Supplement to:

Effectiveness of mammography screening on breast cancer mortality - a study protocol for emulation of target trials using German health claims data

Contents

1 Details on the statistical analysis

Discrete-time cumulative incidence functions (CIFs) will be estimated using the following approach. Let A_{never} (reference), A_{once}, and $A_{regular}$ be indicator variables for the screening strategies "never screened", "screened at least at baseline", and "screened at baseline and every two years afterwards". The discrete-time (cause specific) hazard is modelled using pooled logistic regression adjusted for baseline covariates:

$$logit\left(\mathbb{P}\left(Y_{t+1}\middle|\bar{Y}_t = 0, \bar{C}_t = 0, \bar{D}_t = 0, A_{once}, A_{regular}, X\right)\right)$$
$$= f_1(\theta_1', t) + f_2(\theta_2', t, A_{once}) + f_3\left(\theta_3', t, A_{regular}\right) + \theta_4 A_{once} + \theta_5 A_{regular} + \theta_6' X'.$$

The above model includes flexible functions $f(.)$ of time t, regression coefficients θ for (transformed) time and, possibly, interaction terms between time and screening strategy. The functions $f(.)$ will be determined by visual inspection so that the unadjusted parametric CIF estimated via pooled logistic modelling approximates the non-parametric Aalen-Johansen curves reasonably well. The binary variable Y_t denotes the outcome event breast cancer death at time t. The binary variable C_t denotes censoring status at time t and the binary variable D_t contains the event status of the competing event (death by other causes) at time t. Baseline covariates and interactions between covariates are denoted by X. The prime notation $(.)'$ denotes vectors. The history of a variable is denoted by overbars as $\overline{(.)}$. The above model is a marginal structural model and contains baseline covariates, but no time-varying covariates. Adjustment for time-varying confounding by X_t is achieved by inverse probability weighting, where time-varying weights are calculated for each screening strategy $A \in \{A_{never}, A_{once}, A_{regular}\}$ separately as

$$W_t^A = \prod_{k=1}^{t} \frac{1}{\widehat{\mathbb{P}}(A_k|\bar{A}_{k-1}, \bar{X}_k, \bar{Y}_{k-1} = \bar{C}_{k-1} = 0)},$$

truncating weights at the 99th percentile. Here A_k is the actual screening status at time k and is, by definition, consistent with the strategy A as individuals will otherwise be censored. For efficiency the above weights can be replaced by stabilized weights (see Cain et al. (2010) for a description of stabilized weights). Analogous weights are used for censoring due to competing events when estimating the direct effect. Below, upper indices refer to counterfactuals, e.g. the probability of breast cancer death under screening even if a portion of the study subjects did not experience screening, i.e. exposure is set to a value possibly contrary to the observed exposure (Hernan & Robins, 2020). The cumulative incidence function $\widehat{CIF}_{i,t}^{A=a}$ for clone $i = 1, \ldots, m$, at time point t under screening strategy $A = a$ will then be estimated using one of the approaches (i.e. based on modelling either subdistribution or cause-specific hazard) described in Young et al. (2020), depending on computational

cost. This cumulative incidence will be standardized to the empirical distribution of baseline confounders as

$$\widehat{CIF}_t^{A=a} = \frac{1}{m}\sum_{i=1}^{m} \widehat{CIF}_{i,t}^{A=a}.$$

As a function of time *t*, the above cumulative incidence function allows an assessment of how the effect of screening evolves over the whole of follow-up.

