Forecasting, Interventions and Selection

The Benefits of a Causal Mortality Model

Snorre Jallbjørn[1,2] & Søren F. Jarner[1,2] & Niels R. Hansen[1]
2021-08-14, sjb@atp.dk

1. Department of Mathematical Science, University of Copenhagen, Copenhagen, Denmark
2. Danish Labour Market Supplementary Pension Fund, Kongens Vænge 8, Hillerød, Denmark

Agenda

#1 Motivation

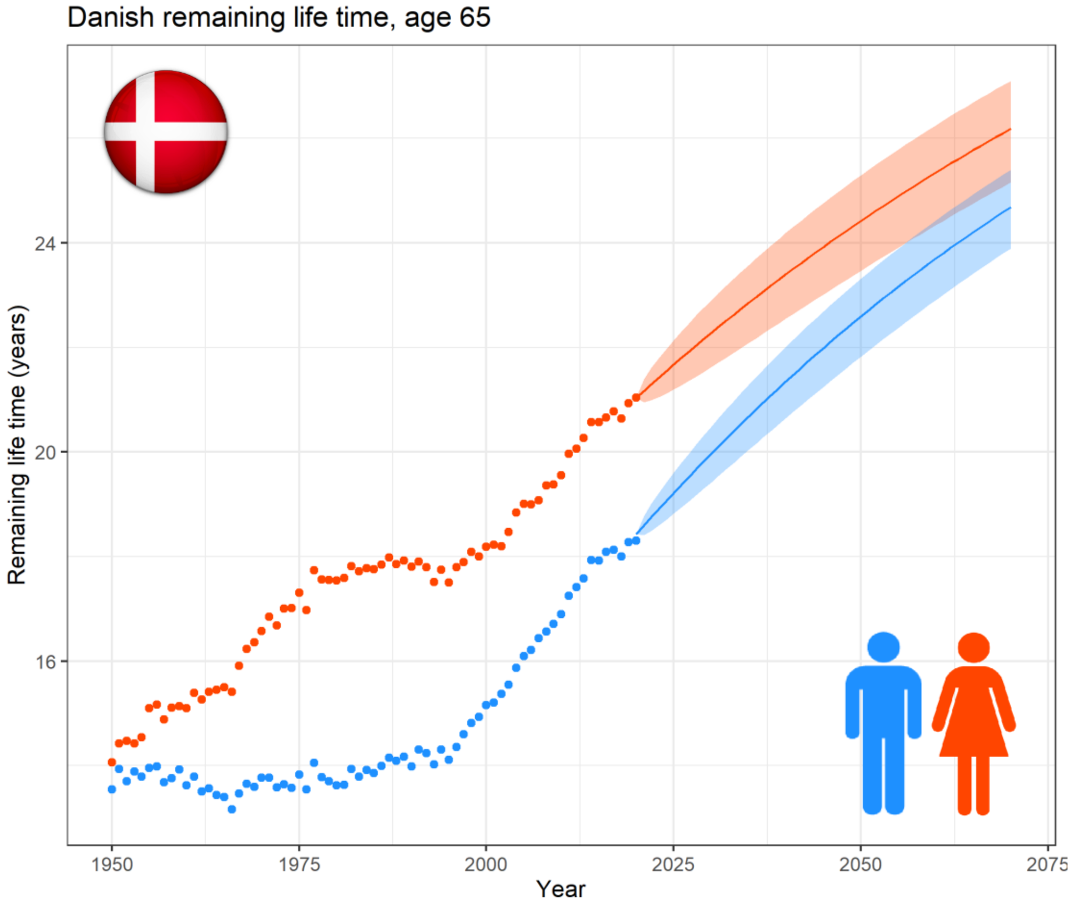
#2 *What-if* scenarios

#3 Framework

#4 Example: Application to US data

Motivation: Communicating longevity risk (1/2)

- **Managing and understanding longevity risk is key!**
  - **Life expectancy is rising**
    - Puts pressure on guaranteed products
  - **We show the Board of Representatives plots like the one on the right …**
    - Use a stochastic mortality model to quantify uncertainty
    - Put a number on longevity risk using e.g. a Value-at-Risk measure
  - **… but do they understand what is going on “under the hood”?**
    - Difficult to engage in a discussion about varying abstract (model) assumptions



Motivation: Communicating longevity risk (2/2)

Objective:

- Formulate scenarios in a straightforward, verbal manner
- Quantify both the direct and indirect effects of changing conditions

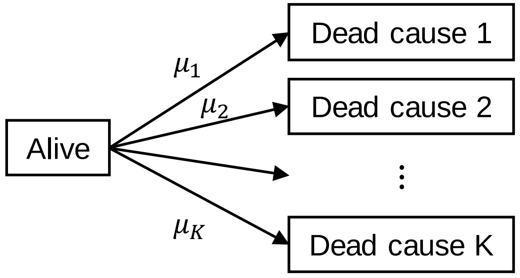
- **Two types of scenarios (interventions)**
  1. Cause of death elimination (or reduction)
  2. Alternative risk prevalence distributions

