analysis is one-at-a-time. Each alternative specification perturbs one external input block while leaving the rest of the empirical pipeline unchanged. The relative-risk perturbations shift the incidence relative risk by -0.20 or $+0.20$ on the logarithmic scale, approximately multiplying the baseline sex-specific values by $\exp(-0.20)$ or $\exp(0.20)$. The risk-reversion perturbations alter the time scale of the excess-risk reversal schedule, generating slower and faster reversion alternatives. The post-diagnosis perturbations change the timing of mortality after diagnosis by shifting 10% of the kernel mass toward later or earlier post-diagnosis intervals. This last perturbation changes the temporal distribution of post-diagnosis mortality rather than scaling all probabilities up or down.

Table D15 reports cumulative projected lung-cancer deaths within the common endogenous projection horizon. Results are shown for the frozen-prevalence and quit scenarios, separately by sex.

Table D15: Sensitivity to fixed external inputs.

Sex	Scenario	Current	RR		Reversal		Postdx timing	
			Low	High	Slow	Fast	Later	Earlier
Male	Frozen at 2022	20 837	20 895	20 788	20 716	20 936	20 635	20 917
	Quit	12 834	13 133	12 579	14 025	11 852	12 779	12 861
Female	Frozen at 2022	13 151	13 264	13 054	13 162	13 154	13 026	13 214
	Quit	9 001	9 367	8 678	9 616	8 518	8 933	9 038

Notes: Values are cumulative projected lung-cancer deaths within the common endogenous projection horizon. The current specification uses the baseline external inputs. RR Low and RR High shift the incidence relative risk by -0.20 and $+0.20$ on the logarithmic scale. Reversal Slow and Reversal Fast perturb the timing of risk reversion after cessation. Postdx Later and Postdx Earlier shift 10% of the post-diagnosis mortality kernel toward later or earlier intervals. The analysis is deterministic and one-at-a-time; the columns should not be interpreted as posterior uncertainty intervals. *Source:* Own elaboration.

The sensitivity patterns are consistent with the structure of the model. Under frozen prevalence, cumulative mortality changes only modestly across the perturbations because the smoking trajectory is held fixed and the mortality layer remains anchored by the observed historical outcome series. Under the quit scenario, the risk-reversion schedule is more influential, especially among males, because the speed at which excess risk declines after cessation directly governs how quickly the intervention affects incidence and downstream mortality. Perturbing the post-diagnosis timing has a smaller effect on cumulative deaths within the common horizon, since it redistributes deaths across years rather than changing the overall scale of post-diagnosis mortality.

The main empirical conclusion is stable across these alternatives. In both sexes, projected cumulative deaths under the quit scenario remain substantially below the frozen-prevalence projection under every perturbation considered. The exercise should therefore be read as a diagnostic for dependence on fixed external inputs. It is not a full probabilistic propagation of epidemiological-parameter uncertainty.

D.10 Multisite implementation details

The multisite exercise in Section 3.3 applies the same sequential architecture to nine smoking-related cancer sites for which the broader implementation contains site- and sex-specific external inputs. These include relative-risk parameters, risk-reversion schedules, and post-diagnosis mortality schedules documented in Appendix B. The multisite figure in the main

text reports projected mortality across smoking-related cancer sites. It should be interpreted as a composition exercise over sites epidemiologically associated with smoking, without applying population-attributable fractions.

The multisite exercise shows how the framework can be used to study the composition of future mortality across tumour sites. Once the external inputs are organised by site and sex, the same machinery used for lung cancer can be applied to a broader set of outcomes. This makes it possible to examine whether changes in smoking prevalence alter not only the level of future mortality, but also the distribution of projected deaths across cancer sites.

The main text reports only the frozen-prevalence and quit scenarios for the multisite display. This choice keeps the figure focused on the contrast between a baseline smoking path and a strong cessation counterfactual. Additional scenario-specific multisite outputs can be generated from the same pipeline, but they are not needed for the main argument of the empirical illustration.

References

A. Altieri, C. Bosetti, R. Talamini, S. Gallus, S. Franceschi, F. Levi, L. Dal Maso, E. Negri, and C. La Vecchia. Cessation of smoking and drinking and the risk of laryngeal cancer. *British Journal of Cancer*, 87(11):1227–1229, 2002. ISSN 0007-0920. doi: 10.1038/sj.bjc.6600638.