2 Illustration of assignment to screening strategies

S: screening in 1st quarter

C: breast cancer diagnosis in 1st quarter

D: death in 1st quarter

S&C: screening & breast cancer diagnosis in 1st quarter

S&D: screening & death in 1st quarter

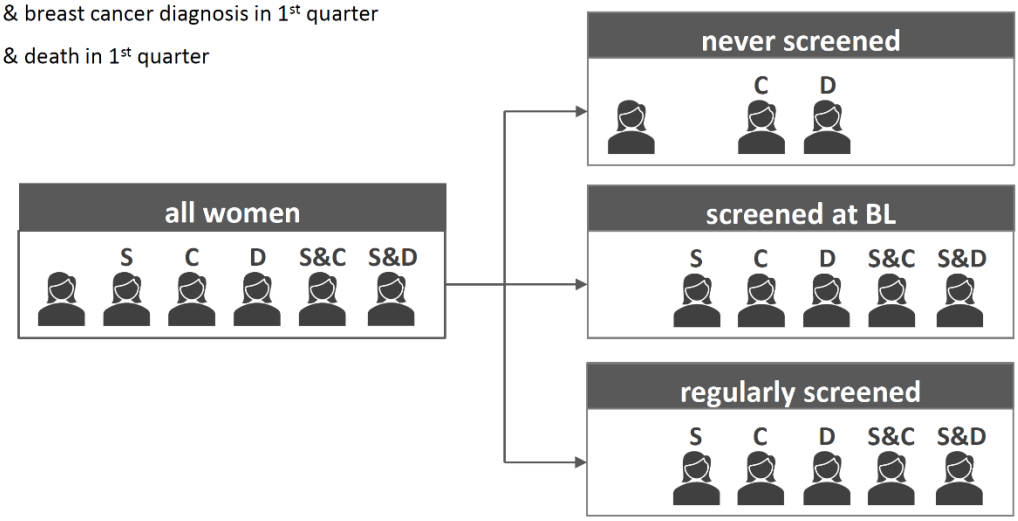


Figure S1: Illustration of cloning of women into the screening strategies. Assignment of clones to screening strategies is based on screening behaviour from the calendar quarter of baseline. Women with a breast cancer diagnosis or recorded death in the first quarter are cloned into all screening strategies, since they were compliant with all screening strategies until the diagnosis/death occurred.