#2

What-if scenarios

The dominant approach to *what-if* scenarios

- **Competing risk framework**
  - $\mathcal{K} = \{1, \dots, K\}$ mutually exclusive and exhaustive causes of death
  - $\mu_k(x,t)$ the force of mortality for cause *k* at age *x*, time *t.*
  - $\mu(x,t) = \sum_{k\in\mathcal{K}} \mu_k(x,t)$ the force of mortality



- **The dominant approach is to manipulate directly the death rates of interest**
  - E.g., eradicate cause *k* by setting $\mu_k = 0$, but leave remaining rates unaffected
  - Implicitly assumes that those we “save” follow the same pattern of mortality as that observed prior to the intervention
  - Only quantifies the “direct” or “main” effect
  - Formally: stochastic independence of latent cause specific life times

The missing effect: Selection-induced feedback

- **Consider, as an example, eliminating deaths due to lung cancer**
  - Initially, a lot of cancer deaths are avoided …
  - … but the individuals who would have died of lung cancer are typically smokers
  - The progressive buildup of smokers in the population indirectly affects (i.e., increases) the death rates of other smoking attributable causes (at the population level)
  - The initial gain is thus partly offset by increased mortality from other diseases

Conclusion: Need to model selection to quantify the indirect effect of changing conditions

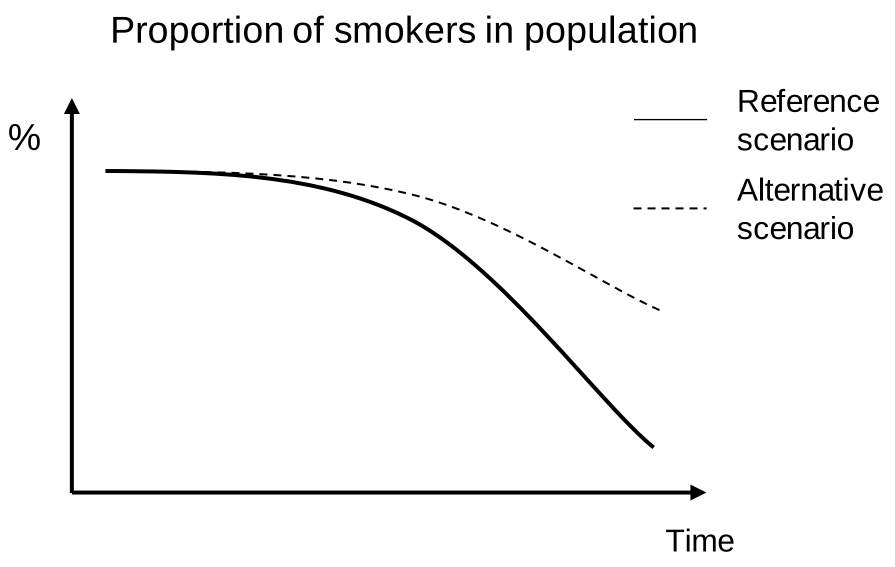
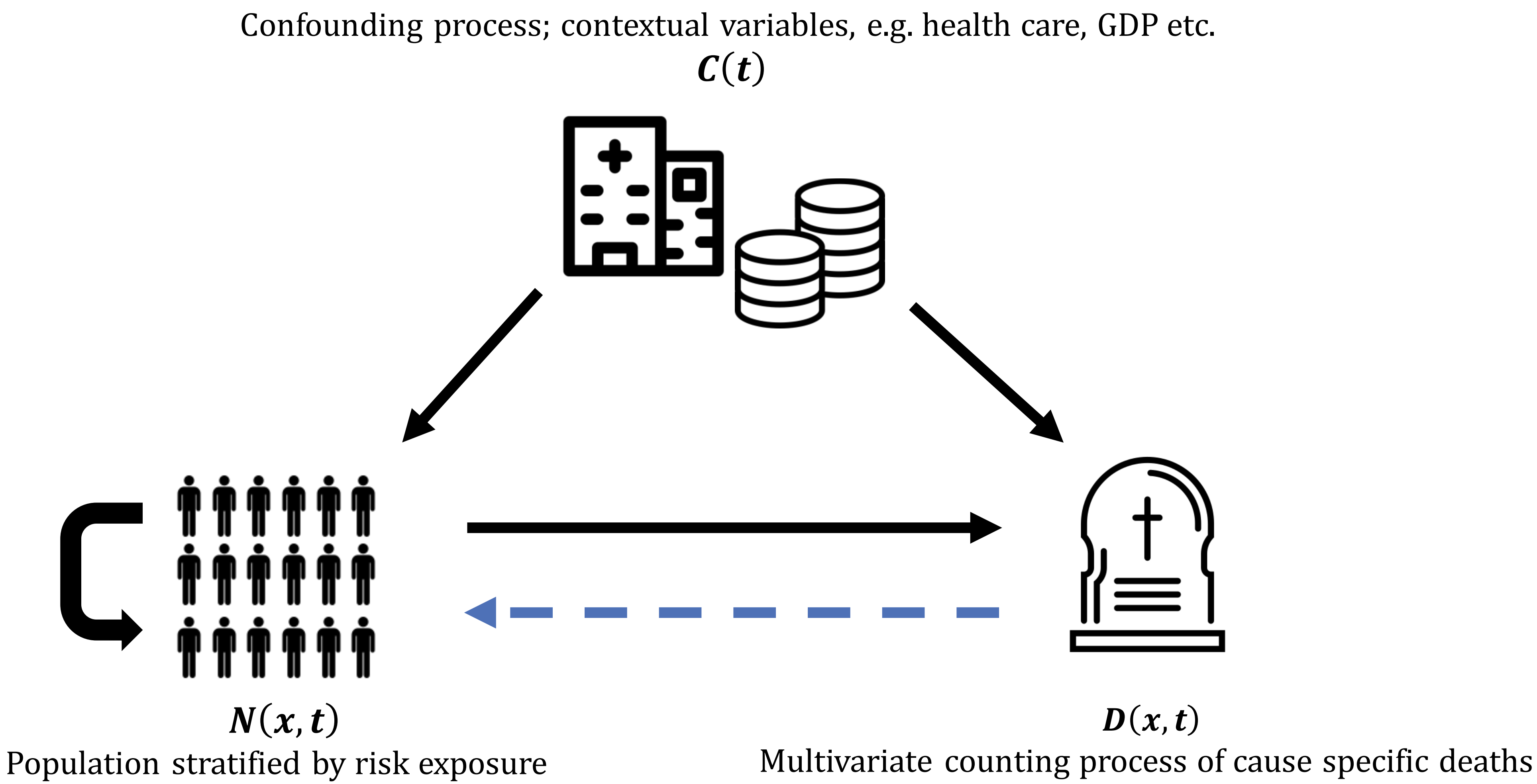


