

# CANTRANce Screening Aim

Leslie Mallinger

lmalling@fhcrc.org

Etzioni Lab, Fred Hutchinson Cancer Research Center

August 9, 2013

## 1 Background

The fundamental assumption underlying cancer screening is that early detection of a tumor will lead to correspondingly improved prognosis and disease-specific survival. A key consequence of early detection is the reduction in the incidence of advanced cancers associated with the screening test. Using user-specified inputs, CANTRANce will link comparative effectiveness study results on the reduction in the incidence of advanced tumors in the presence of a screening test into changes in disease-specific mortality and years of life saved under screening.

The general CANTRANce model for screening interventions is first discussed, then an example is provided using data from the National Lung Screening Trial (NLST).

## 2 Model Components

### 2.1 Model inputs

#### 2.1.1 Comparative effectiveness study data

The comparative effectiveness study data are entered into CANTRANce in one of two ways: 1) as individual-level data describing patient characteristics, or 2) as marginal distributions for patient characteristics. CANTRANce assumes a study design in which individual-level characteristics are linked with stage at clinical incidence, and a stage at incidence due to screening is assigned to each individual during the modeling process.

If individual-level data are available, the user will provide a data set in which the rows are individuals and columns contain data for each applicable patient characteristic.

If marginal distributions are to be used, the user may specify truncated normal distributions for continuous covariates and/or proportional distributions for categorical covariates.

Regardless of entry method, required inputs are:

1. *male*: sex of individual

2. *age* or *agegroup*: age of individual at time of clinical incidence. If *age* is provided, users may specify desired age groups by which to perform analyses. If *agegroup* is provided, single year ages are assigned to individuals by sampling from a uniform distribution within the given range.
3. *study\_year*: year of clinical incidence
4. *stage*: stage at time of **clinical** incidence

Study data will often also include covariates that are expected to affect risk of disease-specific death or other-cause death.

### 2.1.2 User specifications and assumptions

Several choices must be made in how to use the comparative effectiveness study results to project mortality:

1. **Population.** The user can choose to model the impact on the original study population, or on a population that is of a different size and/or has a different composition of covariates.
2. **Number of simulations.** Uncertainty due to modeling assumptions is computed by summarizing the results of many simulations. Users may choose to run between 50 and 1,000 simulations.
3. **Length of time to project.** CANTRANce will estimate mortality from the time of treatment through the length of time  $t$  requested by the user.
4. **Population-level stage distribution in presence of screening.** Stage at screened incidence is assigned during the modeling process based on clinical stage and the population-level stage distribution in the presence of screening.

In addition to the proportion of individuals in each stage in the presence of screening, the user must provide the order of the stages during natural progression of the disease, where 1 denotes the earliest, or least invasive stage of the disease, 2 denotes the next progression, etc.

CANTRANce assigns stage at screened incidence using a cascade approach beginning with individuals in the most invasive stage of the disease at the time of clinical incidence. Some of these individuals will remain in the most invasive stage during screen-detection, and others will be shifted to an earlier stage. Thus, the appropriate number are sampled to remain in the highest stage, and the remainder are temporarily re-assigned to the second highest stage. Of those individuals in the second highest stage, a number are sampled to remain in that stage and the remainder are re-assigned to the next stage down in severity. The process continues for all remaining stages. In this way, some individuals undergo a “stage shift” due to screening, while the stage of others is unaffected by screening.

5. **Time from clinical incidence to cause-specific death.** CANTRANce models time from clinical incidence to cause-specific death assuming an exponential process for the baseline cause-specific survival curve, with the option of applying modifications to the baseline curve for specified covariates.

The user specifies the baseline cause-specific survival curve based on  $k$ -time survival, median survival, mean survival, or mortality rate in the absence of competing risks. Modified survival curves for up to three covariates may be specified either via hazard ratios or via survival statistics directly.

6. **Time from clinical incidence to other-cause death.** For a typical population, time to other-cause death is approximated using US cohort lifetables matched to each individual by sex and year of birth. If the study population is thought to be healthier or less healthy than the US population, the user can specify a hazard ratio to adjust the lifetable-based survival estimates.