3 Illustrations of artificial censoring schemes per screening strategy

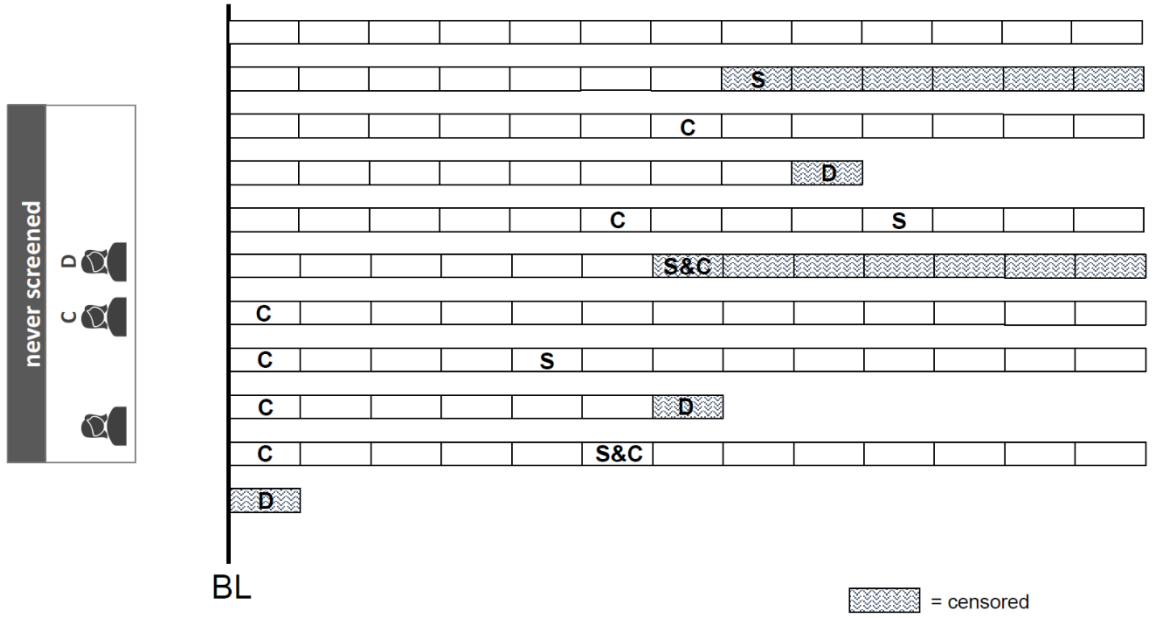
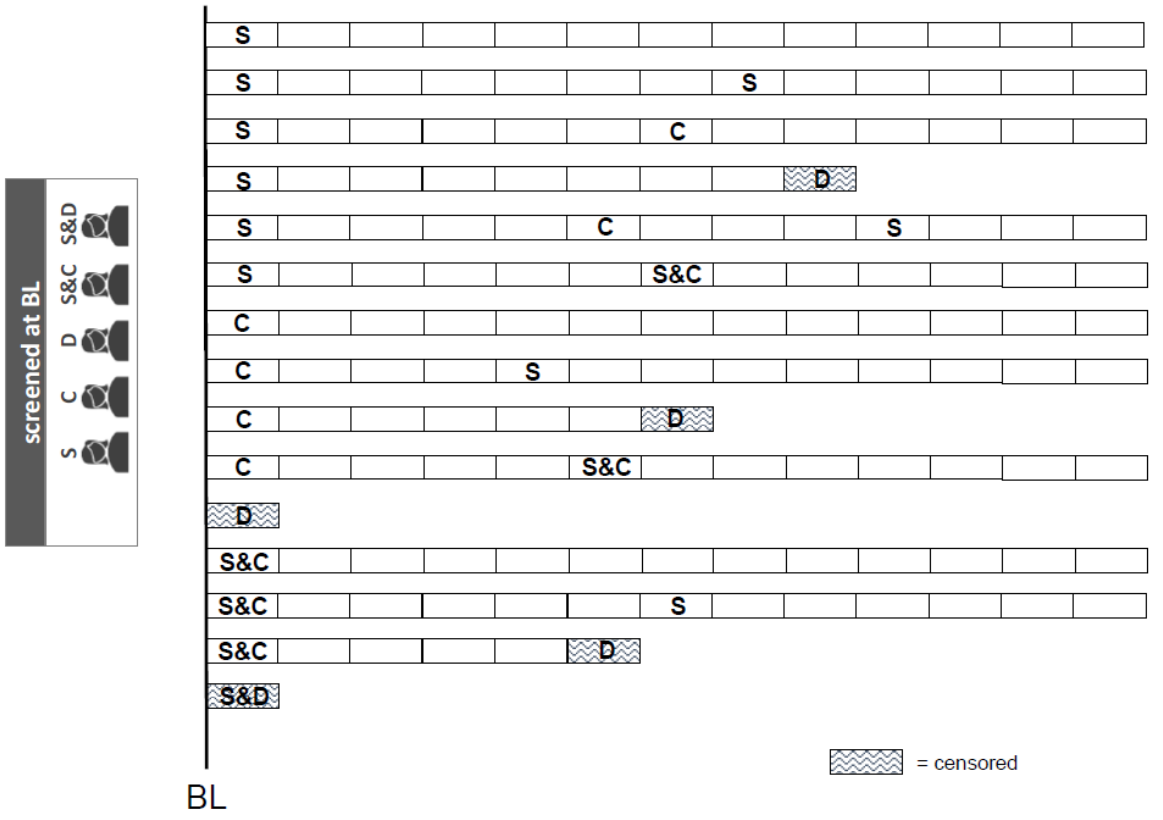


Figure S2: Illustration of artificial censoring scheme under screening strategy "never screened". Follow-up time is discretized into calendar quarters, with rectangles denoting individual quarters. The rationale for censoring is described in depth in the main body of the paper. Note that when a woman is censored, the time of censoring is set to the beginning of the calendar quarter that led to censoring. In the above illustration, the last woman is censored at baseline because she dies in the baseline quarter, i.e. she is censored at time point 0 with reason of censoring being death. C = breast cancer, D = death, S = screening, S&C = screening and cancer in the same quarter.