Figure: Stylized example of smoking prevalence following an elimination of lung cancer deaths.

Model and forecasting considerations

- **Want to produce scenarios that conform with real-world implementations of interventions**
  - **The effect of an intervention is a causal query**
    - Relies on a mechanistic understanding of how risk exposure translates to changes in the risk of death
  - **Requires a model embedded with causal relations**
    - The model must describe the link between individual risk factors, e.g. lifestyle choices and health conditions, and cause-specific mortality
- **The conventional projection methodologies for mortality models with covariates are unsuited**
  - **Risk prevalence is typically treated as an exogenous process**
    - Determined and projected outside the mortality model
    - One-sided dependence structure: risk prevalence affects mortality but not vice versa

Local independence graph of a model with and without feedback



#3

Framework

Individual level mortality model

The mathematical framework for formulating assumptions for causal inference is that of potential outcomes

- Let $\bar{z}$ be a possible (deterministic) covariate trajectory
- What would the cause *k* death rate be, had covariate exposure been $\bar{z}$?

$$\mu_k^{\bar{z}}\big(x, t; C(t)\big) = \underbrace{\mu_{0k}\big(\mathrm{x}; \mathrm{C(t)}\big)}_{\text{Baseline rate}} \underbrace{\exp\Big(\beta_k^{\top}(x) z(x,t)\Big)}_{\text{Relative risk}}$$

where

- $z(x,t) \in \mathbb{R}^{\mathrm{V}}$ is the individual's covariates
- $\beta_k(x) \in \mathbb{R}^{\mathrm{V}}$ are age-specific coefficients
- $C(t)$ is a confounding process (contextual variables, e.g. health care, GDP, etc.)
  - Let $C(t) = t$ such that calendar time acts as a *surrogate* confounder for various unmeasured factors that change over time.

Population level mortality model

Trace out the consequences of changing conditions for the population level death rate