## 2.2 Model Outputs

Model inputs are processed according to the steps in Figure 1. From treatment to the time  $t$  specified by the user, CANTRANce will return:

1. A table of mean, median, and  $t$ -time all-cause survival, and mean, median, and  $t$ -time net and crude disease-specific survival, presented as time from treatment to the outcome of interest. While all-cause survival does not distinguish between disease-specific and other-cause death, net disease-specific survival censors other-cause deaths and represents the hypothetical scenario of survival from the disease in the absence of other-cause death, assuming that the two types of death are independent. Crude disease-specific survival describes survival as it is observed; other-cause death is considered a competing event that precludes death from the disease, so other-cause deaths are not considered at risk after their death. It follows from this definition of at-risk time that crude disease-specific survival is always higher than net disease-specific survival; in the latter, the censoring of other-cause death implies the possibility of disease-specific death at future times<sup>2</sup>.
2. Survival curves for all-cause survival, net disease-specific survival, and crude disease-specific survival.
3. A table of person-years saved by the screening intervention among the study population during the period from enrollment to time  $t$ .
4. A table of cause-specific mortality statistics describing the mortality trends observed in the simulated population, stratified by covariate and mode of detection when applicable. This table allows the user to verify that simulated times to event are consistent with the times to event specified as user inputs.
5. A table of statistics describing the time by which cause-specific death was delayed, stratified by stage at clinical diagnosis.

## 2.3 Model Validation

We should brainstorm ways to validate the model.