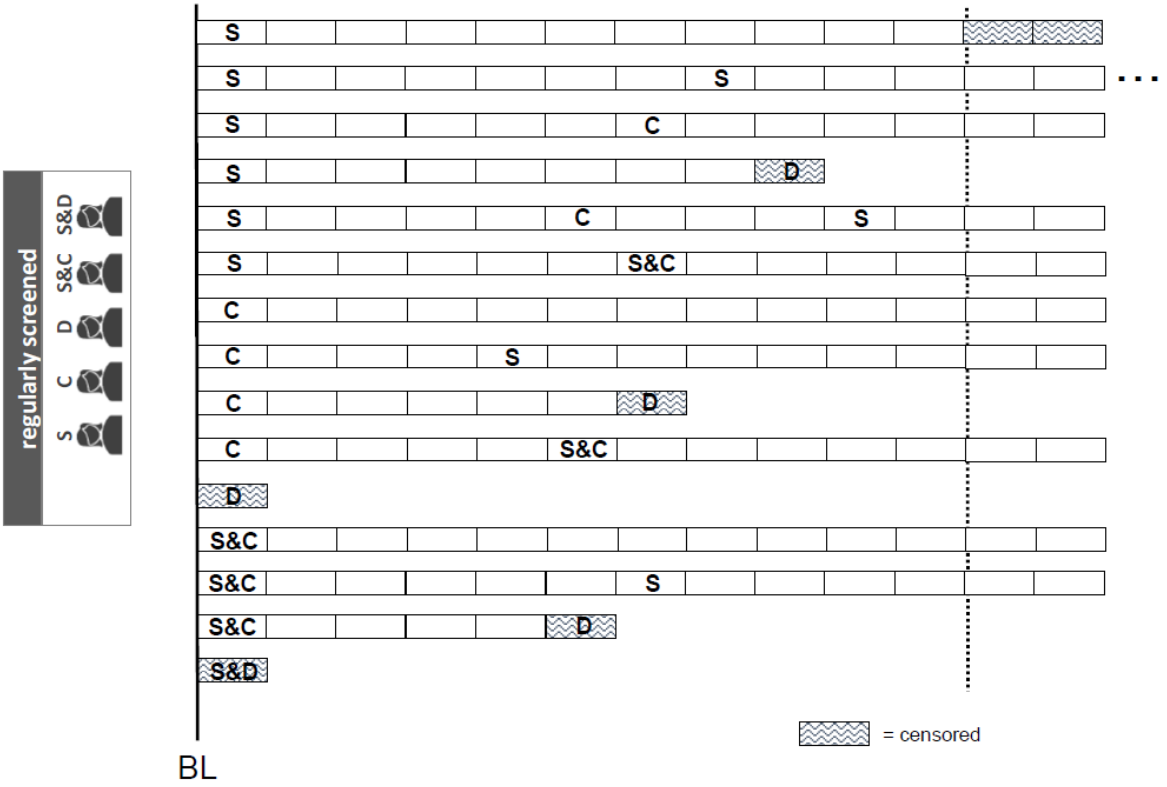


62

63

Figure S3: Illustration of artificial censoring scheme under screening strategy "screened at baseline". Follow-up time is discretized into calendar quarters, with rectangles denoting individual quarters. The rationale for censoring is described in depth in the main body of the paper. C = breast cancer, D = death, S = screening, S&C = screening and cancer in the same quarter, S&D = screening and death in the same quarter.