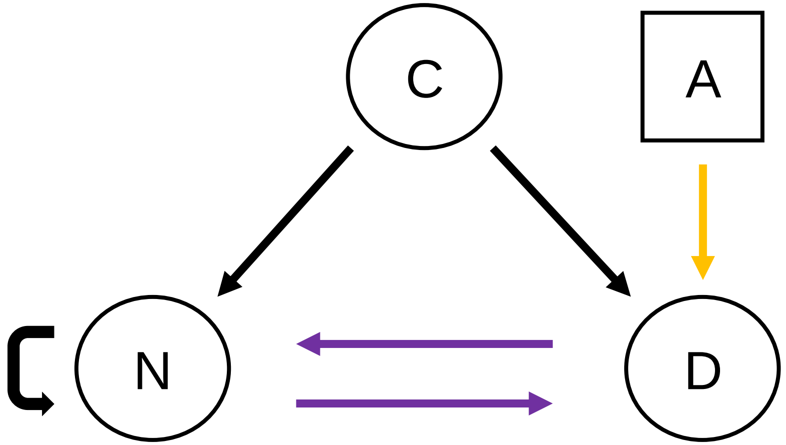
- In practice, risk factors are typically reported as categorical variables, even when the underlying exposure is continuous
  - Let Z be multinomial with G different covariate configurations
  - Let $\Pi^{\mathcal{K}}(x,t) = \left(\pi_1^{\mathcal{K}}(x,t), \dots, \pi_G^{\mathcal{K}}(x,t)\right)^{\top} \in \{p \in \mathbb{R}_+^G \mid \sum_g p_g = 1\}$ be the risk factor composition
  - Let $R_{kg}(x)$ denote the relative risk from cause k in group g.
- At the population level, the aggregated death rate becomes

$$\mu^{\mathcal{K},\Pi^{\mathcal{K}}}(x,t) = \sum_{k \in \mathcal{K}} \mu_{0k}(x,t) \sum_{g=1}^{G} \pi_g^{\mathcal{K}}(x,t) R_{kg}(x)$$

Interpretation: $\mu^{\mathcal{K},\Pi^{\mathcal{K}}}$ is the death rate in a world where risk prevalence is $\Pi^{\mathcal{K}}$ and causes $\mathcal{K}$ are operating.

Cause elimination: Decomposition of effects

- **Cause of death elimination**
  - Consider a forecast with only causes $\mathcal{K}^* \subset \mathcal{K} = \{1, \dots, K\}$ operating
  - We can think of this action (A) as an intervention in the graph



The total effect is

$$\begin{aligned}\text{TE} &= \mu^{\mathcal{K},\Pi^{\mathcal{K}}} - \mu^{\mathcal{K}^*,\Pi^{\mathcal{K}^*}} \\ &= \underbrace{\mu^{\mathcal{K},\Pi^{\mathcal{K}}} - \mu^{\mathcal{K}^*,\Pi^{\mathcal{K}}}}_{\textbf{DE}} + \underbrace{\mu^{\mathcal{K}^*,\Pi^{\mathcal{K}}} - \mu^{\mathcal{K}^*,\Pi^{\mathcal{K}^*}}}_{\textbf{IE}}\end{aligned}$$

with **direct** and **indirect** effects

$$\textbf{DE}(x,t) = \sum_{k \in \mathcal{K} \setminus \mathcal{K}^*} \mu_{0k}(x,t) \sum_{g=1}^{G} \pi_g^{\mathcal{K}}(x,t) R_{kg}(x)$$

$$\textbf{IE}(x,t) = \sum_{k \in \mathcal{K} \setminus \mathcal{K}^*} \mu_{0k}(x,t) \sum_{g=1}^{G} \left[\pi_g^{\mathcal{K}}(x,t) - \pi_g^{\mathcal{K}^*}(x,t)\right] R_{kg}(x)$$

#4

Example: US data

US data example (1/5): Setup

- **Analysis restricted to two major lifestyle related risk factors: cigarette consumption and obesity**
  - Investigation period: 1999-2018
  - Age range: 20-84

- **Data sources**
  - Relative risks and risk-outcome pairs: Global Burden of Disease Initiative (Murray et al., 2020)
  - Cause-specific occurence-exposure data: CDC WONDER database
  - Risk prevalence: IPUMS – National Health Interview Survery database.