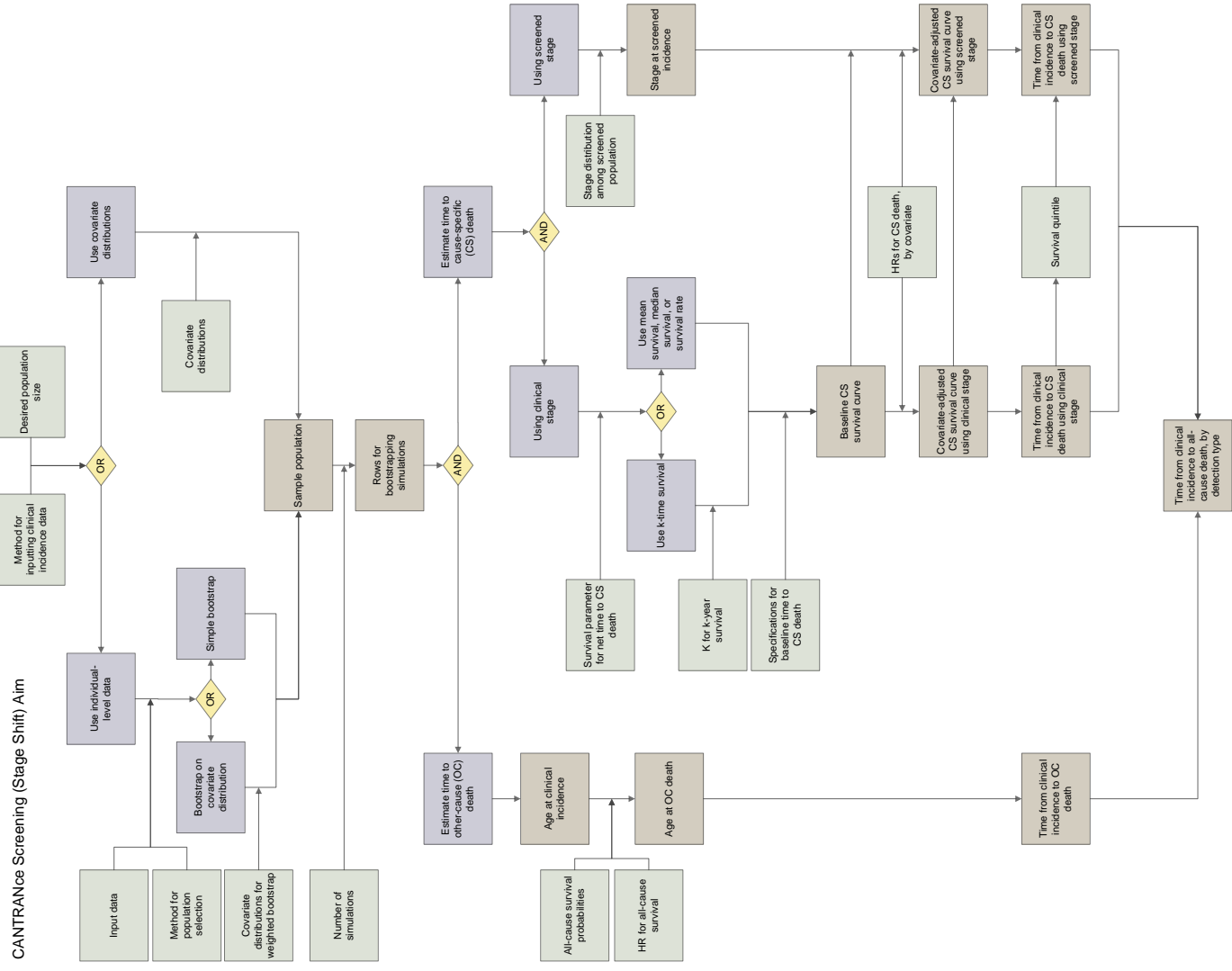


Figure 1: Flowchart representing the steps in the CANTRANce model for screening interventions. Green=user inputs. Purple=actions. Brown=intermediate results.

### 3 Example: National Lung Screening Trial (NLST)

The facility with which CANTRANce can produce meaningful mortality projections can be seen in an application to a comparative effectiveness study of the use of lung cancer screening with low-dose helical computed tomography (CT) compared to the current standard of screening with chest radiography<sup>1</sup>.

Previous studies of screening with chest radiography showed no evidence of reduction in lung cancer mortality, but preliminary results of screening with low-dose CT have been promising. The NLST aimed to determine whether screening with low-dose CT would reduce lung cancer mortality among individuals at high risk of lung cancer<sup>1</sup>.

Over the approximately 7-year follow-up period, 24.2% of low-dose CT screens and 6.9% of radiography screens were positive. Lung cancer incidence was 645 per 100,000 person-years in the CT screening group and 572 per 100,000 person-years in the radiography group. Deaths were 247 and 309 per 100,000 person-years in the CT and radiography groups, respectively<sup>1</sup>.

This example will illustrate the use of CANTRANce to estimate the effect of low-dose CT screening on lung cancer mortality via detection in earlier stages of disease. Stage-specific mortality data were obtained from SEER.

#### 3.1 Model Inputs

##### 3.1.1 Comparative effectiveness study data

Study participants consisted of 1000 men assigned characteristics at clinical incidence according to the distributions in Tables 1-3. The date of study entry was assumed to be 2003 for all participants.

Individuals were assigned to five-year age groups between 55 and 74 according to the proportions specified in Table 1, then assigned single-year ages from a uniform distribution within the five-year range. Tables 2 and 3 give the proportion of individuals assigned to each sex and clinical stage.

Age Group	Proportion
55-59	0.428
60-64	0.306
65-69	0.178
70-74	0.088

Table 1: Population characteristics for age (years)

Sex	Proportion
Female	0.4
Male	0.6

Table 2: Proportions of individuals in each sex

Note that CANTRANce will assume age, sex, and stage to be independent for simulation purposes because they are specified separately rather than as joint distributions. Although this may

Clinical Stage	Proportion
I	0.31
II	0.08
III	0.25
IV	0.36

Table 3: Proportions of individuals in each clinical stage

not directly reflect the study population, it is the only approach available without more specific information about the study population.

### 3.1.2 User specifications and assumptions

1. **Population.** The model population for each simulation consisted of 1000 individuals with clinical incidence in 2003.
2. **Number of simulations.** 50
3. **Length of time to project.** Mortality was projected for 20 years from the time of treatment.
4. **Population-level stage distribution in presence of screening.** The proportion of individuals presenting in each stage in the presence of screening is shown in Table 4, as is the order of stage progression.

Order	Screened Stage	Proportion
1	I	0.50
2	II	0.07
3	III	0.21
4	IV	0.22

Table 4: Proportions of individuals in each stage in presence of screening

Recall that the stage distribution among individuals at clinical incidence was 31.1% in stage I, 7.9% in stage II, 24.9% in stage III, and 36.1% in stage IV.

The stage shift was implemented as follows: of the 36.1% that were clinically incident in stage IV, 60.1% (21.7% of the total population) were sampled to stay in stage IV for their screened incidence, and the remaining 39.9% (14.4% of the total population) were temporarily shifted to stage III, bringing the population total in stage III to  $24.9\% + 14.4\% = 39.3\%$ . Then, of the 39.3% of individuals in stage III, 54.2% (21.3% of the total population) were sampled to stay in stage III for their screened incidence, and the remaining 45.8% (18.0% of the total population) were shifted to stage II. The process continued until all individuals were assigned a stage at screened incidence.

Note that individuals with clinical stage IV disease could be assigned to any of the four screened stages, but individuals with clinical stage I disease stayed in stage I upon screening.

