

Implementing the G-methods in neurocognitive diseases

L. Paloma Rojas Saunero

PhD Student

Promotor: Arfan Ikram **Co-promotor:** Sonja Swanson

Epidemiology Department

Causal Inference Group - NeuroEpi Group



What are the G-methods?

G-methods

- G methods are a family of methods that include the g-formula, marginal structural models, and structural nested models.
- They provide consistent estimates of contrasts of average potential outcomes under a less restrictive set of identification conditions than standard regression methods.
- Specifically, standard regression requires no feedback between time-varying treatments and time-varying confounders, while G-methods do not.

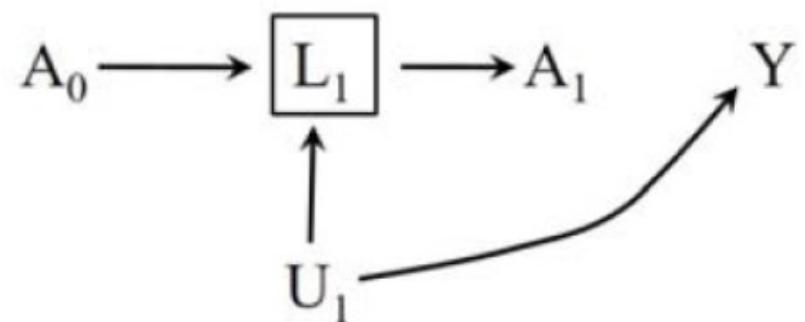
Time-fixed interventions

- One dose vaccine
- Surgery
- Traumatic brain injury



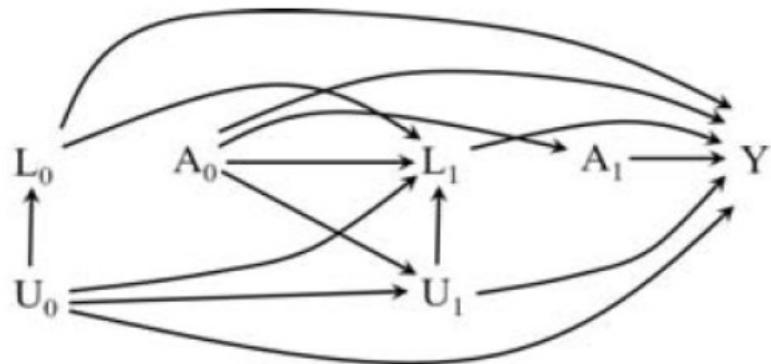
Time-varying interventions

- Medication
- Diet
- Smoking
- Exercise



- A (exposure), L (confounders), Y (outcome), U (unmeasured confounders).
- DAG from Hernan MA, Robins JM (2019). Causal Inference.

Randomized controlled trials (RCT's)



Intention to treat (ITT) effect:

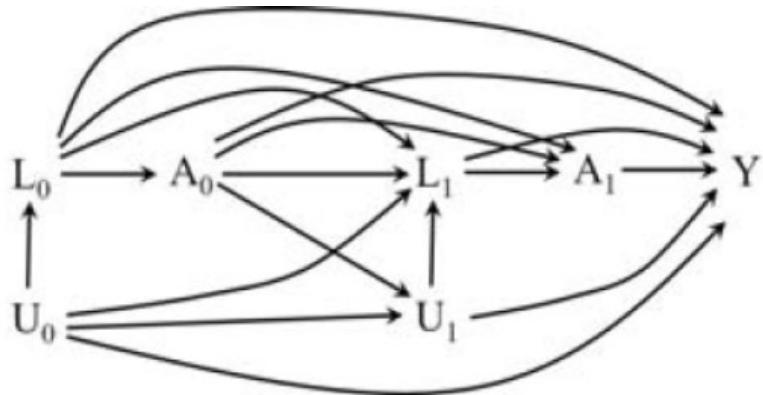
What is the effect of randomization?

Per protocol (PP) effect:

What is the effect of the intervention had everyone followed the strategy over time?

- A (exposure), L (confounders), Y (outcome), U (unmeasured confounders).
- DAG from Hernan MA, Robins JM (2019). Causal Inference.

Observational studies



- Cohorts, electronic clinical records, registries
- Under certain assumptions we could estimate the causal effect of an exposure as if we were conducting an RCT.

- A (exposure), L (confounders), Y (outcome), U (unmeasured confounders).
- DAG from Hernan MA, Robins JM (2019). Causal Inference.

Hypothetical interventions on systolic blood pressure and smoking to reduce the risk of stroke and dementia

Motivation

- Stroke and dementia have a important burden in public health and share risk factors.
- Trials have focused on selected group of patients, have limited follow up and focus primarily on ITT.
- There is limited evidence on the **sustained effect** of *joint* interventions in systolic blood pressure and quitting smoking

Methods

- Participants from Rotterdam Study I, recruited between 1990-1993 and followed during 1993-1995, 1997-1999 and 2002-2005 (n = 7983)
- Eligibility criteria:
 - Persons below 80 years old
 - No prior history of stroke and no prior history of dementia diagnosis or MMSE below 26
 - Complete information at baseline
 - Final sample size of **5113** participants

Hypothetical interventions

- Natural course
- Maintain systolic blood pressure (SBP) < 120 mmHg
- Maintain SBP < 140 mmHg
- Reduce SBP by 10% if > 140
- Reduce SBP by 20% if > 140
- Quit smoking (if current smoker)
- Joint interventions of lowering SBP and quit smoking

Outcome and Follow up

- Stroke
- Dementia
- Follow up: 15 years since baseline
- Lost to follow up is a **censoring** event (*skipped visits are part of lost to follow up*)
- Death is treated as a **competing** event

Covariates

Baseline covariates:

- Age
- Education
- History of diabetes
- History of heart disease
- Baseline Systolic blood pressure

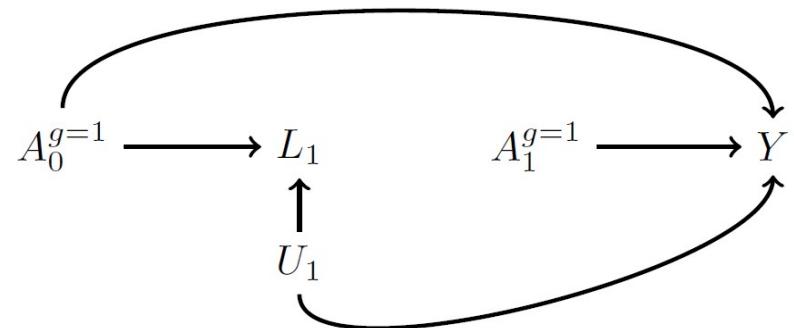
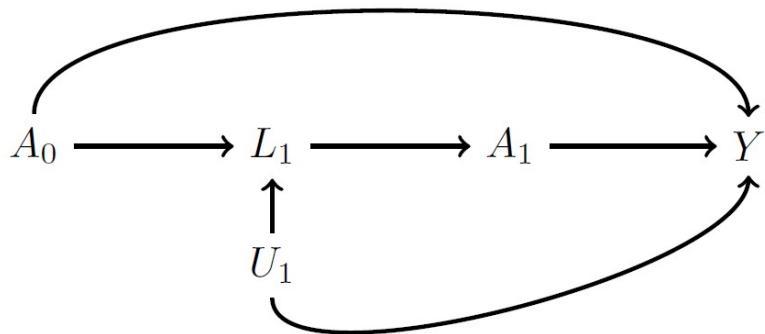
Time-varying covariates:

- Visit process
- Systolic blood pressure
- Body mass index
- Cigarette smoking
- Alcohol intake (g/day)
- Cholesterol
- Hypertensive medications
- Development of:
 - Diabetes
 - Heart disease
 - Cancer
 - Stroke/Dementia

Analysis

What would have happened had everyone was randomized and had adhered to each intervention (g)?

- Parametric G-formula
- Allows presence of measured time-varying confounders (in addition to baseline confounding). Measured time-varying confounders are themselves affected by past exposure



- A (exposure), L (confounders), Y (outcome)
- $g=1$: hypothetical intervention. (ex. maintain blood pressure < 140)

Parametric G-formula

1. Model each variable (A, L, Y, D) using the covariate history.
2. Select a random sample from the cohort and simulate covariate and treatment history using estimated coefficients from step 1 using Monte Carlo simulation.
3. Intervene by assigning the treatment value at each timepoint according to the hypothetical intervention.
4. Estimate the predicted probability of the outcome based on the updated intervention
5. Calculate the average of the subject-specific risks and bootstrap CI.
6. Repeat steps 2-5 for each hypothetical intervention.

- A (exposure), L (confounders), Y (outcome), D (competing event)

Assumptions

Identifiability assumptions

- Exchangeability: *No unmeasured confounding*
- Positivity: $\Pr(A|L > 0)$
- Consistency: *Well-defined intervention*

Modeling assumptions

- No model misspecification

The pitfalls

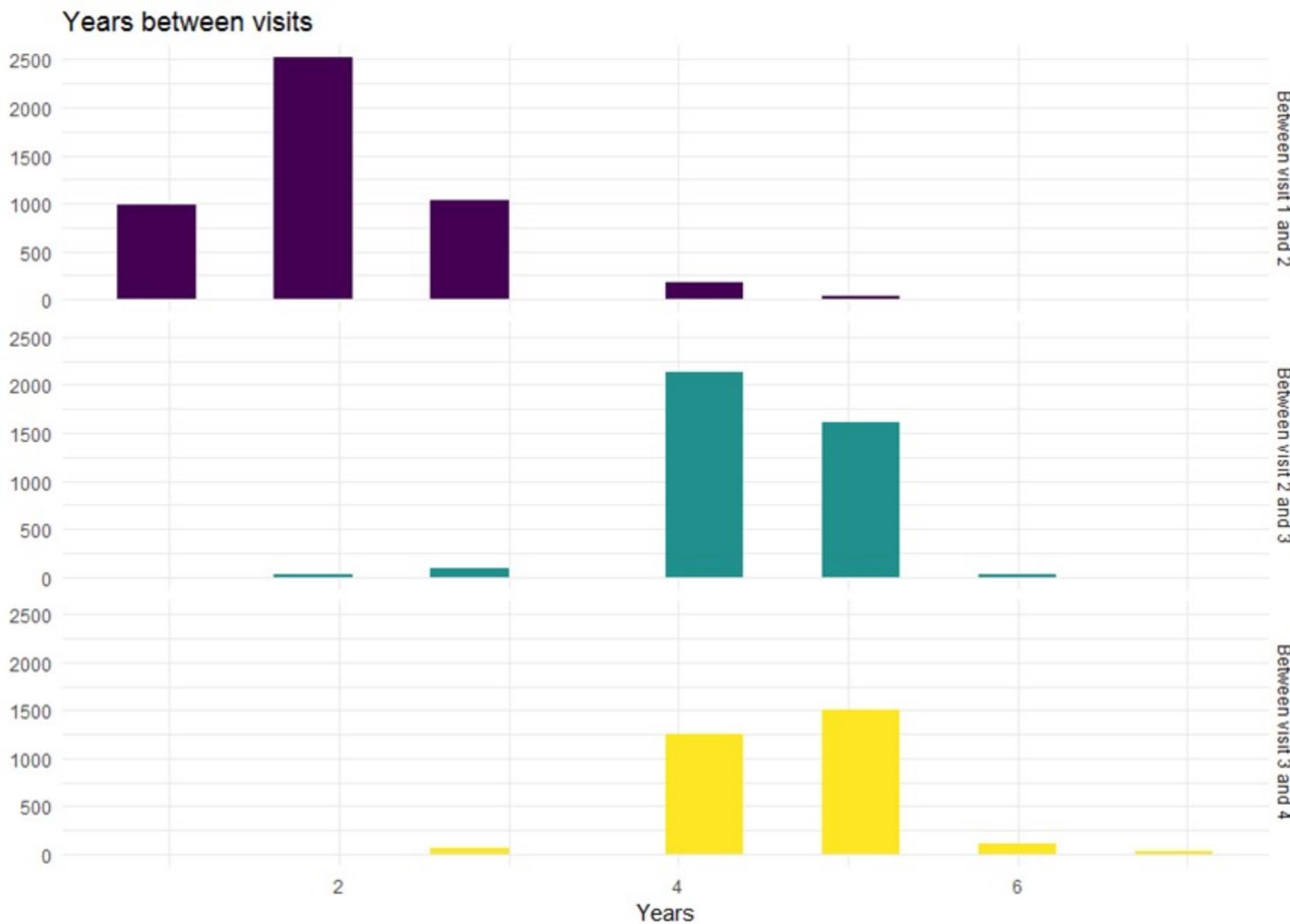
Time



- Intervals between *visits* are not symmetric.



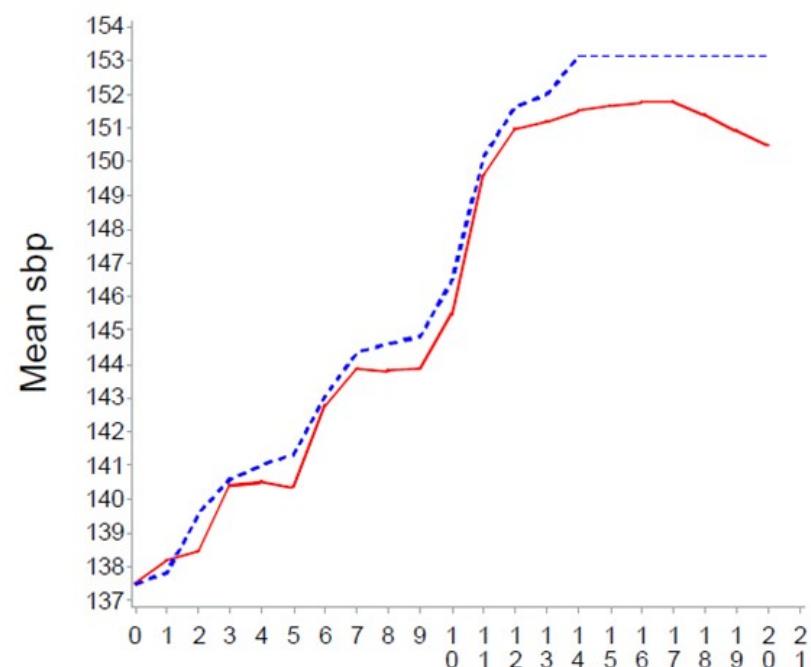
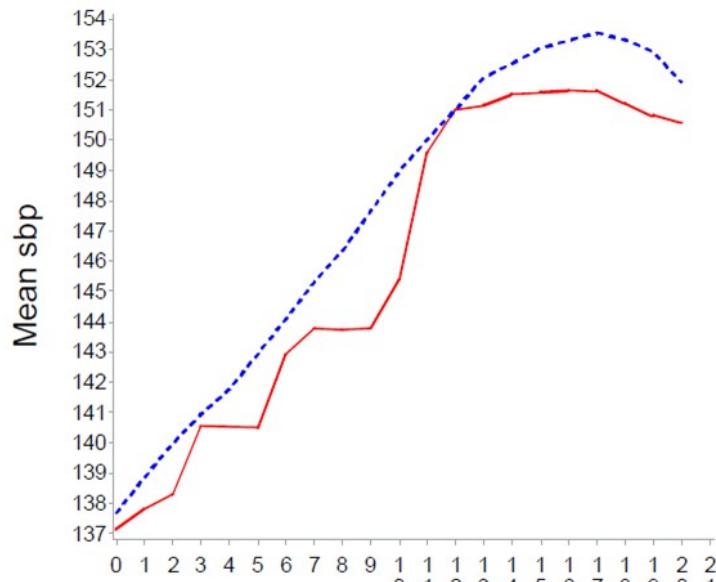
- Interval within 2 visits are not symmetric across *individuals*.