US data example (2/5): Risk prevalence from IPUMS

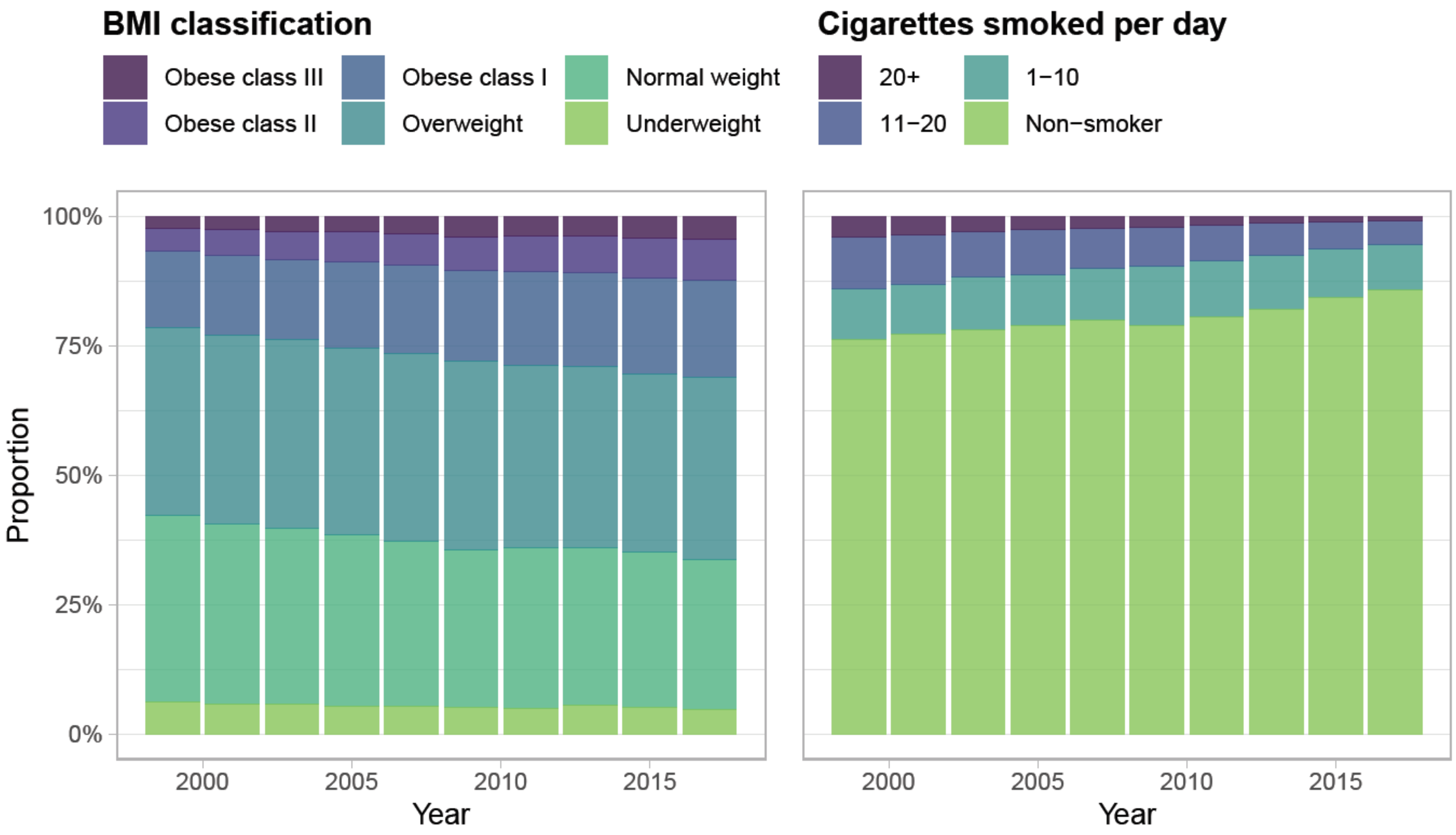