5. **Time from clinical incidence to cause-specific death.** Data on lung cancer survival following clinical incidence were obtained from the SEER 9 registries. The study population was restricted to individuals aged 55-74 years of age, diagnosed between 2000 and 2005 with

Clinical Stage	$\lambda$
I	0.0851
II	0.1593
III	0.3950
IV	0.7950

Table 5: Stage-specific mortality rate ( $\lambda$ ) following clinical incidence in SEER population

Sex	$\lambda$
Male	0.4147
Female	0.3501
Both	0.3848

Table 6: Sex-specific mortality rate ( $\lambda$ ) following clinical incidence in SEER population

cancer of the lung and bronchus and known stage at incidence. Lung cancer screening with low-dose CT was rare enough during that time that stage at incidence was assumed to be representative of stage at clinical incidence.

Net cause-specific survival was obtained for each year up to ten years following incidence, calculated separately by stage and by sex. Each  $k$ -year survival value was converted to  $\lambda$  for an exponential distribution ( $\lambda = \frac{-\log(ksurv)}{k}$ ) and the average  $\lambda$  across years was used as the estimate for stage- and sex-specific mortality rates. Table 5 gives the average  $\lambda$  by stage, and Table 6 gives the average  $\lambda$  by sex, as obtained from SEER.

For the baseline survival curve, cause-specific survival rate was 0.395, that observed in stage III cases in SEER. Survival statistics were specified separately by clinical stage and sex (Table 7). Because CANTRANce applies survival parameters multiplicatively when present for multiple covariates, the stage-specific rates from SEER were used directly in CANTRANce while the sex-specific hazard ratio was used as a ratio to the whole population (both sexes).

For this example, stage-specific survival was specified in terms of a cause-specific survival rate, which CANTRANce used to calculate survival rates and corresponding hazard ratios to modify baseline survival. Sex-specific survival was already provided by the user in terms of hazard ratios. The hazard ratio for each individual was calculated as the product of hazard ratios for each group to which he or she belonged.

Risk	Group	HR
Clinical Stage	I	0.22
	II	0.40
	III	1.00
	IV	2.01
Sex	Female	0.91
	Male	1.08

Table 7: Hazard ratio for cause-specific death, by clinical stage and sex

- 6. Time from intervention to other-cause death.** Time to other-cause death was approximated using unadjusted US cohort lifetables matched by sex and year of birth.

### 3.2 Model Outputs

Measure	Mode	Mean.Survival	Median.Survival
All-Cause Survival	Clinical	4.170 (3.830-4.470)	1.910 (1.740-2.090)
All-Cause Survival	Screened	5.490 (5.200-5.850)	2.860 (2.680-3.060)
Net Survival	Clinical	5.210 (4.710-5.690)	2.030 (1.840-2.220)
Net Survival	Screened	7.070 (6.690-7.590)	3.140 (2.940-3.450)
Crude Survival	Clinical		2.070 (1.850-2.240)
Crude Survival	Screened		3.250 (3.010-3.570)

Table 8: Mean and median survival for three survival metrics, by mode of detection. Parentheses give 2.5% and 97.5% quantiles across simulations. Survival times are measured from time of clinical incidence.

Measure	Mode	Five.Year	Ten.Year	Twenty.Year
All-Cause Survival	Clinical	0.253 (0.229-0.274)	0.123 (0.104-0.138)	0.0305 (0.0229-0.0397)
All-Cause Survival	Screened	0.354 (0.334-0.373)	0.185 (0.169-0.208)	0.0469 (0.0369-0.0577)
Net Survival	Clinical	0.276 (0.252-0.296)	0.153 (0.131-0.171)	0.0586 (0.0436-0.0707)
Net Survival	Screened	0.386 (0.368-0.405)	0.230 (0.210-0.255)	0.0911 (0.0746-0.106)
Crude Survival	Clinical	0.293 (0.274-0.315)	0.186 (0.166-0.206)	0.121 (0.105-0.136)
Crude Survival	Screened	0.403 (0.382-0.420)	0.267 (0.250-0.288)	0.172 (0.156-0.190)

Table 9: 5-year, 10-year, and 50-year survival for three survival metrics, by mode of detection. Parentheses give 2.5% and 97.5% quantiles across simulations. Survival times are measured from time of clinical incidence.

Table 8 gives the number of years for mean and median all-cause survival and net and crude cause-specific survival from the time of clinical incidence. 5-year, 10-year, and 50-year survival proportions are provided in Table 9. Blank cells indicate that the metric could not be calculated given available information. Corresponding survival curves are shown in Figure 2.