64



65

Figure S4: Illustration of artificial censoring scheme under screening strategy "regularly screened every two years". Follow-up time is discretized into calendar quarters, with rectangles denoting individual quarters. A regular screening is defined as having taken place between one year to ten quarters after the previous screening. The rationale for censoring is described in depth in the main body of the paper. C = breast cancer, D = death, S = screening, S&C = screening and cancer in the same quarter, S&D = screening and death in the same quarter. The dotted line indicates the end of the time period in which the second screening would need to take place.

66

67

4 Addressing potential sources of bias

Confounding: Covariates used to adjust for confounding will be derived at baseline and during follow-up. Their selection is based on subject matter knowledge and available literature. Risk factors for breast cancer were considered relevant, even though the outcome variable is breast cancer mortality, since developing breast cancer is a necessary antecedent for breast cancer death. Figure S5 illustrates the causal considerations for covariate selection. Adjustment for confounding will be carried out via standardization and inverse probability weighting.

Given that claims data are not collected for research purposes, direct information on relevant confounders is not always available or only available for extreme cases (e.g. heavy smoking, alcohol abuse). We aim to minimize this problem by using indirect information on these confounders (e.g. diseases resulting from exposure to these risk factors such as smoking-related diseases, or diseases resulting mainly from unhealthy behaviour such as obesity) as well as proxy variables for a health-seeking behaviour (e.g. utilization of preventive services, educational attainment). With respect to family history of breast cancer, the information is restricted to the ICD-10-GM code Z80.3 ("malignant neoplasm of the breast in the family"). It is not clear whether it is primarily coded in patients with a hereditary breast cancer syndrome rather than in those with a "simple" family history. The observed low proportion of women with Z80 codes (Braitmaier et al. 2022) indicates that it might only be used in high-risk subjects who would not be the target group of normal MSP screening. We therefore plan to conduct sensitivity analyses excluding women with this code. In addition, we will conduct a quantitative bias analysis to estimate the impact of unmeasured confounding regarding a "simple" family history of breast cancer.

For some risk factors, no information will be available in our data, for example age at menarche, parity, age at first full-term pregnancy, breastfeeding, age at menopause, height, breast density, exposure to radiation (unrelated to mammography). However, we argue that these risk factors are relatively unknown to the public and it is therefore reasonable to assume that they do not influence the decision to undergo screening.

Healthy screenee bias: Individuals volunteering for screening are generally healthier than individuals who choose not to undergo screening (Weiss & Rossing 1996). In addition to adjustment for confounding, we will address this specific issue by carrying out a subgroup analysis within screening-affine women, defined by their pre-baseline use of other preventive services (research question 2). This subpopulation is more homogenous regarding health seeking behaviour, and we expect an increased internal validity albeit at the cost of generalizability. Therefore, both effects, the one in the

full study sample and the one in the subgroup of screening-affine women, will be important for the evaluation of the screening programme.

Competing events: Death due to causes other than breast cancer is a competing event for the outcome of interest. We will compare the total effect (where death due to other causes is not treated as eliminable) with the direct effect of screening (where the competing event is treated as eliminable and thus censored with appropriate inverse probability of censoring weights, IPCW). Note that adjustment for confounding of the direct effect must also include common causes of the competing event and the study outcome, e.g. by including comorbidities (Young et al. 2020).

Time-related biases: Immortal time and other biases will be minimized by aligning eligibility checks and treatment assignment at time zero, i.e. baseline (Dickerman et al. 2019). Furthermore, women whose screening behaviour in the first quarter after trial start is consistent with more than one screening strategy will be copied and one clone will be assigned to each eligible screening strategy, i.e. women who undergo screening in the baseline quarter will be assigned to all active screening strategies. An alternative, but less efficient approach would be to randomly assign each person to exactly one of the eligible strategies (Garcia-Albeniz et al. 2020). Given that some information in the database used for this study is only available on a quarterly basis (e.g. outpatient diagnosis codes), it is impossible to break down the information into smaller time intervals than quarters. However, the length of follow-up required to observe the effect of screening is large (approx. 7 - 10 years) (Jatoi & Miller 2003). We therefore argue that the extent of bias due to the time units is negligible, as a delay of diagnosis of three months is unlikely to influence the screening effect.