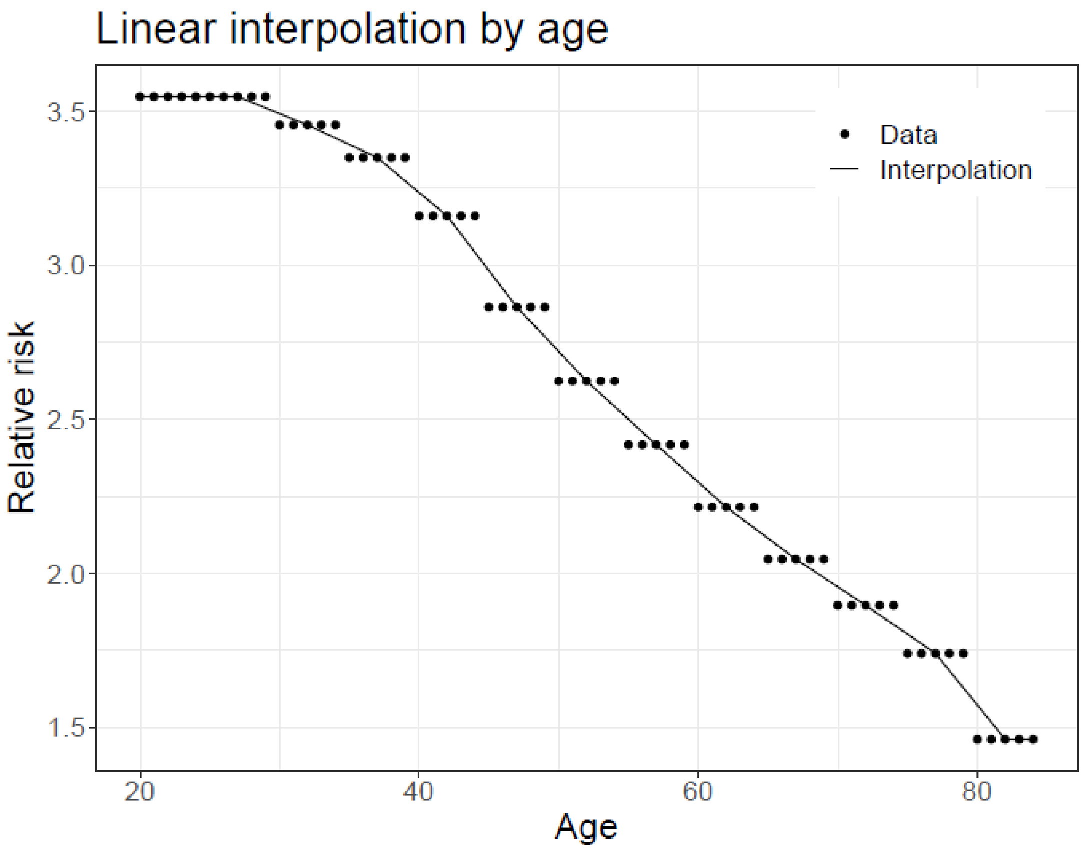
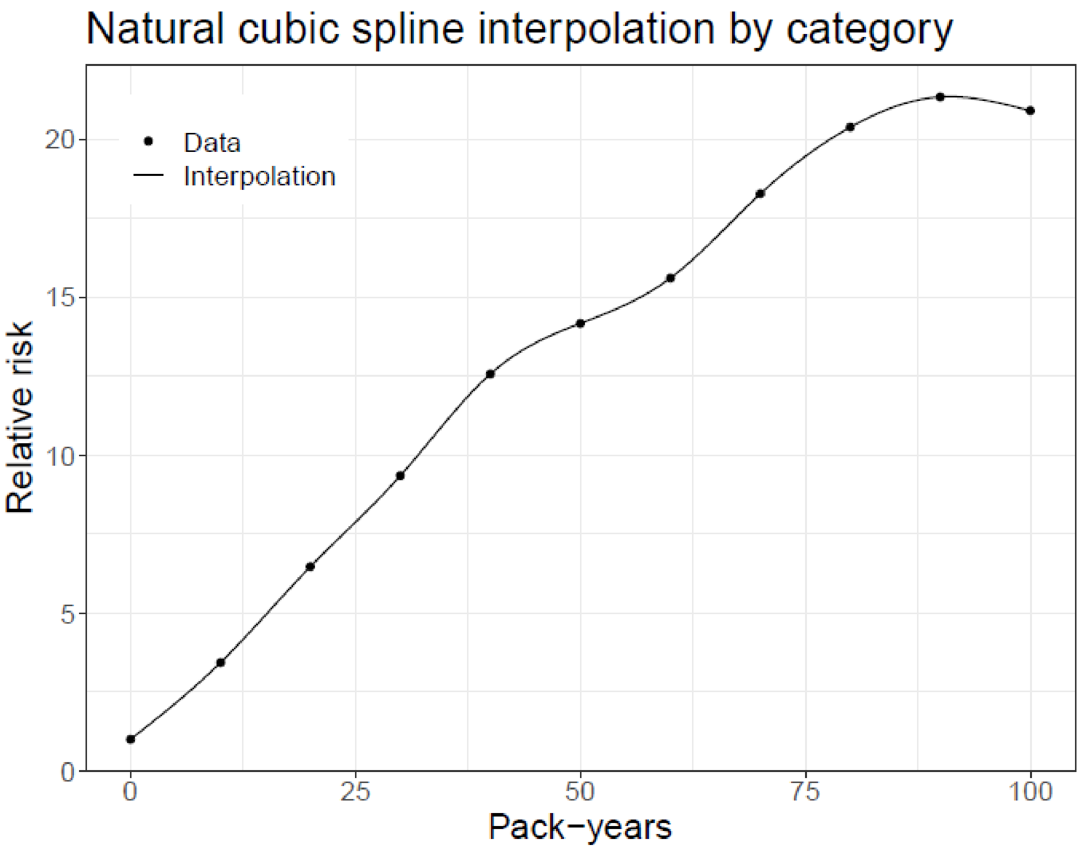