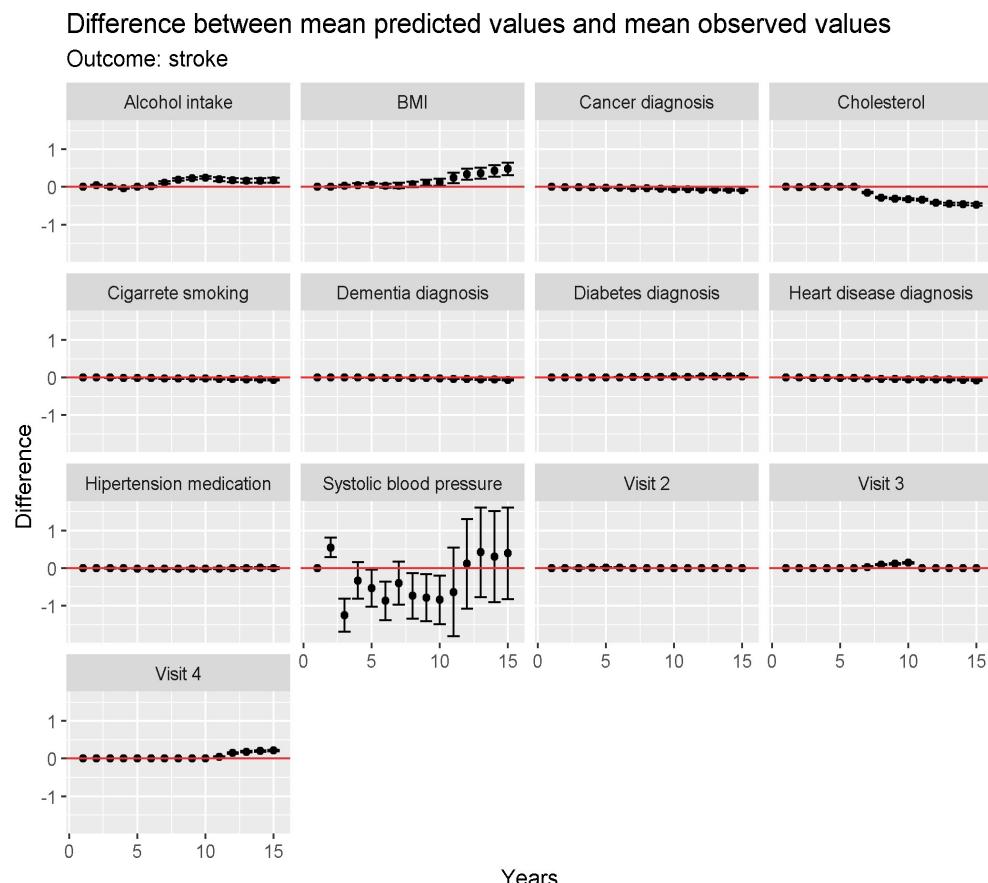
- Some covariates are measured in during the visits, some are captured from clinical records (dates of diagnose).

- Solution:

- Model the visit process as covariates and simulate the visit process.
- Model covariates conditioned on the visit process.



Assessing modeling



Results: Stroke

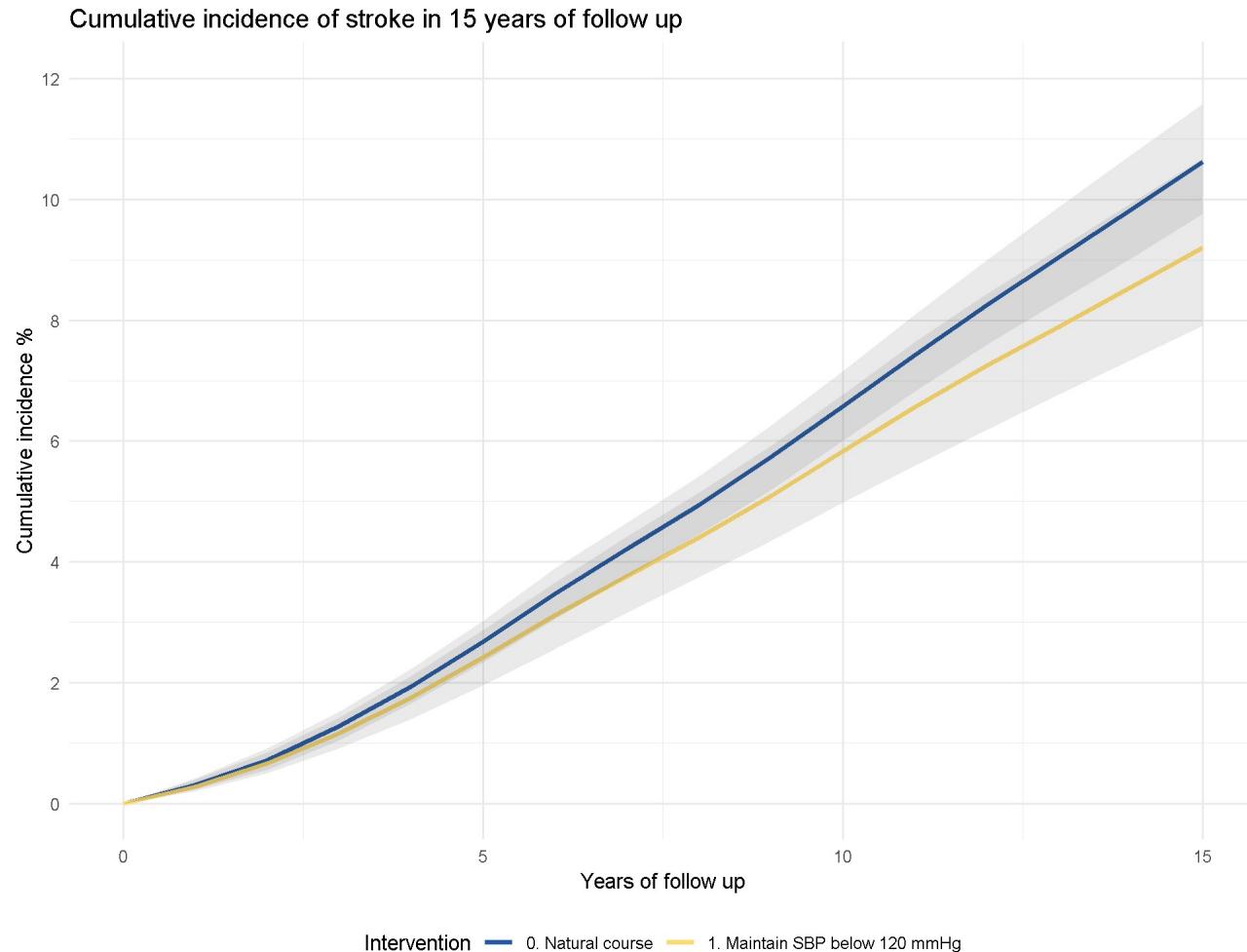
The observed risk at 15 years is **10.7%**.

Intervention	Risk	Risk Ratio	Risk Difference	Total Intervened %
0. Natural course	10.6 (9.6,11.8)	1 (1, 1)	0 (0, 0)	0.0
1. Maintain SBP below 120 mmHg	9.2 (7.7,11.1)	0.9 (0.7, 1)	-1.4 (-2.8, 0.1)	97.9
2. Maintain SBP below 140 mmHg	9.5 (8.3,10.7)	0.9 (0.8, 1)	-1.1 (-2, -0.5)	83.5
3. Reduce SPB by 10% if above 140	9.3 (8.1,10.6)	0.9 (0.8, 1)	-1.3 (-2.3, -0.4)	83.5
4. Reduce SPB by 20% if above 140	9.3 (8,10.9)	0.9 (0.8, 1)	-1.3 (-2.5, -0.2)	83.5
5. Quit smoking (if current smoker)	9.9 (8.8,11.2)	0.9 (0.9, 1)	-0.8 (-1.3, -0.3)	26.5
6. Joint 1 + 5	8.6 (7.1,10.5)	0.8 (0.7, 0.9)	-2.1 (-3.5, -0.6)	99.1
7. Joint 2 + 5	8.7 (7.6,10.1)	0.8 (0.7, 0.9)	-1.9 (-2.8, -1.1)	88.6
8. Joint 3 + 5	8.7 (7.5,10.1)	0.8 (0.7, 0.9)	-2 (-3, -1.1)	88.8
9. Joint 4 + 5	8.6 (7.3,10.3)	0.8 (0.7, 0.9)	-2 (-3.2, -0.8)	88.8

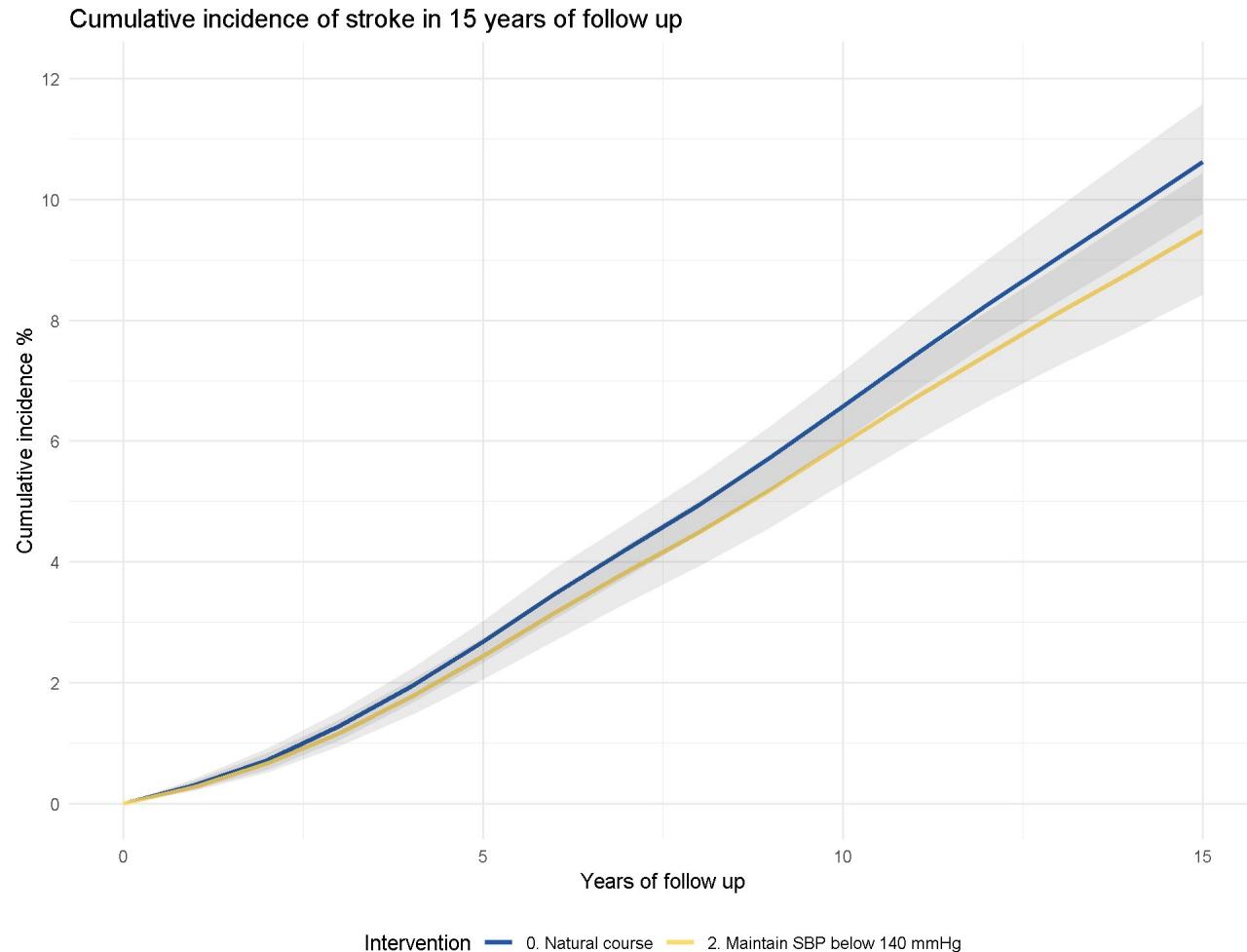
Note:

Results with 500 bootstraps.

