Causal Inference Methods for Secondary Analysis of Randomized Screening Trials

Sudipta Saha (PhD candidate)
Joint work with Olli Saarela & Amy Liu

Dalla Lana School of Public Health,
University of Toronto

June 4, 2018

Background

- Screening detects clinical illnesses (e.g. cancer) at an asymptomatic stage (Miettinen, 2011).
- The aim of screening is to reduce mortality from cancer by providing early treatments to the individuals detected by screening.
- The benefits of screening trials are studied through randomized trials, as non-experimental studies suffer from lead-time, length and confounding biases (Raffle and Gray, 2007, pp. 97).
- However, unlike therapeutic trials, participants in the screening trials are asymptomatic.
- Individuals without symptoms are randomly assigned to receive a series of screening examinations or standard care and subsequently followed for a prespecified period.

Background (Cont'd)

- The primary objective of screening trials is to estimate cancer-specific mortality reduction using intention-to-screen principle.
- Trials also often collect high-quality data, since numerous resources are invested in maintaining the standard of trials (shuang Ying, 2016).
- The high-quality data from trials provide opportunities to test secondary hypothesis.
- One such hypothesis concerns the effect of screening-induced early versus symptom-induced delayed treatments among early diagnosed cases of cancer.
- Miettinen has suggested that the reduction in case-fatality reduction, given the subpopulation of screening-diagnosed cases of cancer, is a definitive measure of benefit due to screening-induced early treatments (Miettinen, 2014).
- Miettinen also has proposed an estimator to estimate the case-fatality reduction (Miettinen, 2014).

Background (Cont'd)

- Instrumental variable (IV) approach has been used to estimate the causal effect in a subpopulation, namely the *compliers* (Angrist et al., 1996).
- In the screening context, the IV approach has been used to estimate the effect of screening in the subpopulation who are able to comply with the screening assignment (Roemeling et al., 2007).
- Altstein and co-authors proposed similar approaches to estimate the causal effects in a latent subpopulation, where the underlying subpopulation is defined based on baseline diagnostic test (Altstein et al., 2011; Altstein and Li, 2013).
- However, the underlying subpopulation in the screening context, the screening-diagnosed of cases of cancer, is partially latent and is accumulated over time.
- We have addressed these issue in our current work which is motivated by the case-fatality reduction.

Objectives

- To outline the early versus delayed treatments in the causal modeling framework, and re-derived Miettinen's estimator for proportional reduction in case-fatality.
- To use the IV principle to derive a new estimator for absolute reduction in case-fatality.
- To illustrate estimators using National Lung Screening Trial (NLST) data in the presence of competing risks and censoring.

Screening trials

- The immediate problem in screening trials is how to define the intervention of interest.
- In practice, participants with positive screening test undergo further laboratory testings (e.g. biopsy) before they receive early treatments for cancer.

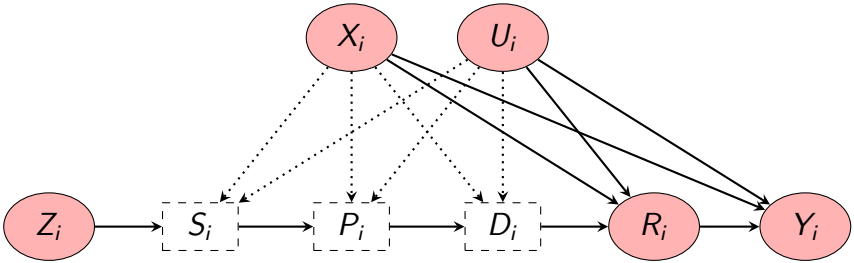
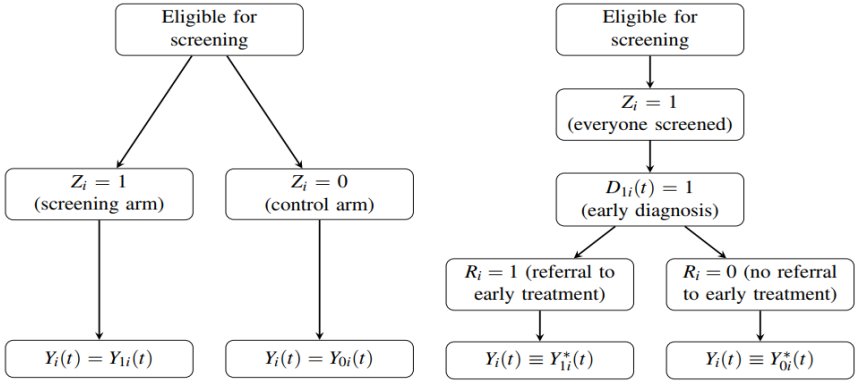