P. Appleby, V. Beral, A. Berrington de González, D. Colin, S. Franceschi, A. Goodill, J. Green, J. Peto, M. Plummer, and S. Sweetland. Carcinoma of the cervix and tobacco smoking: Collaborative reanalysis of individual data on 13,541 women with carcinoma of the cervix and 23,017 women without carcinoma of the cervix from 23 epidemiological studies. *International Journal of Cancer*, 118(6):1481–1495, 2006. ISSN 0020-7136. doi: 10.1002/ijc.21493.

S. A. Bashir and J. Estève. Projecting cancer incidence and mortality using Bayesian age-period-cohort models. *Journal of Epidemiology and Biostatistics*, 6(3):287–296, 2001. ISSN 1359-5229. doi: 10.1080/135952201317080698.

Marta Blangiardo, Michela Cameletti, Gianluca Baio, and Håvard Rue. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 7:39–55, 2013. ISSN 1877-5853. doi: 10.1016/j.sste.2013.07.003.

Freddie Bray and Bjørn Møller. Predicting the future burden of cancer. *Nature Reviews Cancer*, 6(1):63–74, January 2006. ISSN 1474-1768. doi: 10.1038/nrc1781.

C. C. Brown and L. G. Kessler. Projections of lung cancer mortality in the United States: 1985-2025. *Journal of the National Cancer Institute*, 80(1):43–51, 1988. ISSN 0027-8874. doi: 10.1093/jnci/80.1.43.

B. Carstensen. Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*, 26 (15):3018–3045, 2007. ISSN 1097-0258. doi: 10.1002/sim.2764.

Ann Chao, Michael J. Thun, S. Jane Henley, Eric J. Jacobs, Marjorie L. McCullough, and Eugenia E. Calle. Cigarette smoking, use of other tobacco products and stomach cancer mortality in US adults: The Cancer Prevention Study II. *International Journal of Cancer*, 101(4):380–389, 2002. ISSN 0020-7136. doi: 10.1002/ijc.10614.

D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Statistics in Medicine*, 6(4):449–467, 1987a. ISSN 0277-6715. doi: 10.1002/sim.4780060405.

D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. II: Age–period–cohort models. *Statistics in Medicine*, 6(4):469–481, 1987b. ISSN 1097-0258. doi: 10.1002/sim.4780060406.

Mark S. Clements, Bruce K. Armstrong, and Suresh H. Moolgavkar. Lung cancer rate predictions using generalized additive models. *Biostatistics*, 6(4):576–589, 2005. ISSN 1465-4644. doi: 10.1093/biostatistics/kxi028.

G. De Angelis, R. De Angelis, L. Frova, and A. Verdecchia. MIAMOD: A computer package to estimate chronic disease morbidity using mortality and survival data. *Computer*

Methods and Programs in Biomedicine, 44(2):99–107, 1994. ISSN 0169-2607. doi: 10.1016/0169-2607(94)90091-4.

Majid Ezzati, Alan D. López, Anthony Rodgers, and Christopher J. L. Murray. *Comparative Quantification of Health Risks : Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, volume 2. World Health Organization, Geneva, world health organization edition, 2004. ISBN 92 4 158031 3.

Zuleika Ferre, Manuel Flores, and Mariana Gerstenblüth. Harmonized Uruguayan Smoking Survey Dataset, 1998-2024, 2025.

Manuel Flores. Replication materials for "A sequential Bayesian age–period–cohort framework for cancer projections under smoking scenarios". Zenodo, 2026.

Neal D. Freedman, Debra T. Silverman, Albert R. Hollenbeck, Arthur Schatzkin, and Christian C. Abnet. Association Between Smoking and Risk of Bladder Cancer Among Men and Women. *JAMA*, 306(7):737–745, 2011. ISSN 0098-7484. doi: 10.1001/jama.2011.1142.

Connor Gascoigne, Andrea Riebler, and Theresa Smith. Smooth predictions for age-period-cohort models: A comparison between splines and random process. *BMC Medical Research Methodology*, 25(1):177, July 2025. ISSN 1471-2288. doi: 10.1186/s12874-025-02629-8.

GBD 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 1474-547X. doi: 10.1016/S0140-6736(20)30752-2.

(United States Public Health Service Office of the Surgeon General) HHS-OSG and (National Center for Chronic Disease Prevention and Health Promotion) NCCDPHP. *Smoking Cessation: A Report of the Surgeon General*. Publications and Reports of the Surgeon General. US Department of Health and Human Services, Washington (DC), 2020.