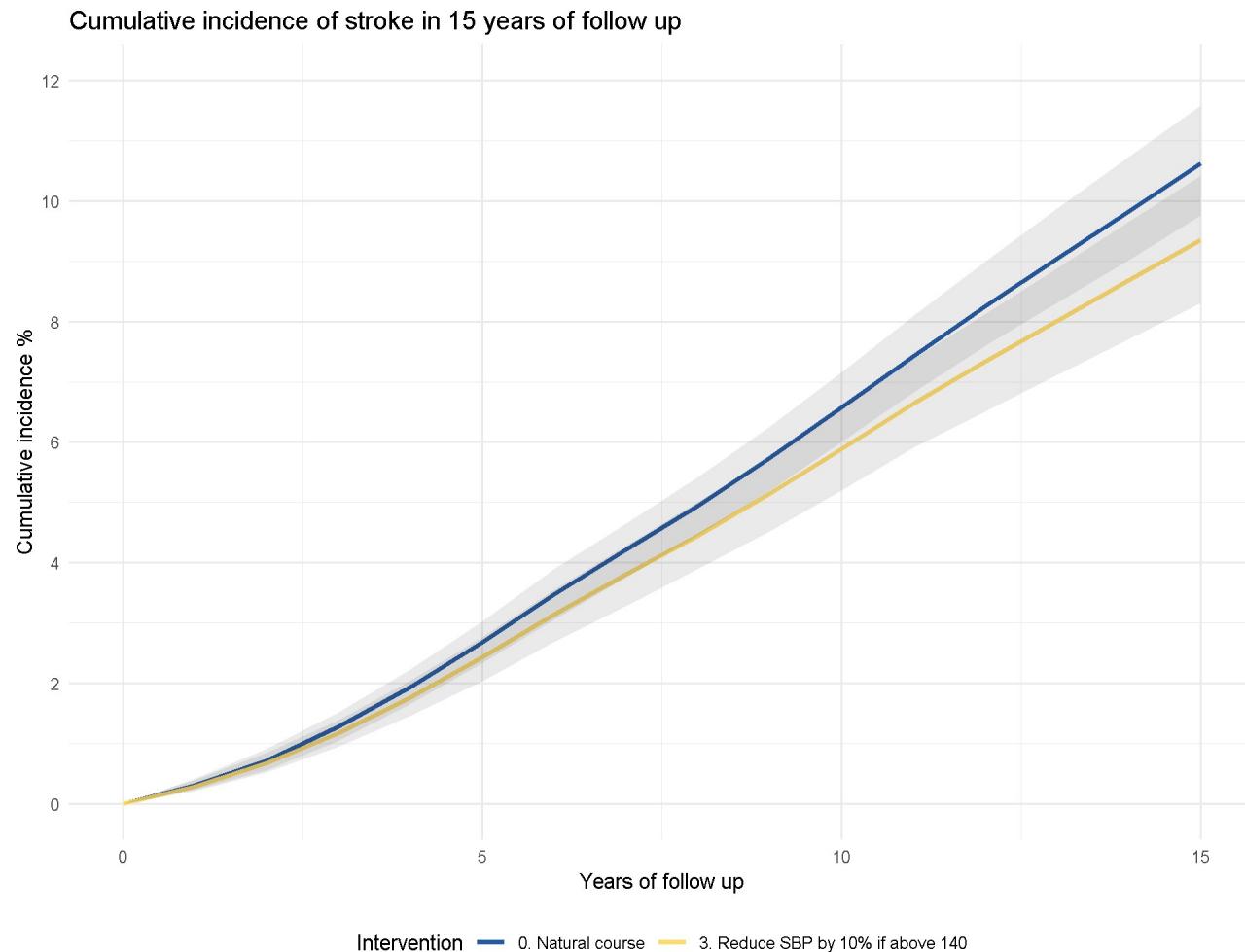
Results: Stroke



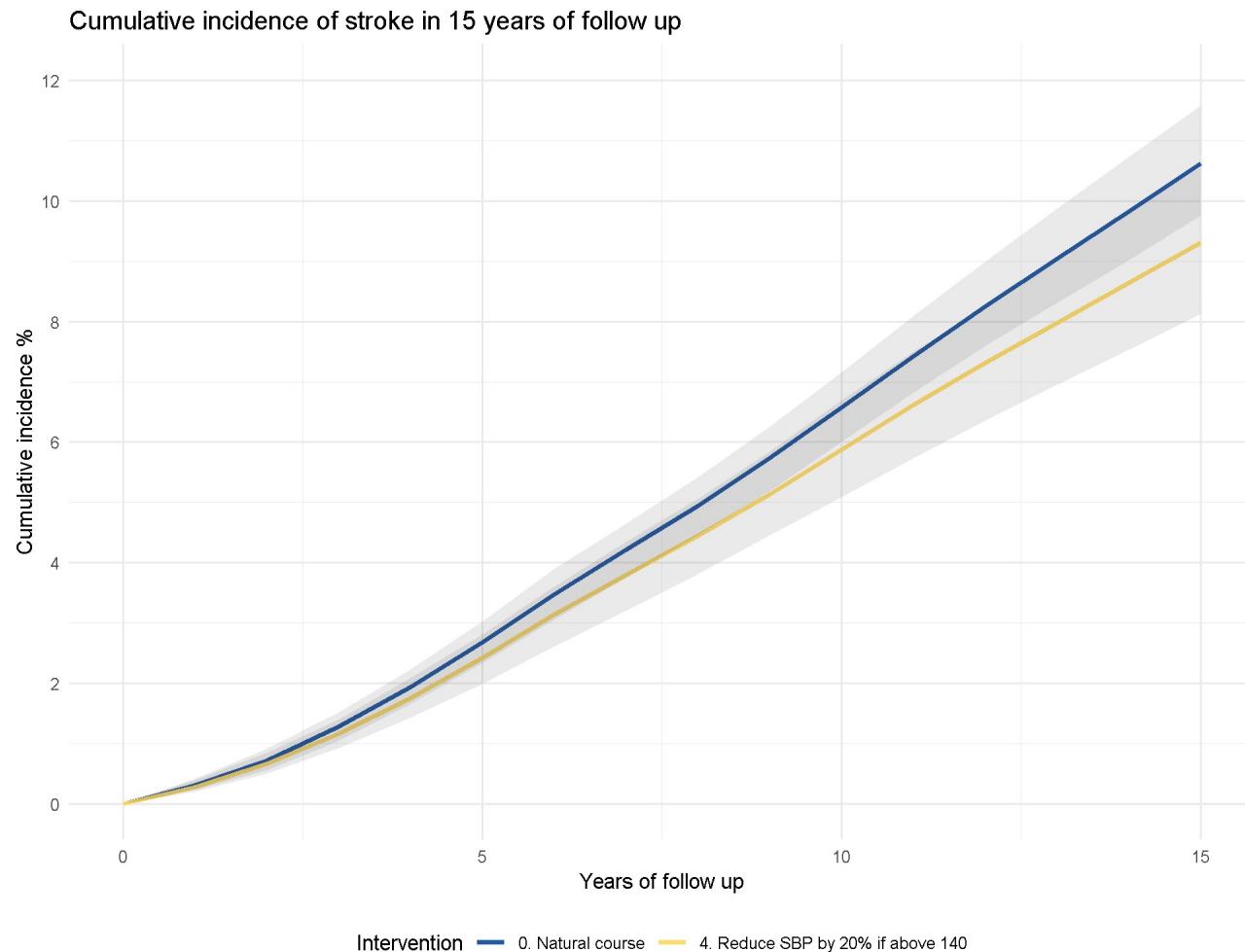
Results: Stroke



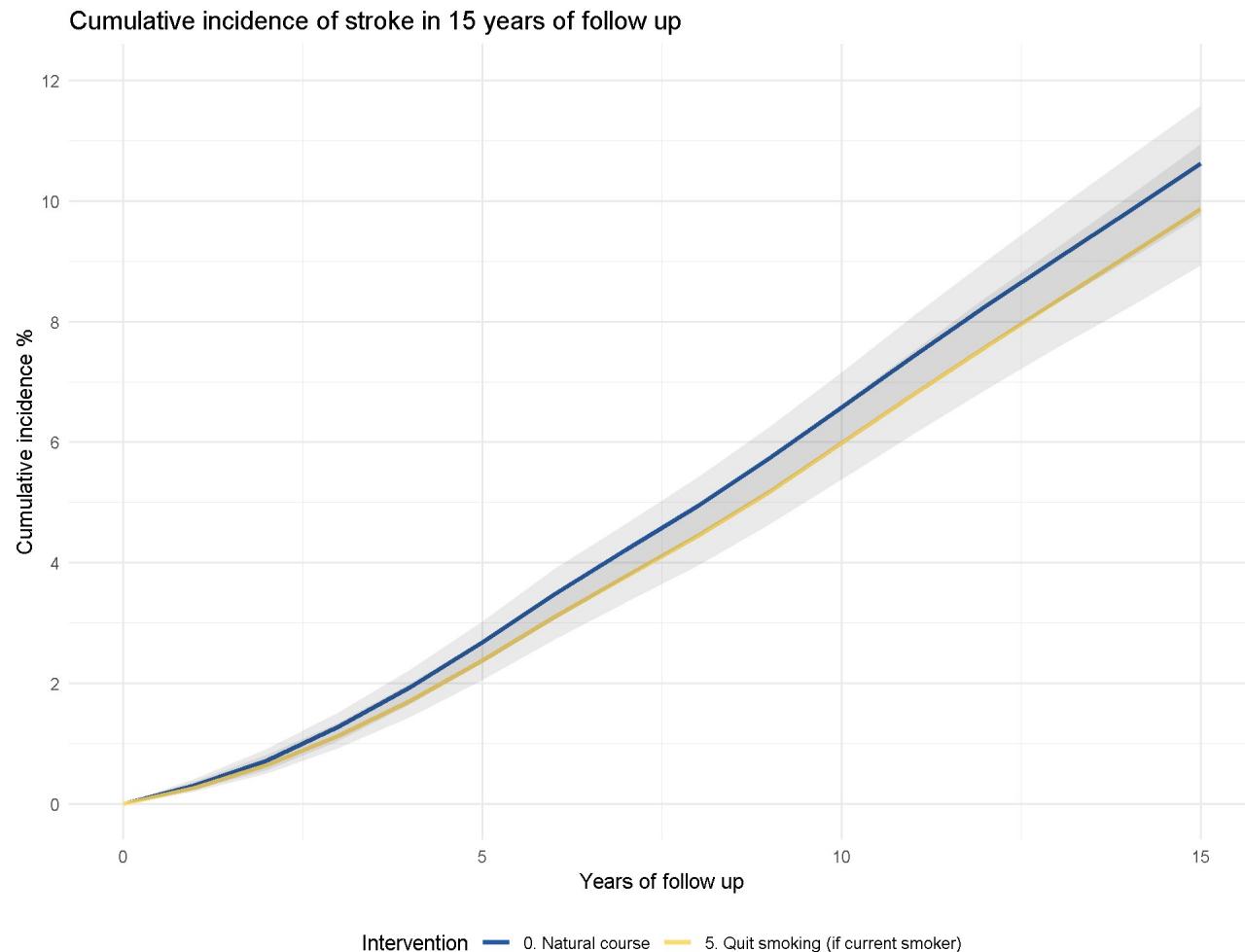
Results: Stroke



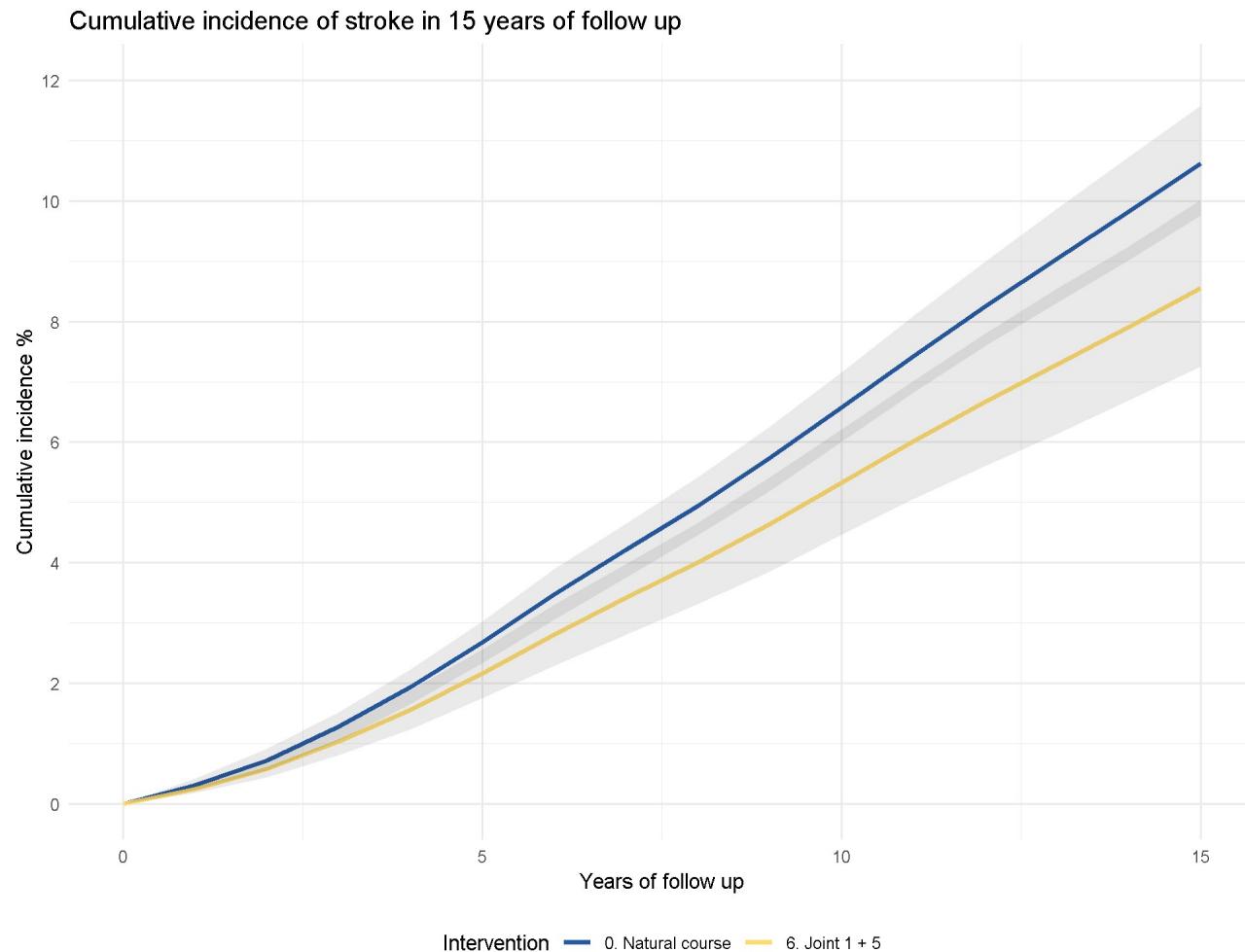
Results: Stroke



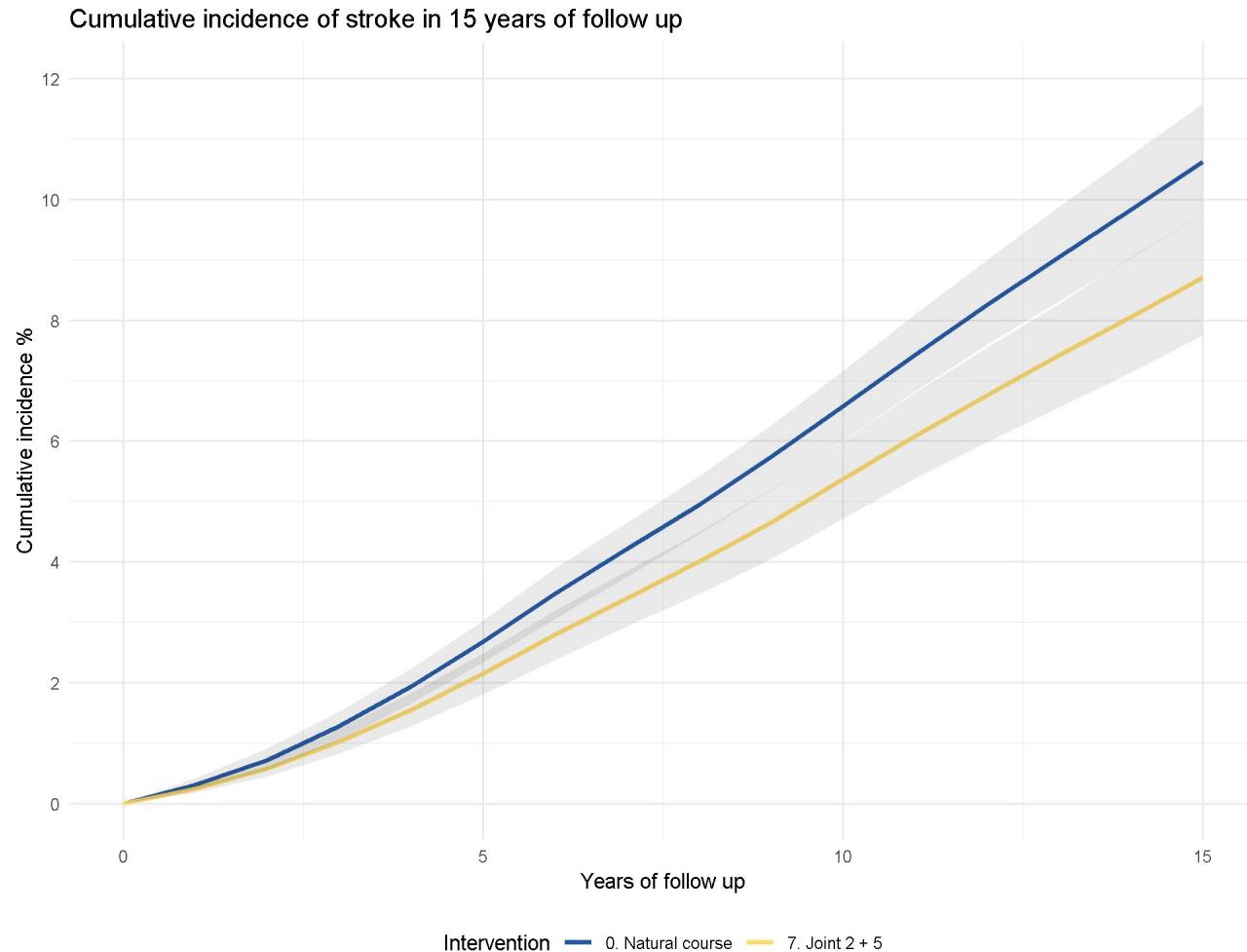
Results: Stroke



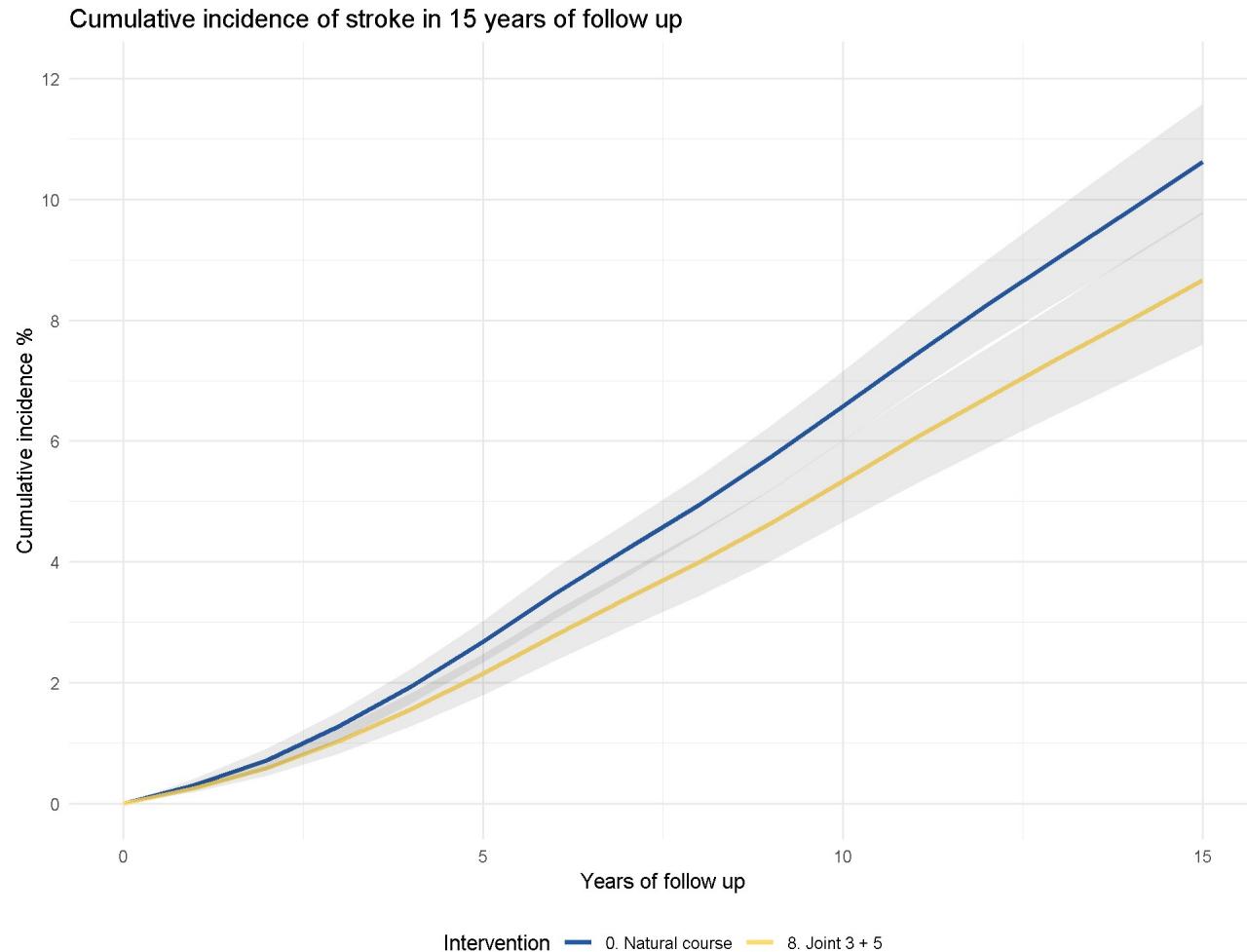
Results: Stroke



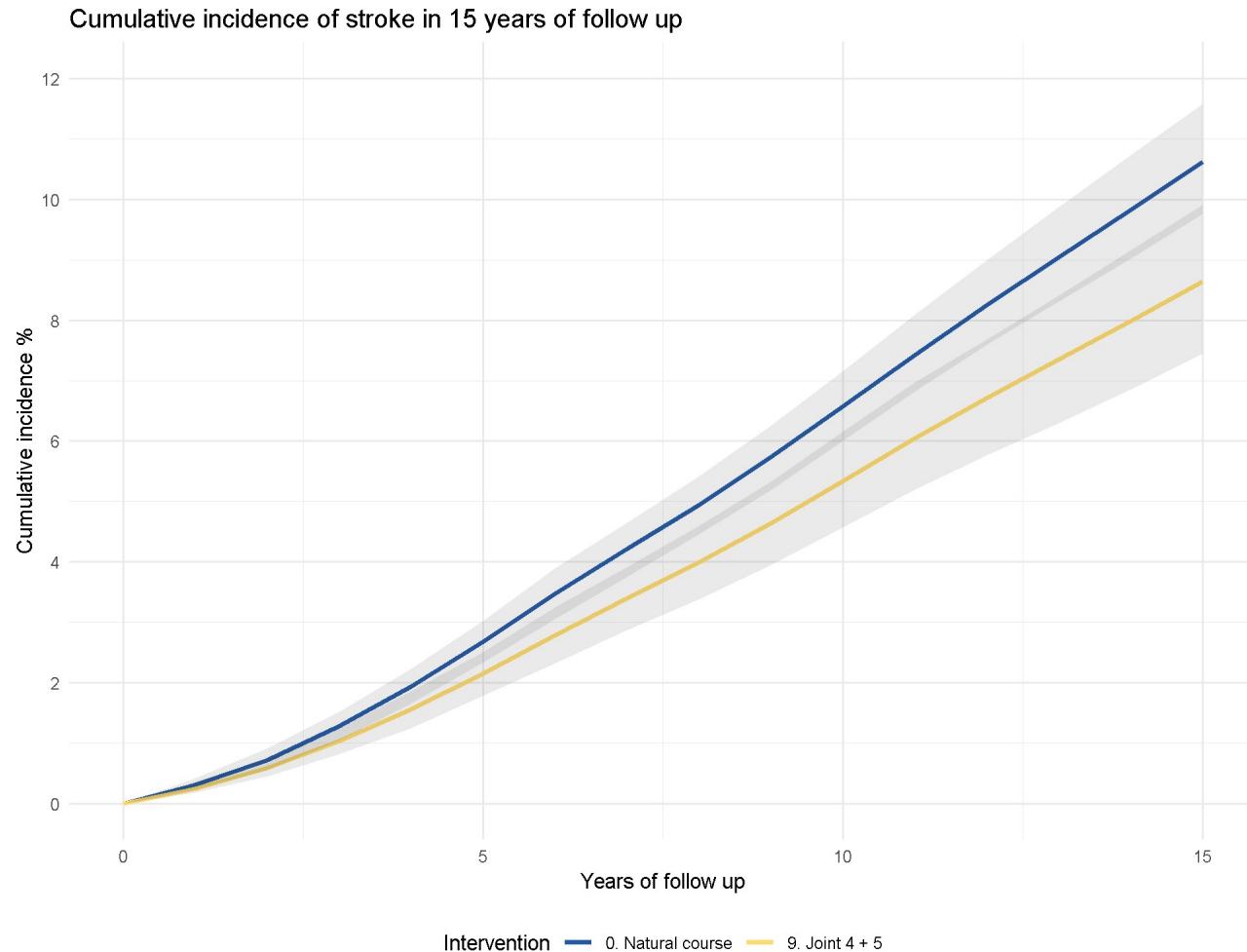
Results: Stroke



Results: Stroke



Results: Stroke



Results: Dementia

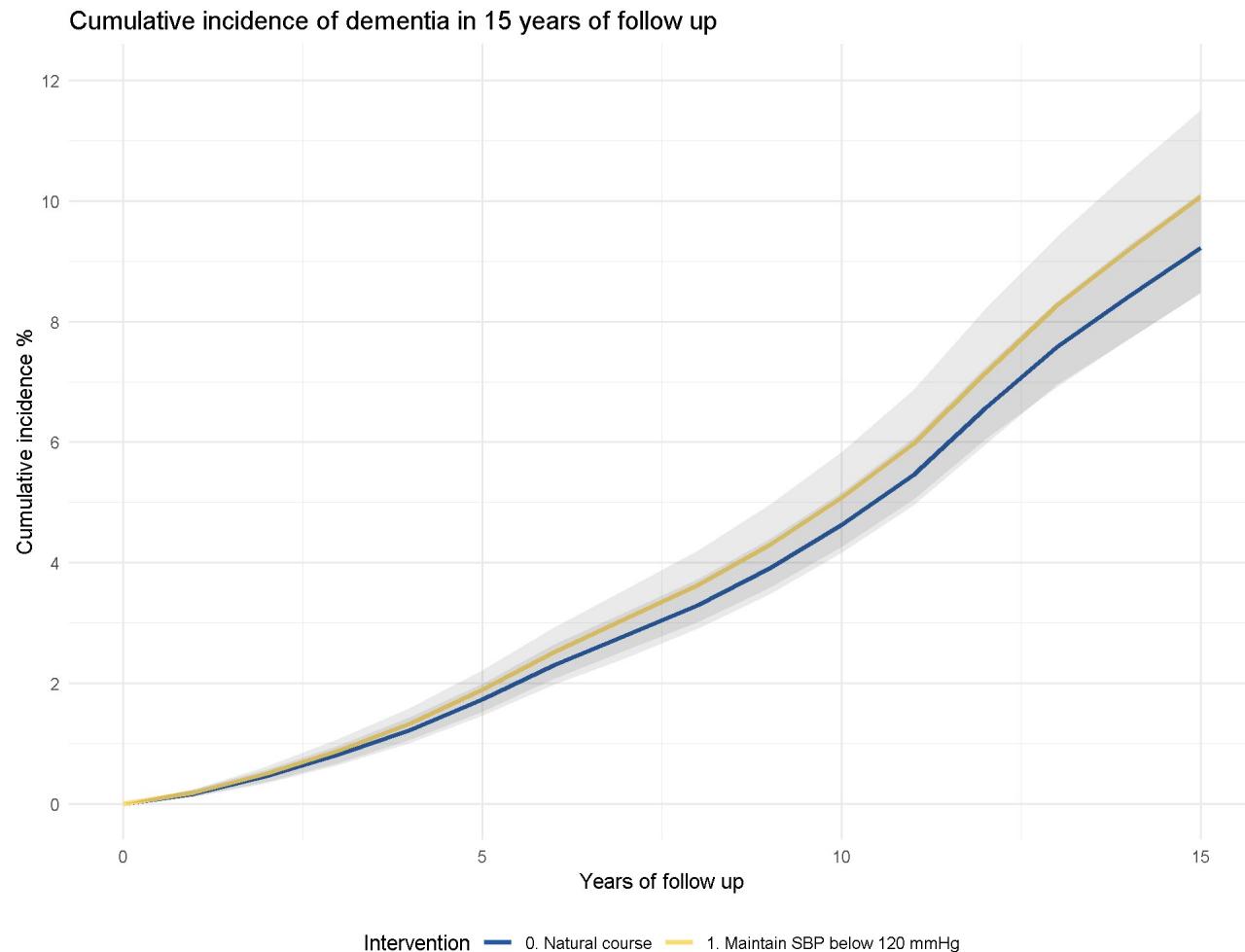
The observed risk at 15 years is **9.15%**.

Intervention	Risk	Risk Ratio	Risk Difference	Total Intervened %
0. Natural course	9.2 (8.3,10.3)	1 (1, 1)	0 (0, 0)	0.0