Figure: A schematic diagram of screening trials. Here Z_i is an indicator of randomized screening assignment, S_i is receiving screening, P_i is positive screening results, D_i is the early diagnosis state (e.g early or no early diagnosis), R_i is subsequent referral to early treatment, and Y_i is the cancer-specific mortality (e.g dead or alive) which is confounderd by observed covariates X_i and unmeasured covariates U_i.

Screening trials (Cont'd)



(a) Conventional screening trial
Contrast: $1 - \frac{E[Y_{1i}(t)]}{E[Y_{0i}(t)]}$
or $E[Y_{0i}(t)] - E[Y_{1i}(t)]$

(b) Miettinen's hypothetical intervention trial
Contrast: $1 - \frac{E[Y^*_{1i}(t)|D_{1i}(t)=1]}{E[Y^*_{0i}(t)|D_{1i}(t)=1]}$
or $E[Y^*_{0i}(t) - Y^*_{1i}(t) \mid D_{1i}(t) = 1]$

Figure: Illustration of conventional screening trial and Miettinen's hypothetical intervention trial in terms of potential outcome.

Screening trials (Cont'd)

- However, such hypothetical trials are completely unethical in practice.
- The hypothetical trial is helpful to define causal quantity of interest.
- The proportional reduction in case-fatality (PRCF) is $1 - \frac{E[Y_{1i}^*(t)|D_{1i}(t)=1]}{E[Y_{0i}^*(t)|D_{1i}(t)=1]}$.
- The absolute reduction in case-fatality (ARCF) is $E[Y_{0i}^*(t) - Y_{1i}^*(t) \mid D_{1i}(t) = 1]$.
- Our goal is to estimate these causal contrasts using data from screening trials.

The probability of being helped by screening

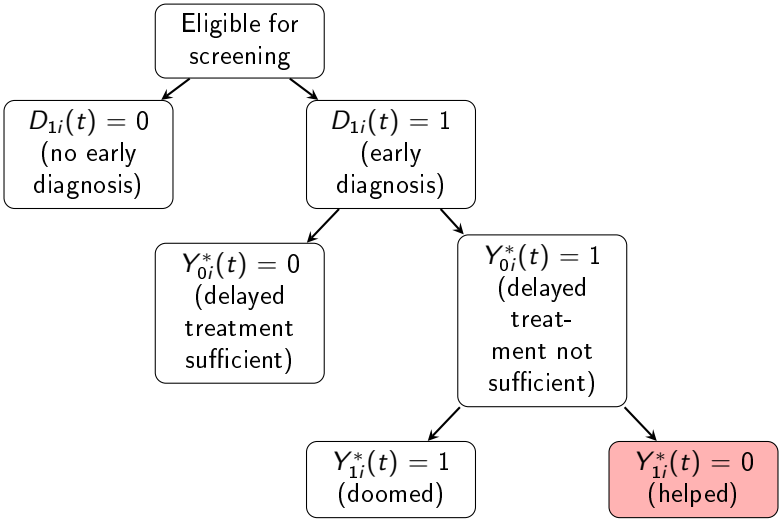


Figure: A probability tree illustrate decomposing the joint probability of benefiting from screening into conditional probabilities.