Misclassification: Health claims data is primarily generated for reimbursement purposes and, therefore, some diagnosis codes might be used inappropriately for the underlying condition or over-used (e.g. diagnosis codes in the outpatient setting). To minimize misclassification, we define most of the diseases based on algorithms that, for example, combine different sources of information (e.g. diagnosis codes in combination with therapy), only use codes with a high validity (such as inpatient diagnosis codes) or only consider codes if recorded repeatedly. There may still be some misclassification of morbidity, but we consider this type of misclassification unlikely to differ between groups and negligible in our analysis. Risk factors that have a delayed impact on breast cancer may not be measured adequately due to a limited length of the available look-back period. For instance, HRT might influence breast cancer risk only after several years. Thus, a woman who stopped HRT treatment five years before baseline would be misclassified as “no HRT use” if her look-back period in the data is only three years. We will systematically describe the available look-back period (stratified by age at baseline) to assess whether this could be a relevant misclassification. Finally, misclassification of the outcome variable of breast cancer related deaths might occur since this variable is not directly

available for much of the data and must be derived based on an algorithm. Langner et al. (2019) reported a sensitivity of 91.3 % and a specificity of 97.4 % for a former version of this algorithm, which is currently being further optimized and will be validated again based on a sample for which the official cause of death is available.

Identifying assumptions: We make the usual assumptions for causal inference from observational data, namely consistency, sequential exchangeability given observed covariates, and positivity. Consistency is fulfilled when the screening strategies being assessed are well-defined and correspond to the screening behaviour observed in the data, e.g. the outcome for a woman who happens to never undergo screening is the same as if she had been assigned to never undergo screening in the target trial. Sequential exchangeability is fulfilled when the observed screening behaviour of a woman at time *t* is independent of her potential outcomes under the strategies given the measured covariates prior to *t;* this can be thought of as no unmeasured baseline or time-varying confounding. Positivity is fulfilled when the probability of observing a screening strategy is greater than zero for all strategies in all covariate strata (Young et al. 2020, Hernan & Robins 2020). Furthermore, censoring competing events to obtain the direct effect requires an assumption of no unmeasured common causes of the different event types.