1. Maintain SBP below 120 mmHg	10.1 (8.2,11.8)	1.1 (0.9, 1.2)	0.8 (-0.6, 2.2)	98.2
2. Maintain SBP below 140 mmHg	9.4 (8.2,10.7)	1 (0.9, 1.1)	0.2 (-0.5, 0.9)	84.3
3. Reduce SPB by 10% if above 140	9.5 (8.2,11)	1 (0.9, 1.1)	0.3 (-0.5, 1.1)	84.0
4. Reduce SPB by 20% if above 140	9.7 (8.3,11.4)	1.1 (0.9, 1.2)	0.5 (-0.6, 1.8)	84.0
5. Quit smoking (if current smoker)	9.4 (8.4,10.4)	1 (1, 1.1)	0.1 (-0.3, 0.5)	26.4
6. Joint 1 + 5	10.1 (8.3,12)	1.1 (0.9, 1.3)	0.9 (-0.6, 2.5)	98.9
7. Joint 2 + 5	9.5 (8.3,10.8)	1 (1, 1.1)	0.3 (-0.5, 1.1)	89.6
8. Joint 3 + 5	9.7 (8.3,11)	1.1 (1, 1.1)	0.4 (-0.5, 1.3)	89.1
9. Joint 4 + 5	9.9 (8.3,11.6)	1.1 (0.9, 1.2)	0.7 (-0.6, 2)	89.1