Figure: U.S. risk proportions of BMI (left) and smoking (right) based on aggregated IPUMS data.

US data example (3/5): Relative risks from the Global Burden of Disease initiative (Murray et al., 2020)





Left panel: Female relative risk for diabetes mellitus by age (dots) using linear interpolation with age-bucket centroids as fixed-points (solid line).

Right panel: Age 60 female relative risk for tracheal, bronchus, and lung cancer by pack-years (dots) using natural cubic spline interpolation (solid line).

US data example (4/5): Cause specific rates from CDC WONDER

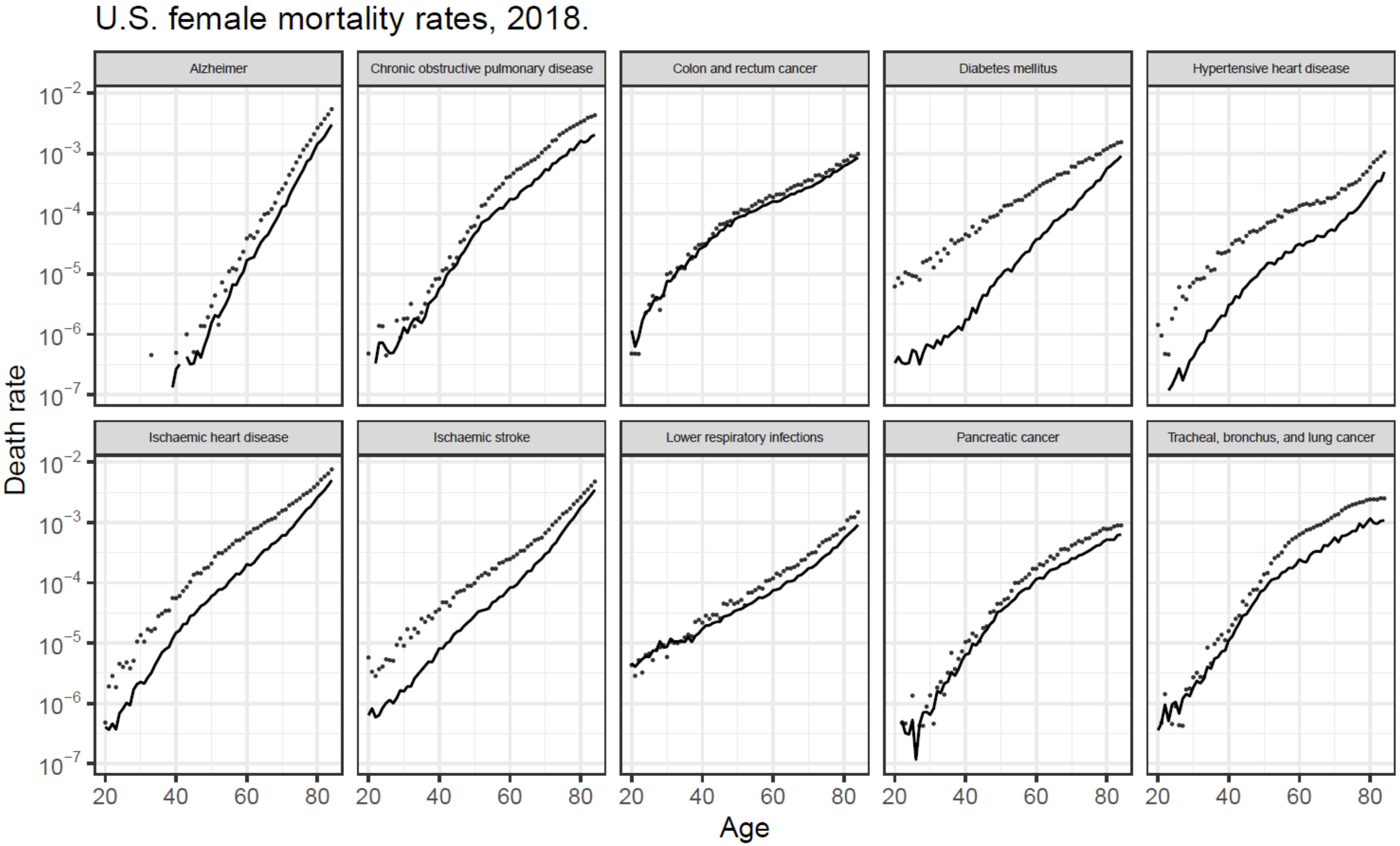


Figure: Empirical (dotted) and baseline (solid) death rates for select leading causes of death affected by smoking and BMI. Baseline model: $\mu_{0k}(x,t) = \exp(\alpha_x + \beta_x t)$.

US data example (5/5): Cure for cancer

Question:

- How quickly will those we save from cancer die of something else?
- And what will they die from instead?

(within 25 years)

Data:

- Population: US females, cohort of 1958
- Risk factors included: Smoking and obesity