5 Illustration of causal considerations for covariate selection

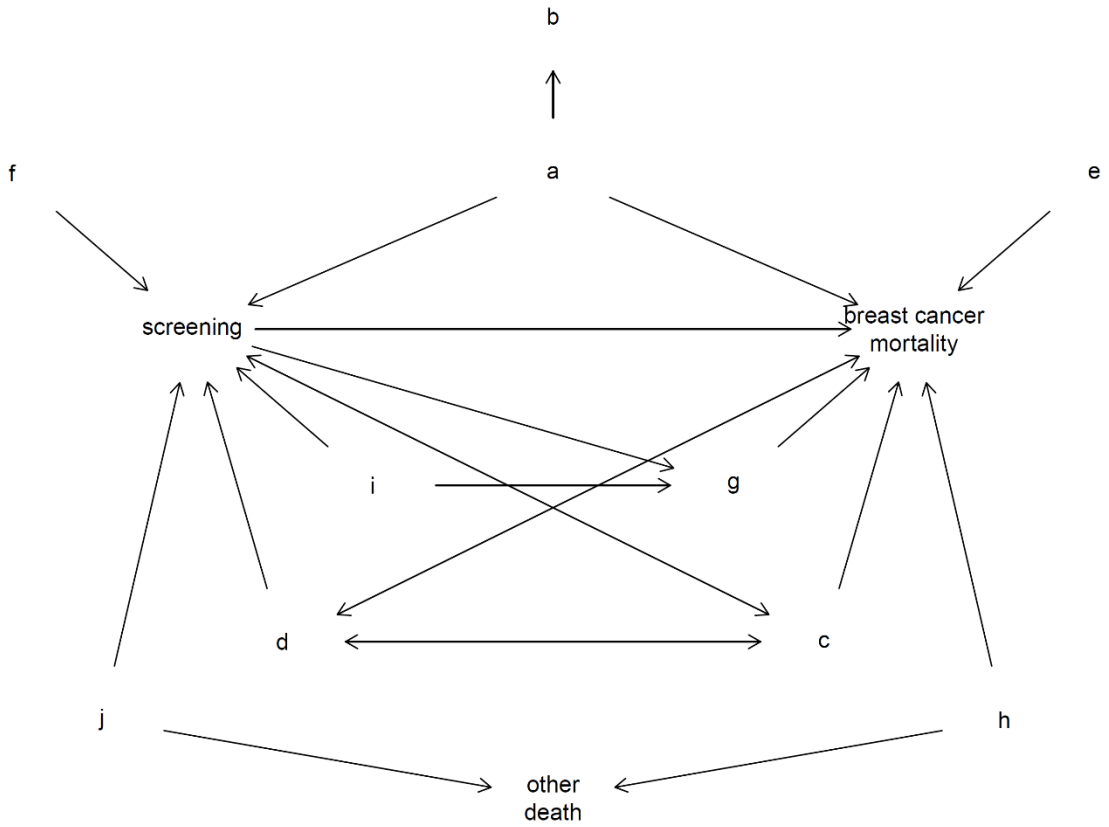


Figure S5: Illustration of variable groups considered for covariate adjustment and their causal connections. Note that this is a simplified graph, ignoring the longitudinal aspect of the study. A directed edge from one variable to another means that the first variable is a direct cause of the second. Screening is the exposure, breast cancer death is the outcome, and other death is a competing event. A bi-directed edge can be interpreted as presence of latent variables between the two connected variables. Variables "a" are common causes of screening and outcome. Variables "b" are proxies for those of category "a". Variables "c" are causes of the outcome that are associated with exposure. Variables "d" are causes of the exposure that are associated with the outcome. Variables "e" are causes of the outcome that are not associated with exposure. Variables "f" are causes of the exposure that are not associated with the outcome. Variables "g" are post-screening variables that are mediators between exposure and outcome. Variables "h" have a causative effect both on the competing event and the outcome. Variables "i" are causes of exposure and mediators. Variables "j" are confounders between exposure and the competing event. Variables "f" should not be included for adjustment, as this can lead to bias-amplification in case of residual unobserved confounding. Variables "g" (e.g. treatment after screening) should not be included for adjustment, as they are on the causal path from exposure to outcome. Variables "a", "b" (if "a" is unmeasured), "c", "d", "h" (only for estimating the direct effect, not for the total effect), "i", and "j" should be included for adjustment to mitigate confounding. Variables "e" are not needed for adjustment but can be included to increase precision of estimation. The variable groups (except "f") are not mutually exclusive, and in fact many variables will fit into more than one of these groups. An example of a covariate of the category "a" would be previous use of menopausal hormone therapy, as this is a known risk factor for breast cancer and physicians might advise women with this risk factor to attend screening. An example of a covariate of the category "j" would be presence of palliative care. An example for "d" might be educational attainment as it may affect awareness of screening and is strongly associated with direct risk factors "c" of breast cancer mortality; educational attainment can also be seen as type "b" proxy for further unmeasured confounders.

6 List of covariates

In Table S1 below, we give an overview of variables used to adjust for confounding. Time-varying covariates will be re-assessed on a quarterly basis. Variables might be added to this list of covariates, if indicated by subject matter knowledge. The list of covariates used in the final analysis will be finalized before data on the study outcome becomes available. Note that this is just an alphabetical list of covariates that will be defined based on the information in the database. Content-wise, a discussion on how confounding as a potential source of bias is considered and how relevant covariates are captured in the data is provided in Supplement 4. Furthermore, Figure S5 illustrates the causal considerations for covariate selection.