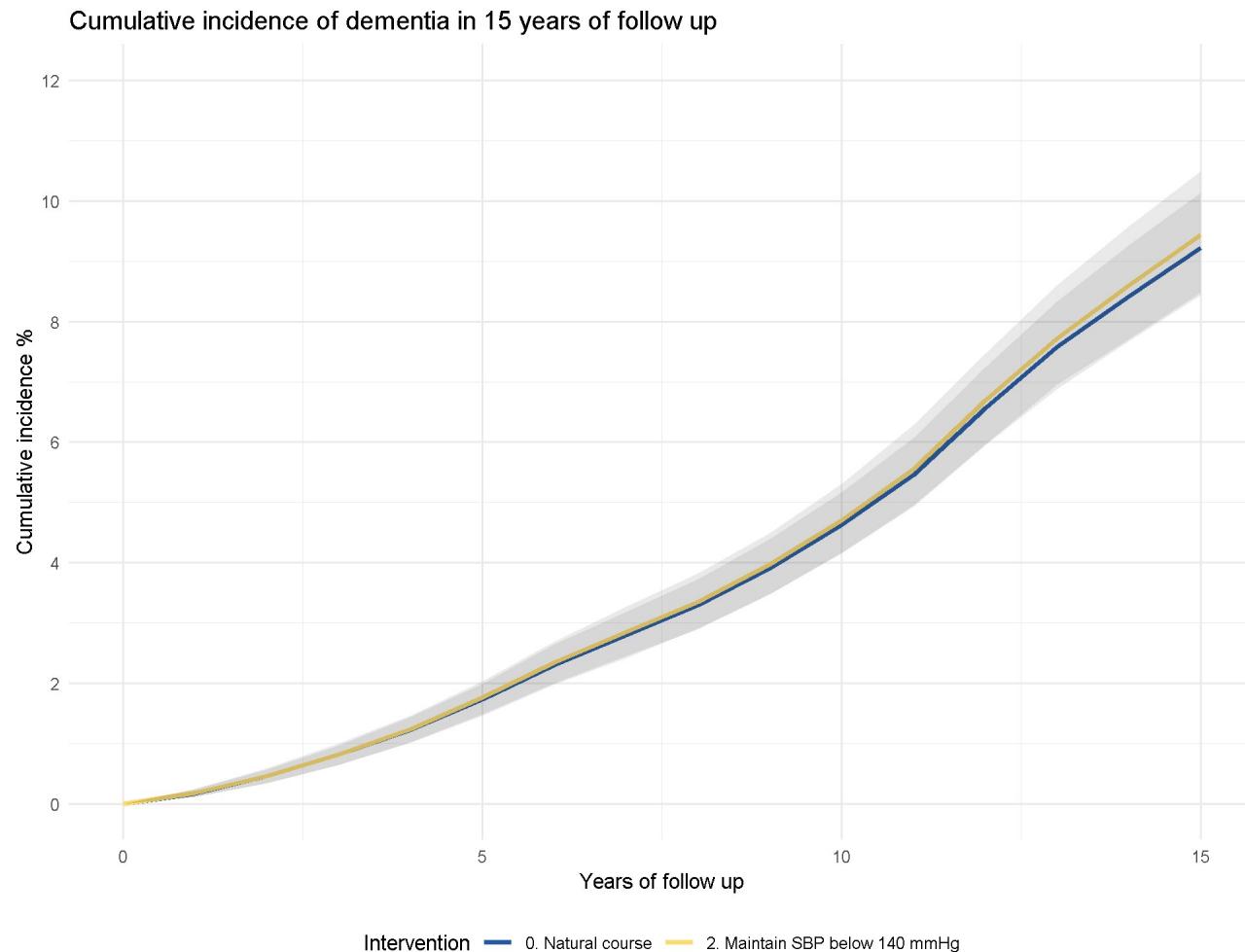
Note:

Results with 500 bootstraps.