Covariate conditional versions of case-fatality

- We introduce the covariates to adjust for possible covariate-dependent censoring.
- Covariate-conditional proportional case-fatality reduction is,

$$1 - \frac{E[Y_{1i}^*(t) = 1 \mid D_{1i}(t) = 1, X_i]}{E[Y_{0i}^*(t) = 1 \mid D_{1i}(t) = 1, X_i]}$$

- The corresponding estimator is,

$$\frac{P(Y_i(t) = 1 \mid Z_i = 0, X_i) - P(Y_i(t) = 1 \mid Z_i = 1, X_i)}{P(Y_i(t) = 1 \mid Z_i = 0, X_i) - P(Y_i(t) = 1, D_i(t) = 0 \mid Z_i = 1, X_i)}$$

- Similarly, the covariate-conditional absolute case-fatality reduction is,

$$E[Y_{0i}^*(t) - Y_{1i}^*(t) \mid D_{1i}(t) = 1, X_i]$$

- And the corresponding estimator is,

$$\frac{E[Y_i(t) \mid Z_i = 0, X_i] - E[Y_i(t) \mid Z_i = 1, X_i]}{E[D_i(t) \mid Z_i = 1, X_i]}$$

Estimation of PCFR

- The quantity $P(Y_i(t) = 1 \mid Z_i = 1, X_i)$ is estimated by fitting Fine & Grey model in the screening arm and estimating the cumulative incidence of cancer-specific mortality.
- Similarly, the quantity $P(Y_i(t) = 1 \mid Z_i = 0, X_i)$ is estimated from the control arm.
- Also, the quantity $P(Y_i(t) = 1, D_i(t) = 0 \mid Z_i = 1, X_i)$ is estimated from the cumulative incidence of cancer-specific death before early diagnosis in the screening arm.

Estimation of ACFR

- Similarly, quantities $E[Y_i(t) \mid Z_i = 1, X_i]$ and $E[Y_i(t) \mid Z_i = 0, X_i]$ are estimated from the cumulative incidence of cancer-specific mortality in the screening and control arms, respectively.
- Also, the quantity $E[D_i(t) \mid Z_i = 1, X_i]$ is estimated from the cumulative incidence of early diagnosis in the screening arm.

Marginal effects under covariate-dependent censoring

- To estimate the marginal effects from conditionals, we derived formulas by averaging over covariate values of X.
- For example, the marginal effect of absolute case-fatality reduction is,

$$
\begin{aligned}
ACFR &= \frac{\int_x \{E[Y_{0i}^*(t) = 1 \mid D_{1i}(t) = 1, x] - E[Y_{1i}^*(t) = 1 \mid D_{1i}(t) = 1, x]\} f(x \mid D_{1i}(t) = 1)\, \mathrm{d}x}{\int_x E[Y_{0i}^*(t) = 1 \mid D_{1i}(t) = 1, x] f(x \mid D_{1i}(t) = 1)\, \mathrm{d}x} \\
&= \frac{1}{|\{i : Z_i = 1, D_{1i}(t) = 1\}|} \sum_{\{i: Z_i = 1, D_{1i}(t) = 1\}} \frac{E[Y_{0i}(t) \mid X_i] - E[Y_{1i}(t) \mid X_i]}{E[D_{1i}(t) \mid X_i]}
\end{aligned}
$$

National Lung Screening Trial

- To illustrate the methodology, we use National Lung Screening Trial (NLST) data.
- A total of $53,452$ high-risk individuals are randomized to receive low-dose helical CT (the screening arm) or standard chest X-ray (the control arm).
- Three rounds of annual screenings were provided to individuals randomized to the screening arm.
- The individuals are followed-up for 7 years.

Descriptive Statistics

- We illustrate how different quantities of the estimators behave as a function of follow-up time.