The covariates in Table S1 are mostly implemented as binary (time-dependent) variables. For most of the variables, algorithms considering different types of information (e.g. diagnosis codes in combination with therapy) will be developed or have been developed, with the aim of maximizing validity and thus minimizing misclassification (see also Supplement 4).

Table S1: Relevant covariates for confounder adjustment.

variable/variable group	**time-varying**
Acute hemorrhagic stroke	yes
Acute ischemic stroke	yes
Acute myocardial infarction	yes
Age at baseline	no
Alcohol abuse	yes
Anaemia	yes
Anticoagulant therapy	yes
Antihypertensive therapy	yes
Antiplatelet therapy	yes
Benign neoplasm of breast	yes
Breast disorders (benign mammary dysplasia, inflammatory disorders of breast, hypertrophy of breast, unspecified lump in breast, other disorders)	yes
Bronchial asthma	yes
Cachexia	yes
Chronic obstructive pulmonary disease (COPD)	yes
Coronary heart disease	yes
Dementia	yes
Diabetes with end organ damage	yes
Drug abuse	yes
Drug-treated (arterial) hypertension	yes
Educational attainment	no
Family history of breast cancer*	yes
Glaucoma	yes
Heart failure	yes
Hemiplegia	yes
Hepatitis B or C	yes
Hip fracture	yes
HIV therapy	yes

Hormone replacement therapy	yes
Lipid-lowering therapy	yes
Liver diseases including chronic viral hepatitis	yes
Mental diseases	yes
Number of hospitalizations	yes
Number of non-screening mammographies	yes
Number of outpatient physician contacts	yes
Number of prescriptions	yes
Number of screening mammographies	yes
Obesity/adiposity	yes
Other cancers	yes
Palliative care	yes
Severe liver disease	yes
Terminal renal disease	yes
Tobacco abuse	yes
Treated diabetes	yes
Treatment for hypothyroidism	yes
Treatment for osteoporosis	yes
Treatment with antidepressants	yes
Treatment with antipsychotics	yes
Treatment with immunosuppressive drugs	yes
Treatment with opioids	yes

170 * Given that information on family history is limited, additional methods will be taken to consider
171 this confounder (see manuscript and Supplement 4).

172

7 References

Braitmaier M, Sarina S, Kollhorst B, et al. (2022) Screening colonoscopy similarly prevented distal and proximal colorectal cancer; A prospective study among 55-69-year-olds; Journal of Clinical Epidemiology; 149: 118-126

Cain LE, Saag MS, Petersen M, et al. (2016) Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy; Int J Epidemiol; 45(5):2038-2049

Dickerman BA, Garcia-Albeniz X, Logan RW, et al. (2019) Avoidable flaws in observational analyses: an application to statins and cancer; Nat Med; 25(10): 1601 – 1606

Garcia-Albeniz X, Hernan MA, Hsu J (2020) Continuation of annual screening mammography and breast cancer mortality in women older than 70 years; Ann Intern Med; 173(3): 247

Hernan MA, Robins JM (2020) Causal Inference – What If. Boca Raton, FL: Chapman & Hall/CRC

Jatoi I, Miller AB (2003) Why is breast cancer mortality declining?; Lancet Oncol; 4(4): 251 – 254

Langner I, Ohlmeier C, Haug U (2019) Implementation of an algorithm fort he identification of breast cancer deaths in German health insurance claims data: a validation study based on record linkage with administrative mortality data; BMJ Open; 9(7): e026834

Weiss NS, Rossing MA (1996) Healthy screenee bias in epidemiological studies of cancer incidence; Epidemiology; 7(3): 319 – 322

Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, et al. (2020) A causal framework for classical statistical estimands in failure-time settings with competing events; Statistics in Medicine; 39(8): 1199 – 1236