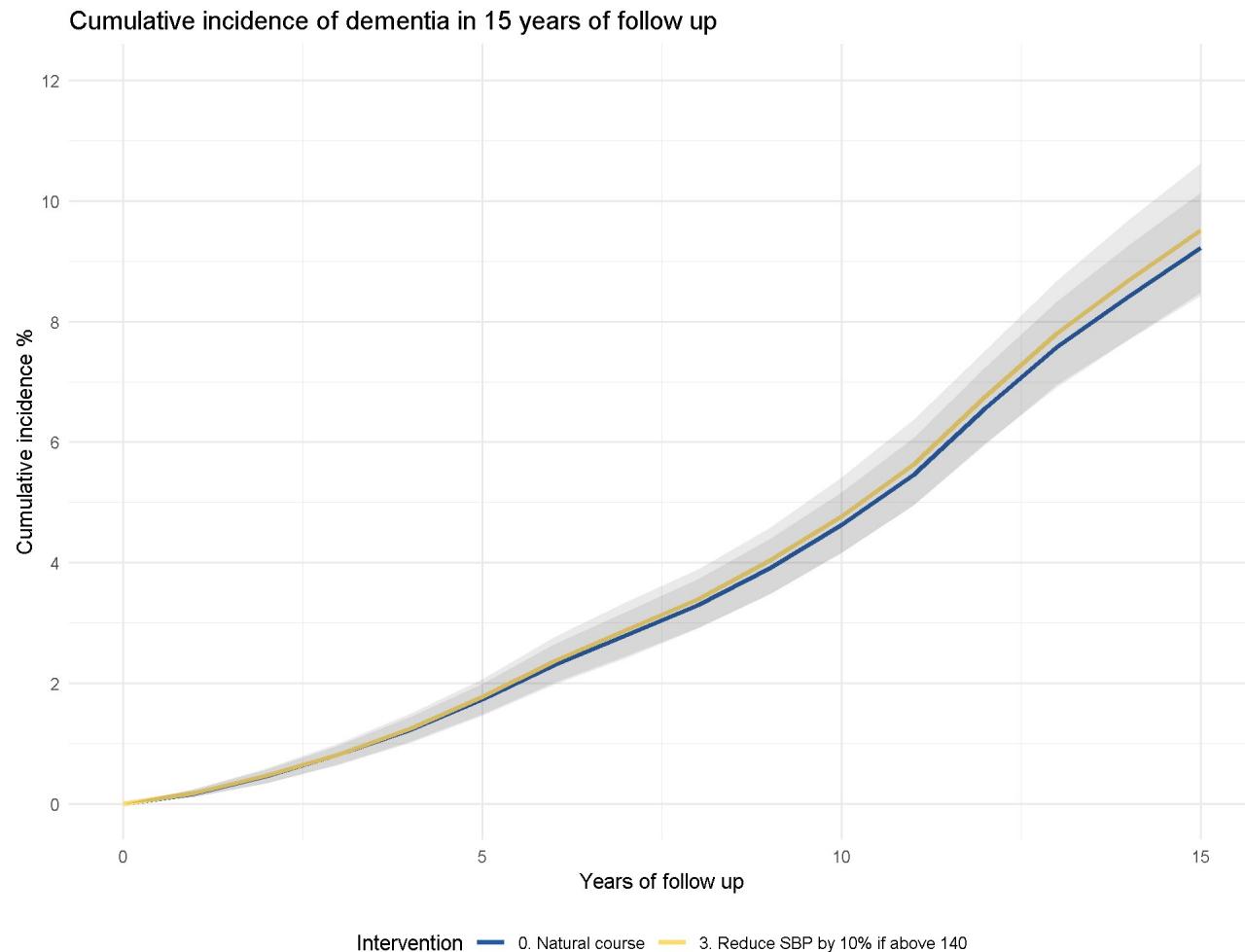
Results: Dementia



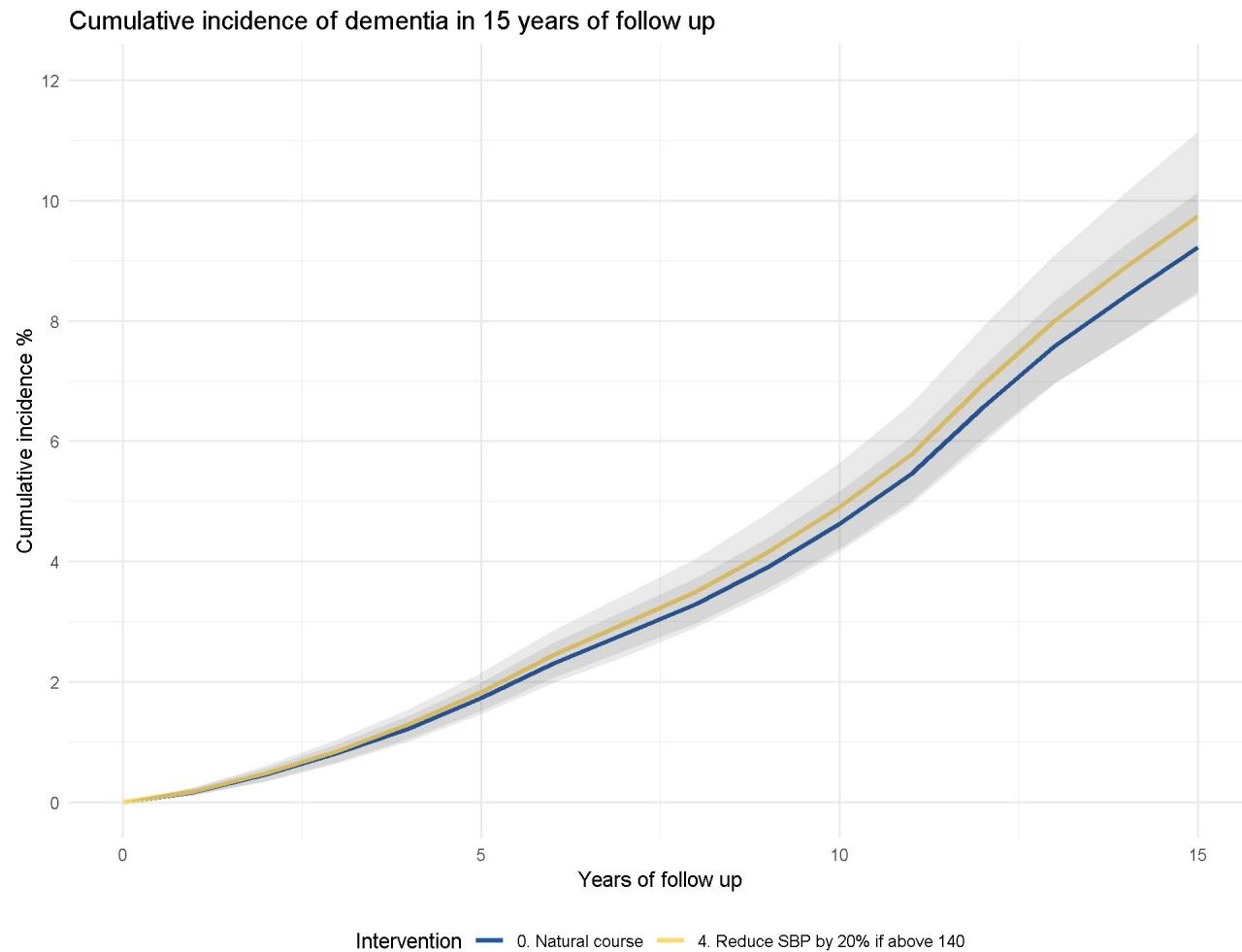
Results: Dementia



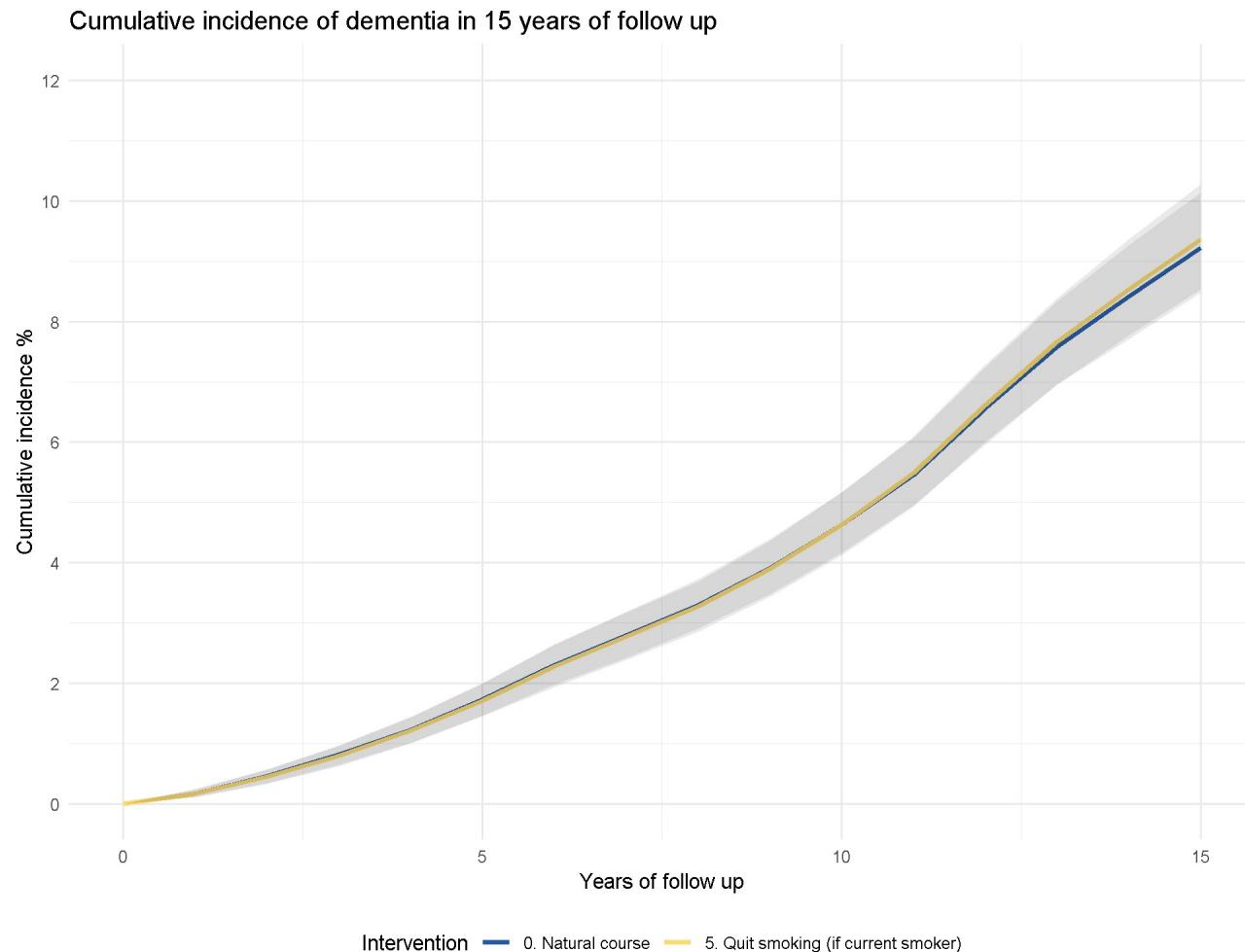
Results: Dementia



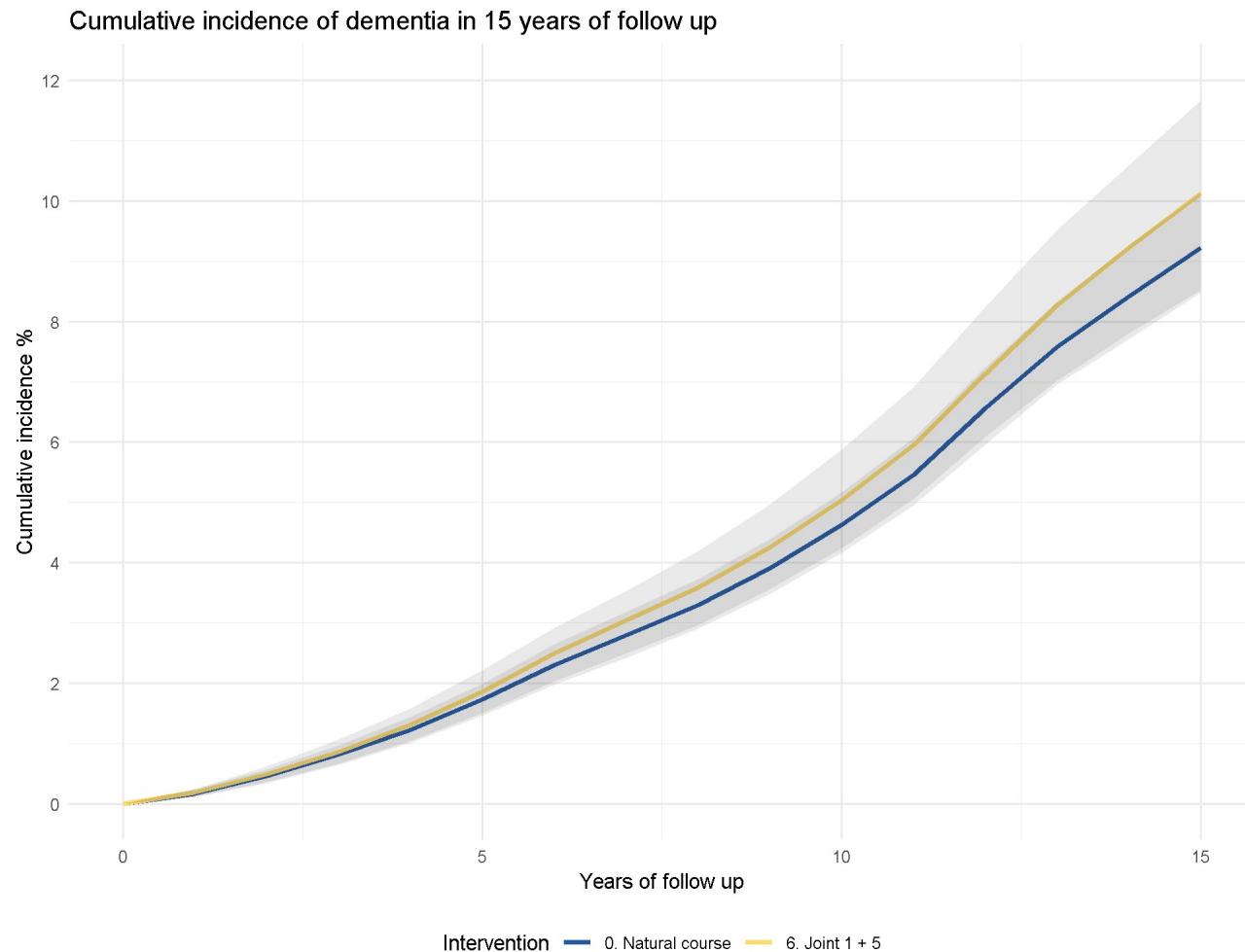
Results: Dementia



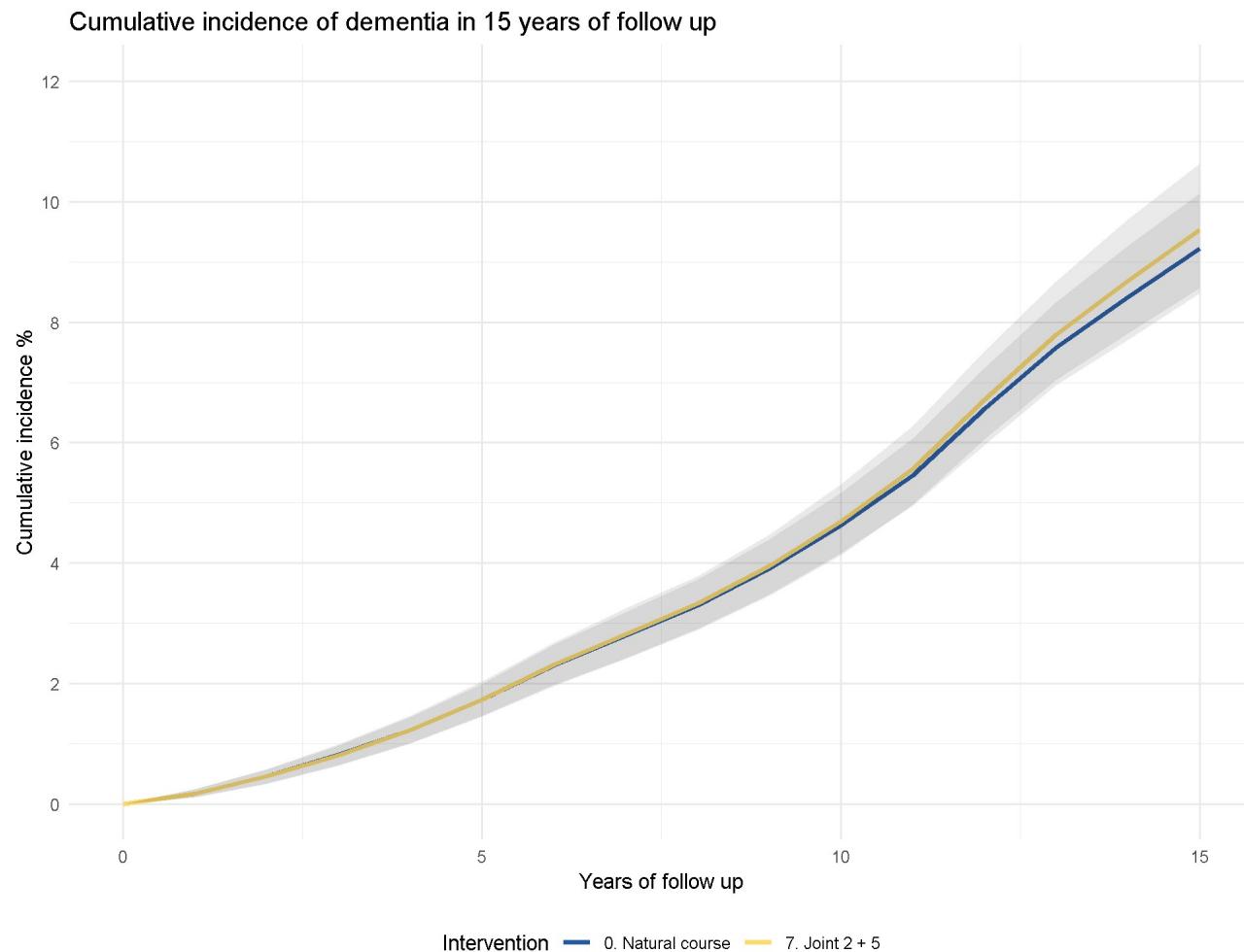
Results: Dementia



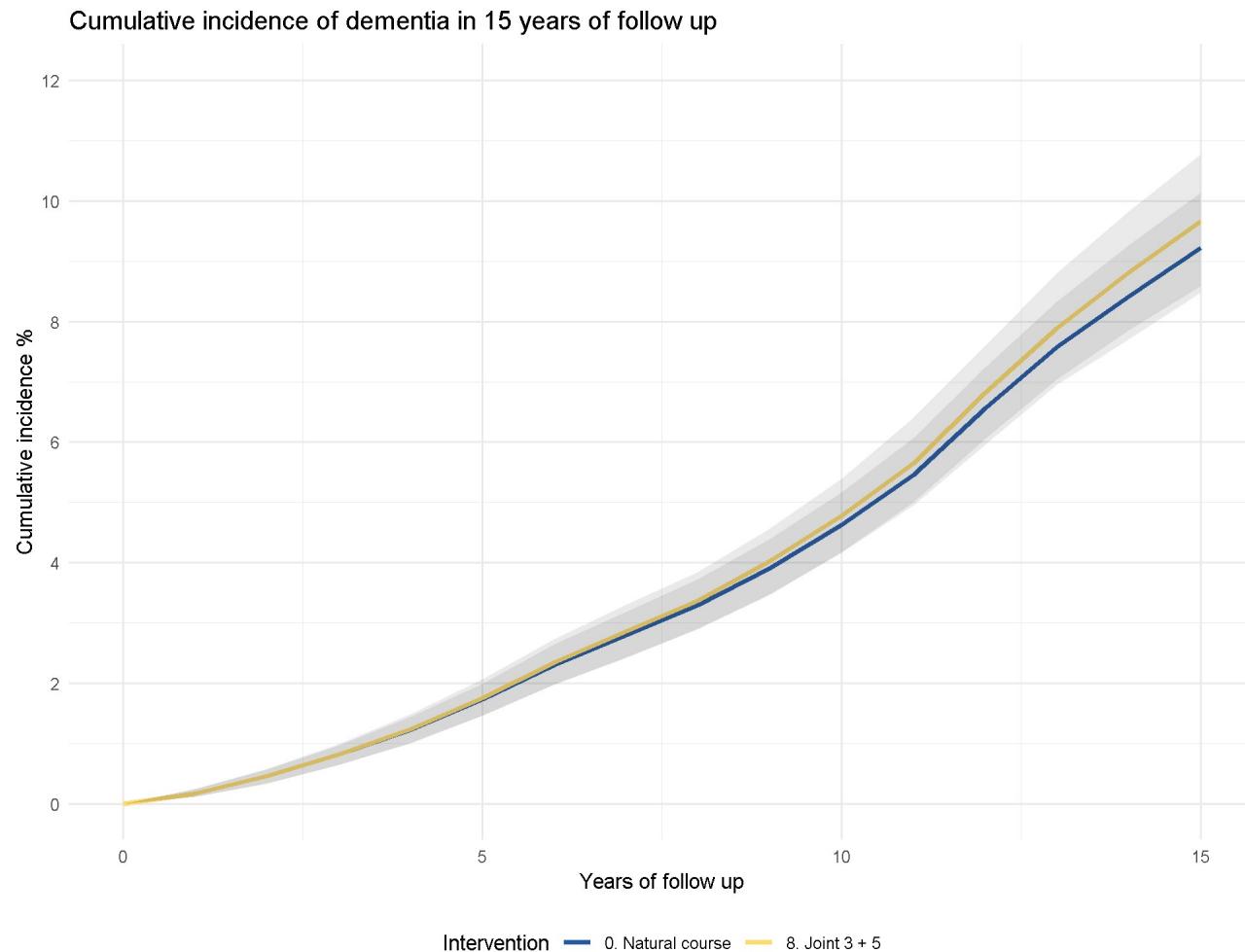
Results: Dementia



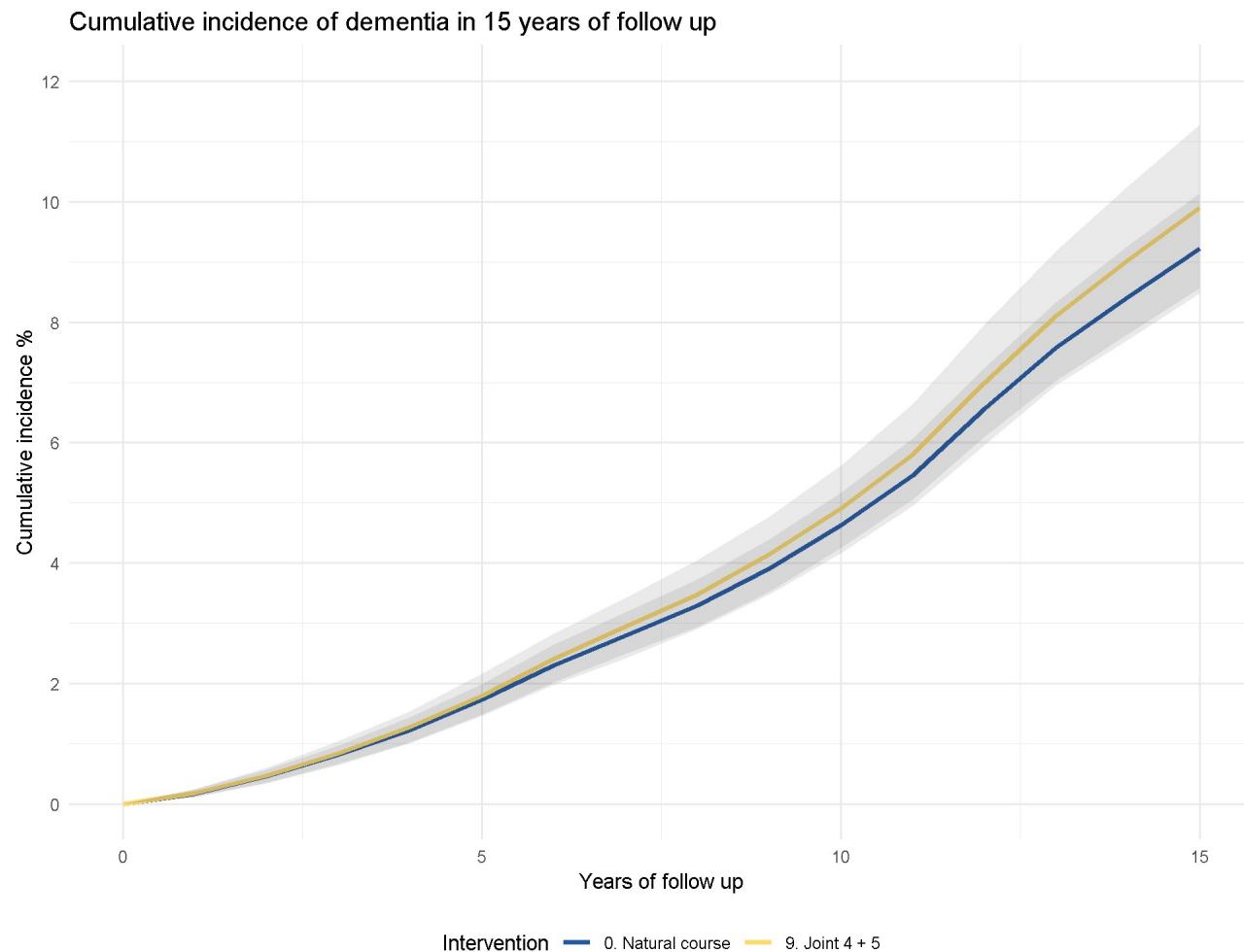
Results: Dementia



Results: Dementia



Results: Dementia

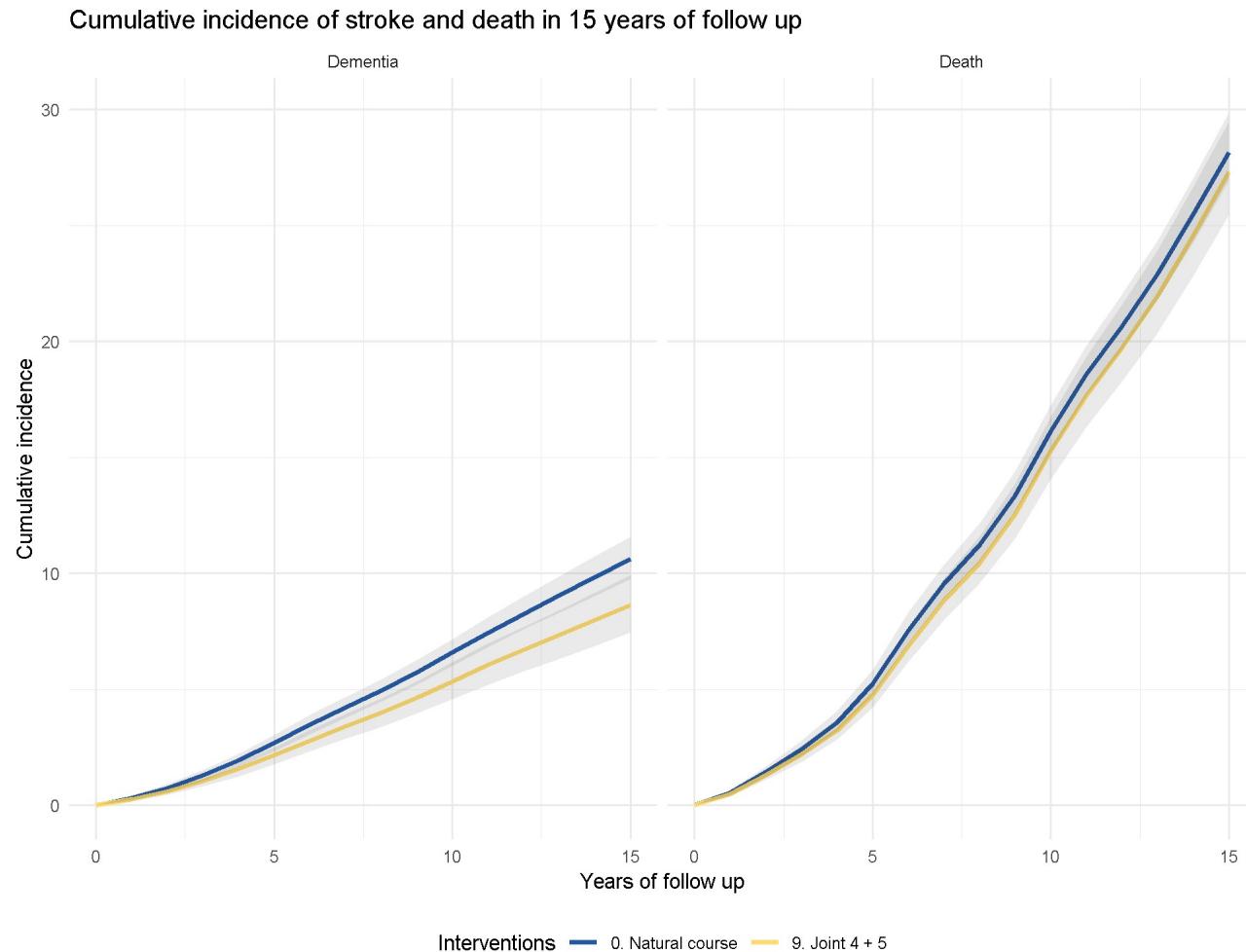


Conclusion

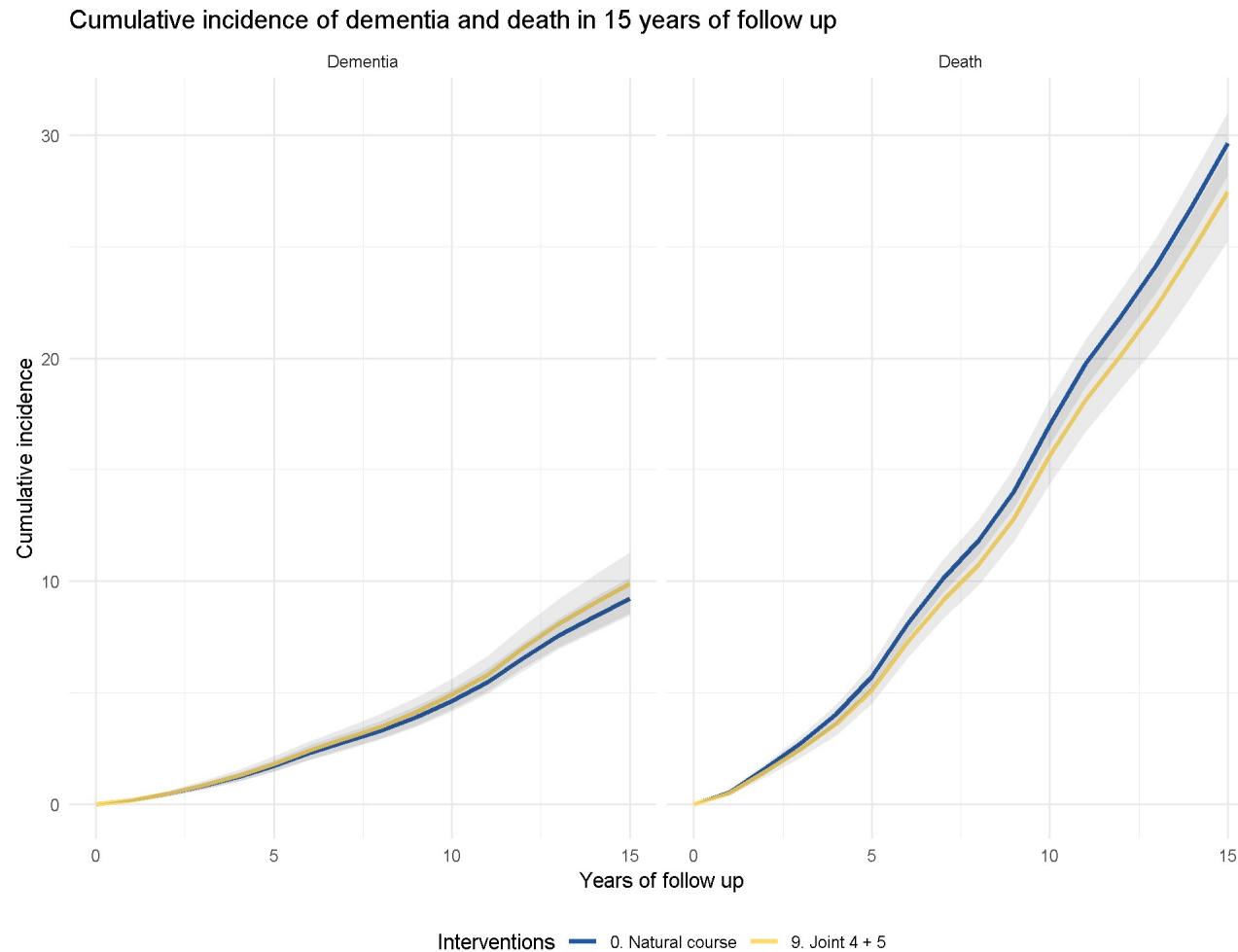
- All interventions decrease the risk of stroke at 15 years.
- Joint interventions decrease the risk of stroke even more.
- There is no effect of interventions in the risk of dementia at 15 years.
- Results with dementia as the outcome may be affected by selection bias.

The problem with the competing events

What happens to the risk of death when stroke is the main outcome?



What happens to the risk of death when dementia is the main outcome?



Beyond the subhazard and cause-specific dichotomy

Table 3. Recommendations for Analyzing Competing Risk Survival Data

-
- Cumulative incidence functions (CIFs) should be used to estimate the incidence of each of the different types of competing risks. Do not use the Kaplan-Meier estimate of the survival function for this purpose.
 - Researchers need to decide whether the research objective is on addressing etiologic questions or on estimating incidence or predicting prognosis.
 - Use the Fine-Gray subdistribution hazard model when the focus is on estimating incidence or predicting prognosis in the presence of competing risks.