Theodore R. Holford, Rafael Meza, Kenneth E. Warner, Clare Meernik, Jihyoun Jeon, Suresh H. Moolgavkar, and David T. Levy. Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964-2012. *JAMA*, 311(2):164–171, 2014. ISSN 1538-3598. doi: 10.1001/jama.2013.285112.

INE. Estimaciones y proyecciones de la población de Uruguay y sus departamentos, 2012–2070, 2025.

INE and CELADE. Estimaciones y proyecciones de la población por sexo y edad, total del país 1950-2050. Revisión 1998., 1998.

International Agency for Research on Cancer. *Personal Habits and Indoor Combustions*, volume 100 E of *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Working Group on the Evaluation of Carcinogenic Risks to Humans, world health organization edition, 2012.

Simona Iodice, Sara Gandini, Patrick Maisonneuve, and Albert B. Lowenfels. Tobacco and the risk of pancreatic cancer: A review and meta-analysis. *Langenbeck's Archives of Surgery*, 393(4):535–545, 2008. ISSN 1435-2451. doi: 10.1007/s00423-007-0266-2.

L. Knorr-Held. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17-18):2555–2567, 2000. ISSN 0277-6715. doi: 10.1002/1097-0258(20000915/30)19:17/18<2555::aid-sim587>3.0.co;2-\#.

Leonhard Knorr-Held and Evi Rainer. Projections of lung cancer mortality in West Germany: A case study in Bayesian prediction. *Biostatistics*, 2(1):109–129, 2001. ISSN 1465-4644. doi: 10.1093/biostatistics/2.1.109.

Carlo La Vecchia, Silvia Franceschi, Cristina Bosetti, Fabio Levi, Renato Talamini, and Eva Negri. Time Since Stopping Smoking and the Risk of Oral and Pharyngeal Cancers. *JNCI: Journal of the National Cancer Institute*, 91(8):726a–728, 1999. ISSN 0027-8874. doi: 10.1093/jnci/91.8.726a.

David T. Levy, Sung-il Cho, Young-Mee Kim, Susan Park, Mee-Kyung Suh, and Sin Kam. SimSmoke Model Evaluation of the Effect of Tobacco Control Policies in Korea: The Unknown Success Story. *American Journal of Public Health*, 100(7):1267–1273, 2010. ISSN 0090-0036. doi: 10.2105/AJPH.2009.166900.

Qingwei Luo, Julia Steinberg, Xue Qin Yu, Marianne Weber, Michael Caruana, Sarsha Yap, Paul B. Grogan, Emily Banks, Dianne L. O'Connell, and Karen Canfell. Projections of smoking-related cancer mortality in Australia to 2044. *Journal of Epidemiology and Community Health*, 76(9):792–799, 2022. ISSN 1470-2738. doi: 10.1136/jech-2021-218252.

Thiago G. Martins, Daniel Simpson, Finn Lindgren, and Håvard Rue. Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, 67:68–83, 2013. ISSN 0167-9473. doi: 10.1016/j.csda.2013.04.014.

Rafael Meza, Jihyoun Jeon, Iakovos Toumazis, Kevin Ten Haaf, Pianpian Cao, Mehrad Bastani, Summer S. Han, Erik F. Blom, Daniel E. Jonas, Eric J. Feuer, Sylvia K. Plevritis, Harry J. de Koning, and Chung Yin Kong. Evaluation of the benefits and harms of lung cancer screening with low-dose computed tomography: Modeling study for the US Preventive Services Task Force. *JAMA*, 325(10):988–997, 2021. ISSN 1538-3598. doi: 10.1001/jama.2021.1077.

Bjørn Møller, Harald Fekjaer, Timo Hakulinen, Laufey Tryggvadóttir, Hans H. Storm, Mats Talbäck, and Tor Haldorsen. Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention*, 11 Suppl 1:S1–96, 2002. ISSN 0959-8278.

Bjørn Møller, Harald Fekjær, Timo Hakulinen, Helgi Sigvaldason, Hans H. Storm, Mats Talbäck, and Tor Haldorsen. Prediction of cancer incidence in the Nordic countries: Empirical comparison of different approaches. *Statistics in Medicine*, 22(17):2751–2766, 2003. ISSN 1097-0258. doi: 10.1002/sim.1481.