Change in cumulative incidence following redistribution of deaths

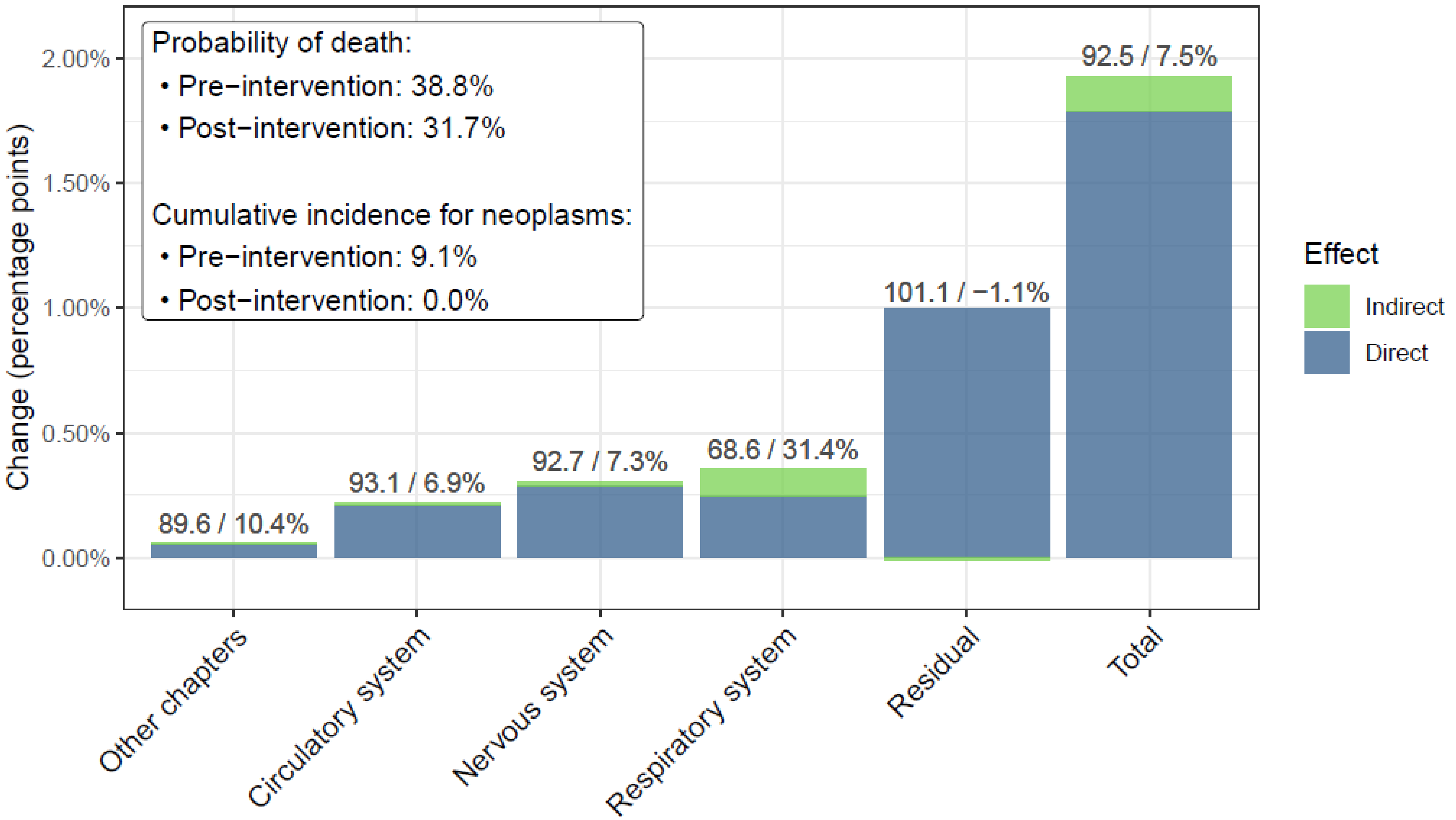


Figure: Change to the cumulative incidence of select disease chapters following an elimination of deaths due to neoplasms caused by smoking and/or obesity.

Conclusions

- **Mortality models with integrated epidemiological information are still in their infancy**
  - Availability and quality of detailed data on risk factors continues to grow
  - Opens the possibility of assessing the impact of preventive measures that reduce disease risk

- **Any model that relies on future risk prevalence to evaluate mortality is fundamentally challenging to forecast**
  - For the model to relay selection-induced effects, risk prevalence must be treated as an intrinsic part of the mortality projection

- **Naïve cause eradication procedures are likely too generous in their estimate of mortality reduction**
  - Removing one or more causes weakens the selection mechanism and results in an accumulation of high-risk individuals in the population

References

CDC WONDER (2020). Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2018 on CDC WONDER Online Database, released in 2020. http://wonder.cdc.gov/ucd-icd10.html.

IPUMS (2019). Blewett, Lynn A. and Drew, Julia A. R. and King, Miriam L. and Williams, Kari C.W. IPUMS Health Surveys: National Health Interview Survey, Version 6.4 [dataset]. Minneapolis, MN: IPUMS, 2019. https://www.nhis.ipums.org.

Jallbjørn, S. and Jarner, Søren F. and Hansen, Niels R. (2021). Forecasting, interventions and selection: The benefits of a causal mortality model. *Working paper*, pp. 1-24.

Murray, C. J., et al. (2020). 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.

Thank you!

Questions or comments?