 - Use the cause-specific hazard model when the focus is on addressing etiologic questions.
 - In some settings, both types of regression models should be estimated for each of the competing risks to permit a full understanding of the effect of covariates on the incidence and the rate of occurrence of each outcome.
-

What if we could frame our causal question in different ways?

$$Pr[Y_{k+1}^{a,c=0,d=0} = 1]$$

What is the risk of the event, if we could prevent death?

Direct Effect: The treatment's effect on the event of interest not mediated by the competing event.

Death is treated as a censoring event

$$Pr[Y_{k+1}^{a,c=0} = 1]$$

What is the risk of the event if we do not eliminate de competing event?

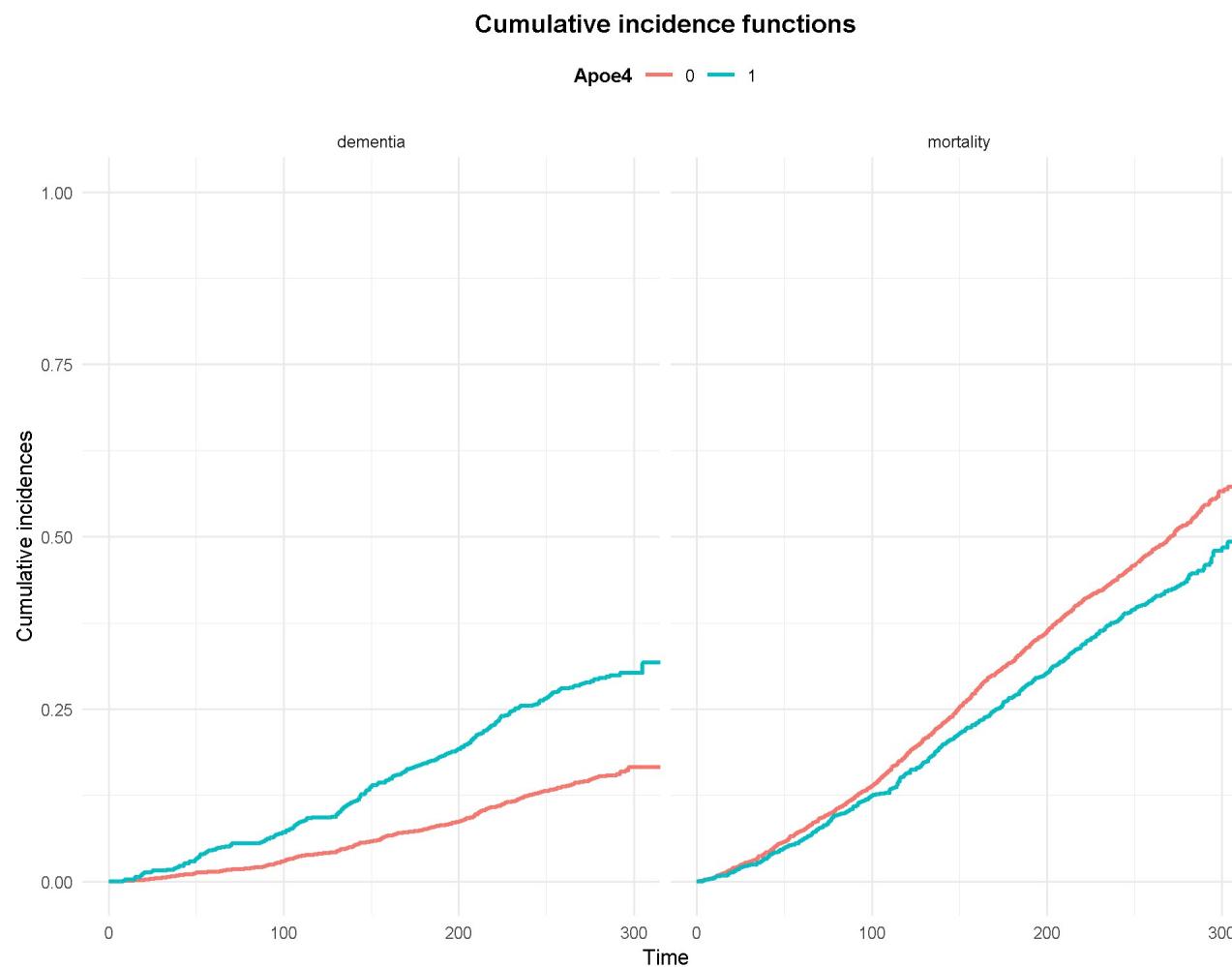
Total Effect: The treatment's effect on the event of interest through all causal pathways between treatment and the event of interest, includung those possibly mediated by the competing event.

Death is not treated as a censoring event, but as a covariate

From theory to the real setting

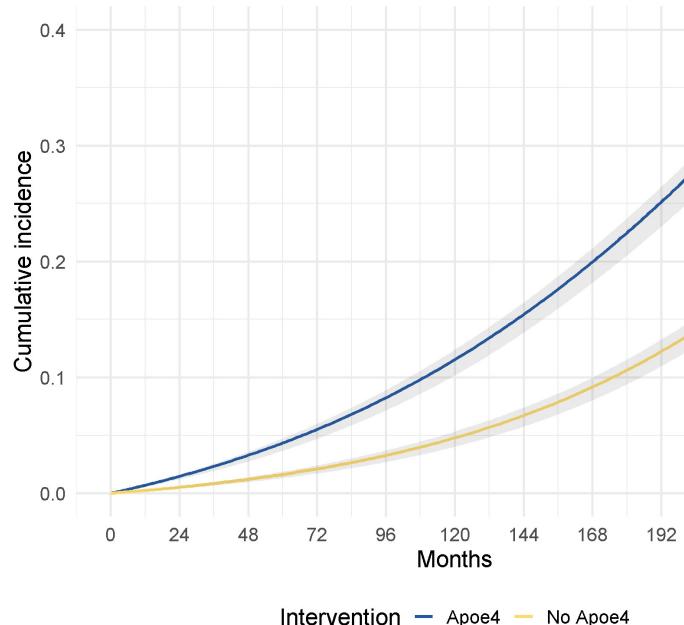
- In aging related research death will always be a competing event to worry about (death can represent 2x/3x the risk of our outcome of interest).
- The magnitude of the difference between the two estimates can be driven by the effect of the intervention in the competing event.
- We need to explore which assumption would we be willing to take in real-case scenarios, and explore in which scenarios these estimates could bring opposite effects.