Suresh H. Moolgavkar, Theodore R. Holford, David T. Levy, Chung Yin Kong, Millenia Foy, Lauren Clarke, Jihyoun Jeon, William D. Hazelton, Rafael Meza, Frank Schultz, William McCarthy, Robert Boer, Olga Gorlova, G. Scott Gazelle, Marek Kimmel, Pamela M. McMahon, Harry J. de Koning, and Eric J. Feuer. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000. *Journal of the National Cancer Institute*, 104(7):541–548, 2012. ISSN 1460-2105. doi: 10.1093/jnci/djs136.

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.

Andrés Pichon-Riviere, Ariel Bardach, Joaquín Caporale, Andrea Alcaraz, Federico Augustovski, Francisco Caccavo, Luz Myriam Reynales-Shigematsu, Belén Sáenz de Miera-Juárez, José Alberto Muños-Hernández, Verónica Gallegos-Rivero, and Esteban Hernández-San Román. Carga de Enfermedad Atribuible al Tabaquismo en México. Technical Report Documento 10, Instituto de Efectividad Clínica y Sanitaria, Buenos Aires, Argentina, 2013.

Andrés Pichon-Riviere, Andrea Alcaraz, Alfredo Palacios, Belén Rodríguez, Luz Myriam Reynales-Shigematsu, Márcia Pinto, Marianela Castillo-Riquelme, Esperanza Peña Torres, Diana Isabel Osorio, Leandro Huayanay, Cesar Loza Munarriz, Belén Sáenz de Miera-Juárez, Verónica Gallegos-Rivero, Catherine De La Puente, María del Pilar Navia-Bueno, Joaquín Caporale, Javier Roberti, Sacha Alexis Virgilio, Federico Augustovski, and Ariel Bardach. The health and economic burden of smoking in 12 Latin American countries and the potential effect of increasing tobacco taxes: An economic modelling study. *The Lancet Global Health*, 8(10):e1282–e1294, 2020. ISSN 2214-109X. doi: 10.1016/S2214-109X(20)30311-9.

REDECAN. *Supervivencia de Cáncer En La Población Adulta En España, 2003-2017*. REDECAN, España, 2025.

Marissa Reitsma, Parkes Kendrick, Jason Anderson, Nicholas Arian, Rachel Feldman, Emmanuela Gakidou, and Vin Gupta. Reexamining Rates of Decline in Lung Cancer Risk after Smoking Cessation. A Meta-analysis. *Annals of the American Thoracic Society*, 17 (9):1126–1132, 2020. ISSN 2325-6621. doi: 10.1513/AnnalsATS.201909-659OC.

Andrea Riebler and Leonhard Held. Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations. *Biometrical Journal. Biometrische Zeitschrift*, 59(3):531–549, 2017. ISSN 1521-4036. doi: 10.1002/bimj.201500263.

Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392, 2009. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2008.00700.x.

Kenji Shibuya, Mie Inoue, and Alan D. Lopez. Statistical modeling and projections of lung cancer mortality in 4 industrialized countries. *International Journal of Cancer*, 117(3): 476–485, 2005. ISSN 0020-7136. doi: 10.1002/ijc.21078.

Daniel Robert Smith, Alireza Behzadnia, Rabbiaatul Addawiyah Imawana, Muzammil Nahaboo Solim, and Michaela Louise Goodson. Exposure–lag response of smoking prevalence on lung cancer incidence using a distributed lag non-linear model. *Scientific Reports*, 11 (1):14478, 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-91644-y.

Qiao-Li Wang, Shao-Hua Xie, Wen-Tao Li, and Jesper Lagergren. Smoking Cessation and Risk of Esophageal Cancer by Histological Type: Systematic Review and Meta-analysis. *JNCI: Journal of the National Cancer Institute*, 109(12):djx115, 2017. ISSN 0027-8874. doi: 10.1093/jnci/djx115.

Xue Qin Yu, Qingwei Luo, Suzanne Hughes, Stephen Wade, Michael Caruana, Karen Canfell, and Dianne L. O'Connell. Statistical projection methods for lung cancer incidence and mortality: A systematic review. *BMJ open*, 9(8):e028497, 2019. ISSN 2044-6055. doi: 10.1136/bmjopen-2018-028497.

J. M. Yuan, J. E. Castelao, M. Gago-Dominguez, M. C. Yu, and R. K. Ross. Tobacco use in relation to renal cell carcinoma. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 7(5):429–433, 1998. ISSN 1055-9965.

documentos de trabajo

Ciencias Sociales
Universidad de la República
URUGUAY

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Facultad de Ciencias Sociales
Universidad de la República
Constituyente 1502 - 2410-6720
comunicacion@cienciassociales.edu.uy
www.cienciassociales.edu.uy