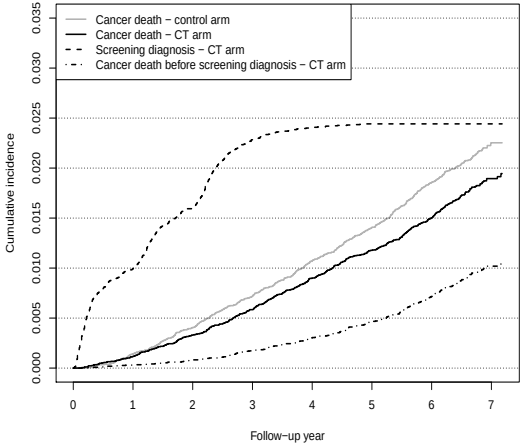
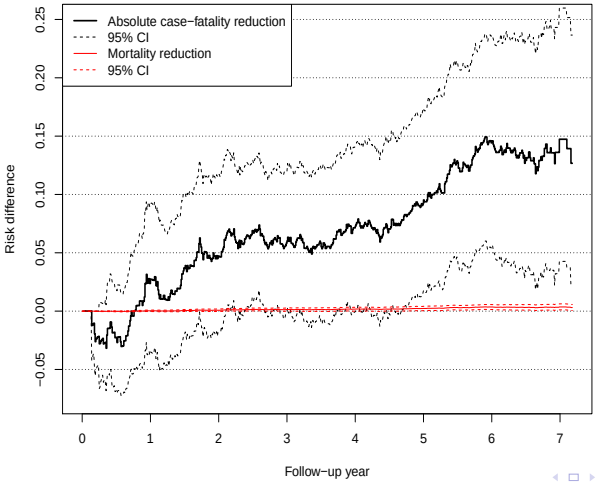


Figure: Cumulative incidences of lung cancer death in control and CT arms, along with early diagnosed lung cancer, and lung cancer death before early diagnosis in the CT arm.

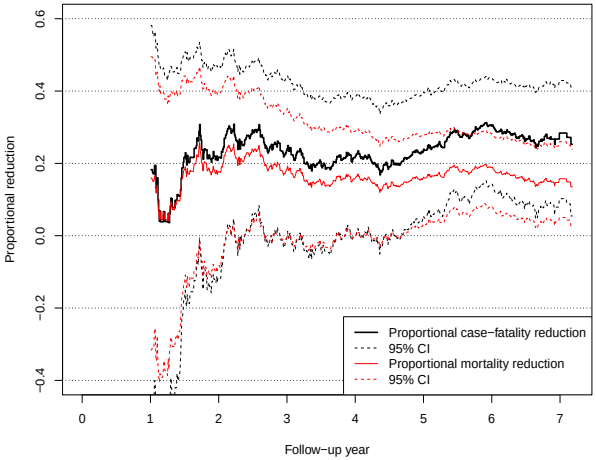
Results

- We illustrate how our proposed measure (absolute reduction in case-fatality) behaves as a function of follow-up window.
- We contrasted our measure with the ITT (i.e. $E(Y_{0i}(t)) - E(Y_{1i}(t))$).



Results (Cont'd)

- We also illustrate how Miettinen's measure, the proportional reduction in case-fatality, behaves as a function of follow-up window.
- We contrasted our measure with the corresponding ITT (i.e. $1 - \frac{E(Y_{1i}(t))}{E(Y_{0i}(t))}$)



Results (Cont'd)

- We reproduce the original results of NLST, which was the 20% proportional reduction at year 6.
- For the NLST, we found that around 7 individuals are needed to be treated early to prevent a cancer death.
- In contrast, we need around 333 individuals to be invited to screen to prevent a cancer death.

References

Lily L Altstein and Gang Li. Latent subgroup analysis of a randomized clinical trial through a semiparametric accelerated failure time mixture model. *Biometrics*, 69:52–61, 2013.

Lily L Altstein, Gang Li, and Robert M Elashoff. A method to estimate treatment efficacy among latent subgroups of a randomized clinical trial. *Statistics in medicine*, 30:709–717, 2011.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

Olli Sakari Miettinen. *Epidemiological research: terms and concepts*. Springer Science & Business Media, 2011.

OS Miettinen. Screening for breast cancer: what truly is the benefit? *Can J Public Health*, 104(7): 435–436, 2014.

Angela E Raffle and JA Muir Gray. *Screening: evidence and practice*. Oxford University Press, 2007.

Stijn Roemeling, Monique J Roobol, Suzie J Otto, Dik F Habbema, Claartje Gosselaar, Jan J Lous, Jack Cuzick, and Fritz H Schröder. Feasibility study of adjustment for contamination and non-compliance in a prostate cancer screening trial. *The Prostate*, 67:1053–1060, 2007.

Gui shuang Ying. Secondary analysis of clinical trials data- a biostatistician's experience. Center for Preventive Ophthalmology and Biostatistics, University of Pennsylvania, 2016. URL `http://www.sctweb.org/public/meetings/2016/slides/Ying-Statistical%20Lessons.pdf`.