Apoe 4: Crude cumulative incidence of dementia and death

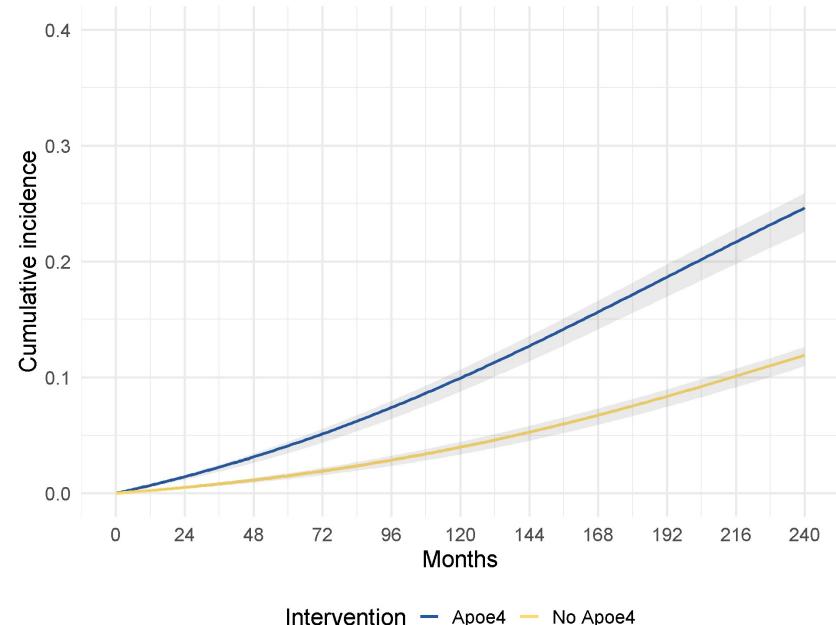


Apoe: Adjusted survival curves

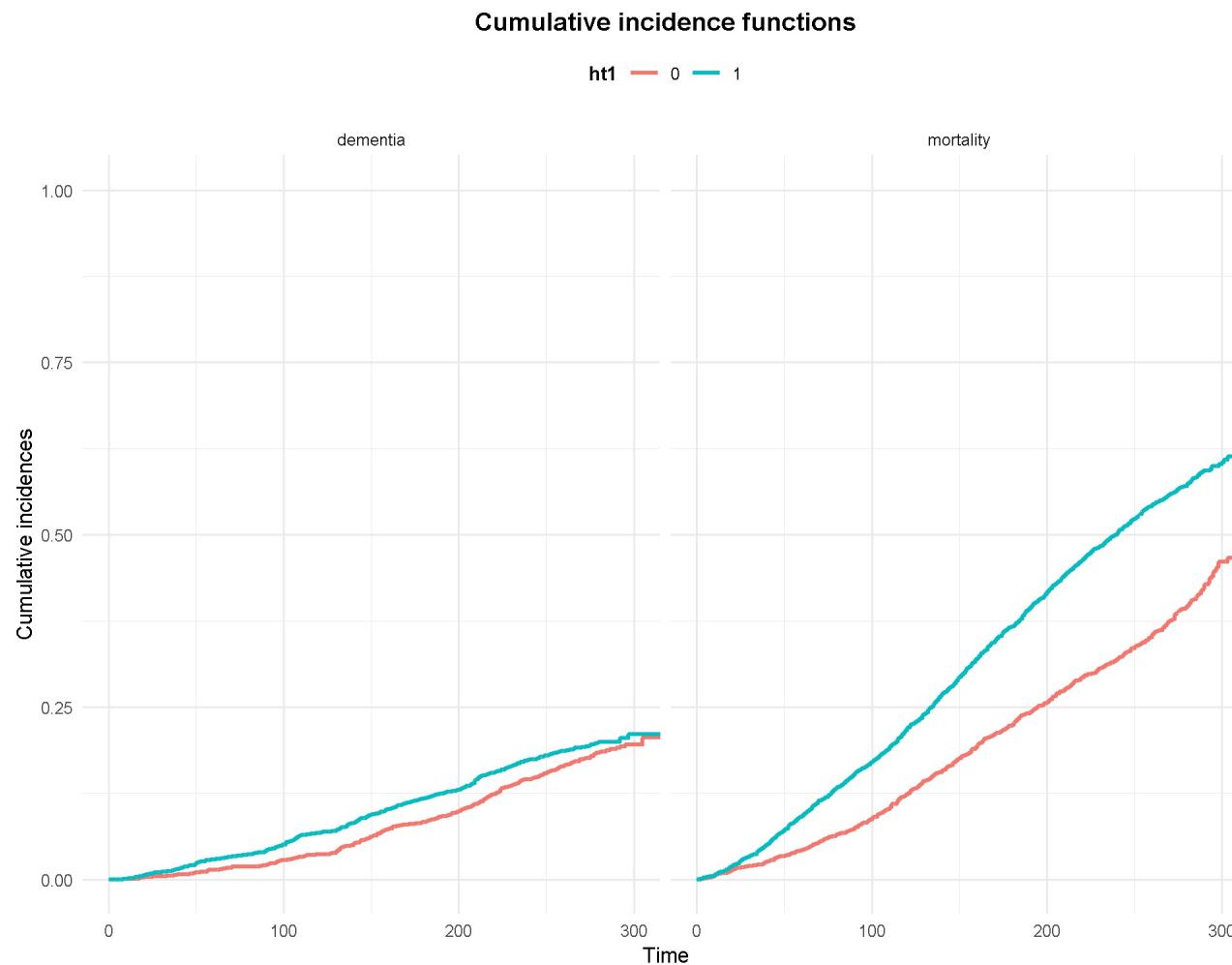
Direct Effect



Total Effect

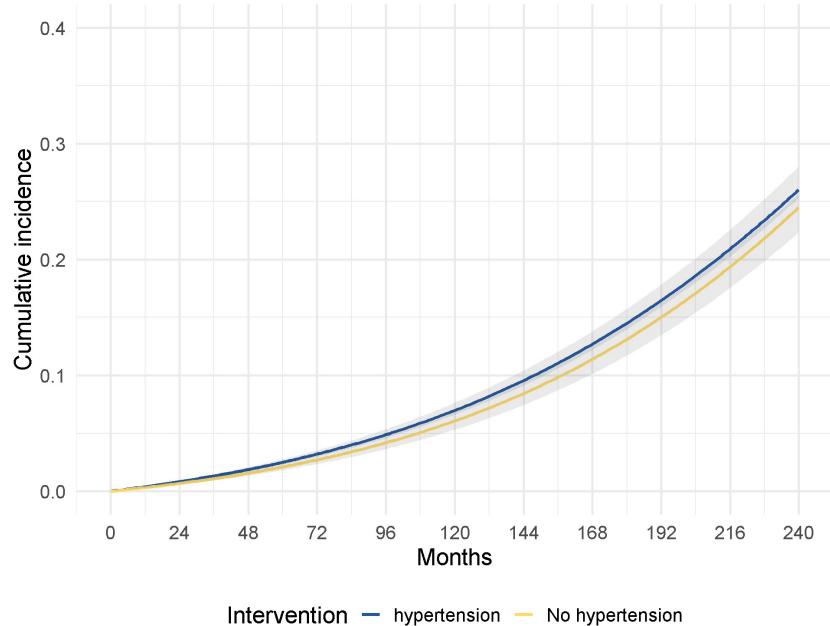


Time-fixed hypertension: Crude cumulative incidence

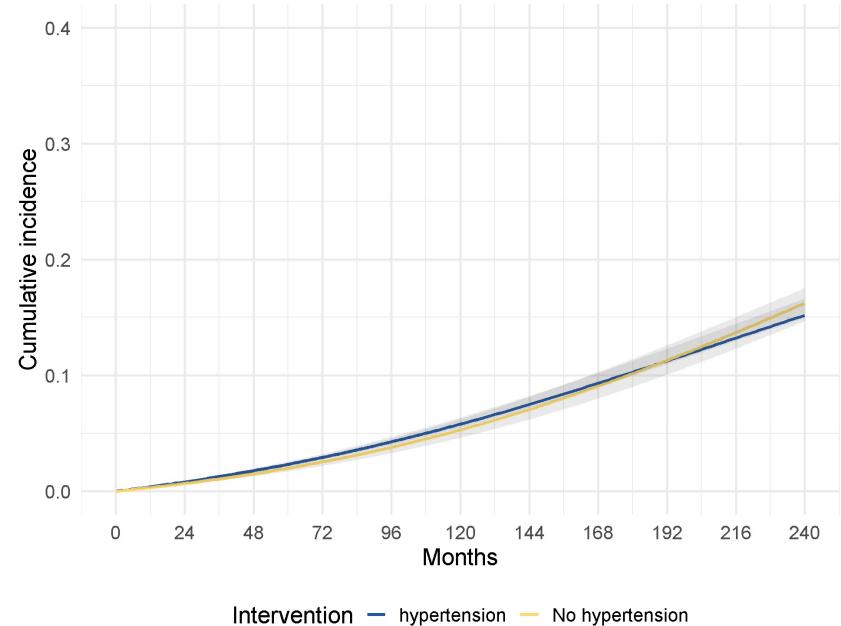


Time-fixed hypertension: Adjusted survival curves

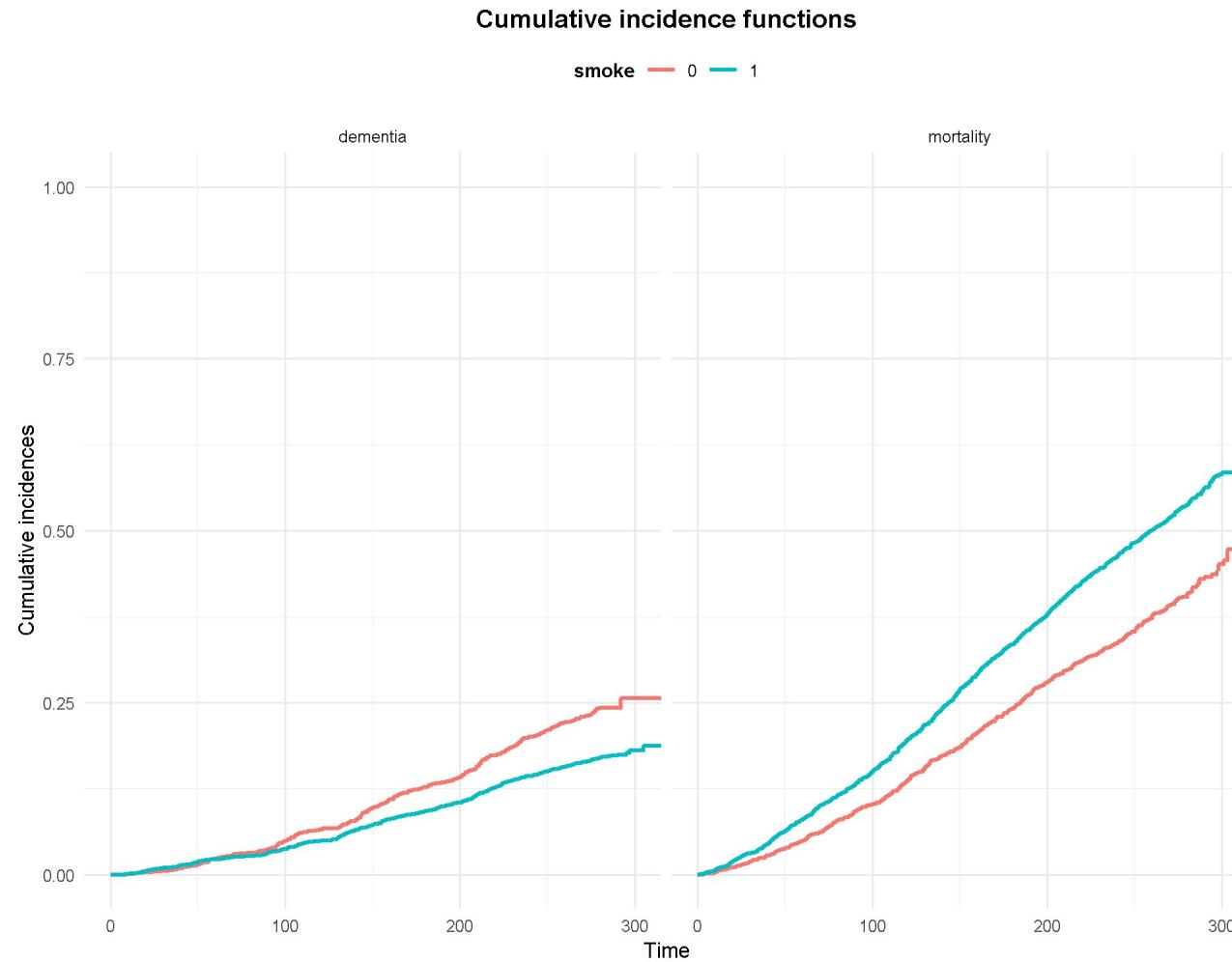
Direct Effect



Total Effect

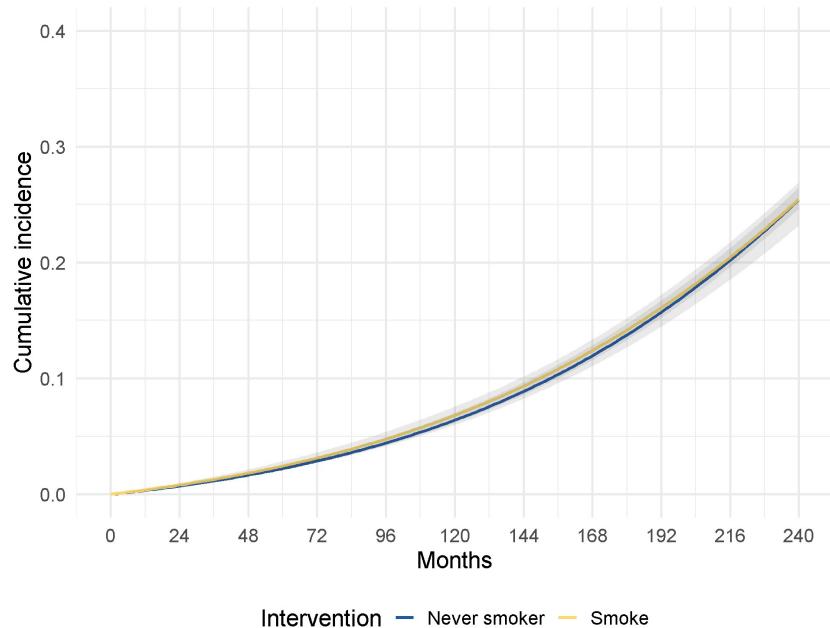


Time-fixed smoking (never vs. current or former): Crude cumulative incidence

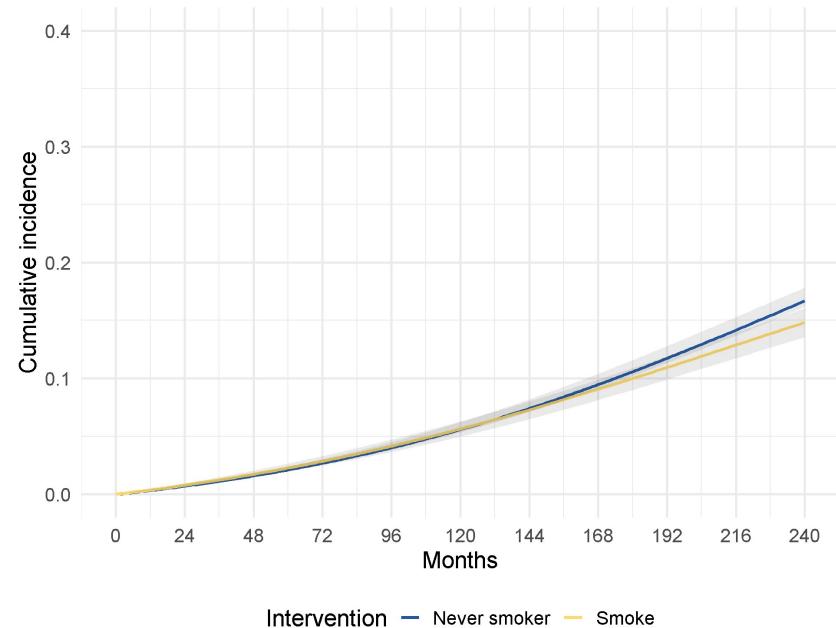


Time-fixed smoking (never vs. current or former): Adjusted survival curves

Direct Effect



Total Effect



Takehome messages

- Observational studies can be used to emulate a target trial.
- G-formula will help us solve the issue with time varying confounding.
- For longitudinal analysis, comprehension of data generating mechanisms is key.
- Competing risk analysis is beyond a statistical choice, it should be adopted as part of the causal question.

Acknowledgements

- Causal inference group
- Neuroepi group

Thank you!