During the 5 years following treatment, 1220 (1001 to 1434) person-years were saved by screening.

Table 10 shows five-year cause-specific survival by risk group and mode of diagnosis, as specified by the user and observed in the simulated population. Observed values are close to those designated in the inputs, and if only one covariate is chosen (not shown), the values are nearly identical.

Table 11 shows the delay in cause-specific mortality due to the stage shift induced by screening, by stage at clinical incidence. As expected, screening has no impact on individuals with localized clinical stage, because they remained in localized stage even under screen-detection.

## References

- [1] Denise R. Aberle, Amanda M. Adams, Christine D. Berg, William C. Black, Jonathan D. Clapp, Richard M. Fagerstrom, Ilana F. Gareen, Constantine Gatsonis, Pamela M. Marcus, JoRean D. Sicks, and Natl Lung Screening Trial Res Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *NEW ENGLAND JOURNAL OF MEDICINE*, 365(5):395–409, AUG 4 2011.



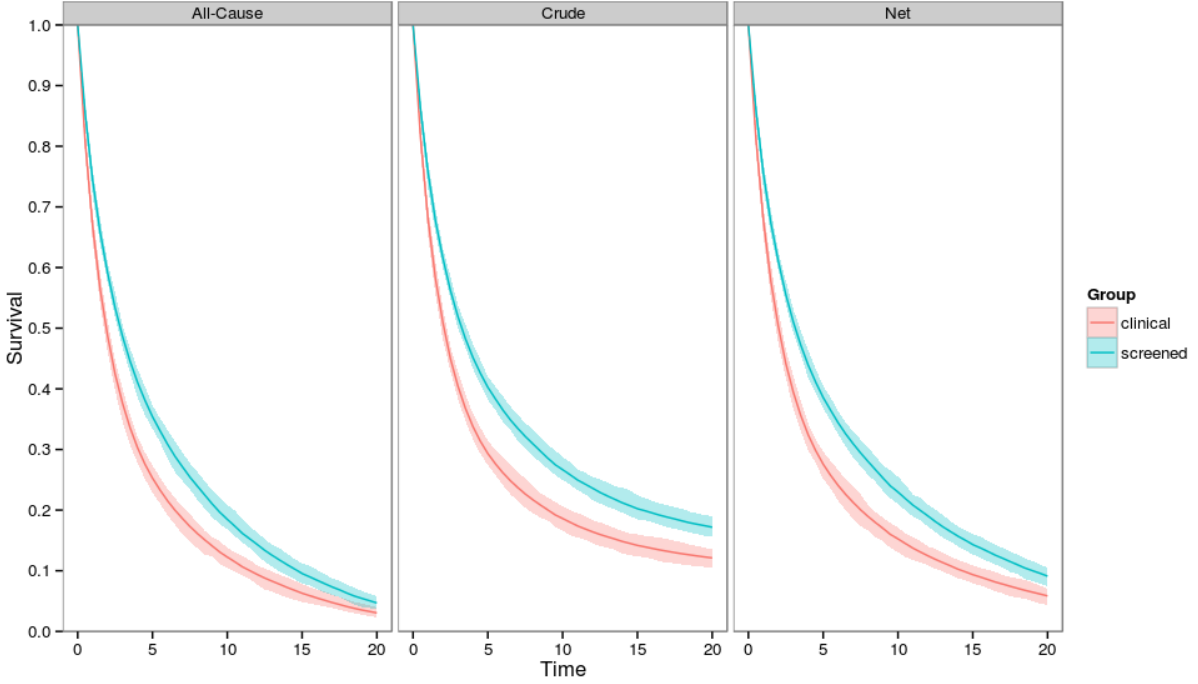


Figure 2: All-cause, crude cause-specific, and net cause-specific survival curves, by mode of detection. Survival times are measured from time of clinical incidence.

Risk	Group	Mode	User Input	Observed
All Individuals	-	Clinical		0.224 (0.209-0.245)
	-	Screened		0.163 (0.155-0.173)
Stage	I	-	0.085	0.0917 (0.0846-0.102)
	II	-	0.159	0.161 (0.131-0.188)
	III	-	0.395	0.403 (0.364-0.455)
	IV	-	0.795	0.802 (0.718-0.876)
Sex	Female	-	0.359	0.172 (0.153-0.193)
	Male	-	0.426	0.204 (0.190-0.221)

Table 10: 5-year cause-specific survival specified by the user and observed in the simulated population, by risk group and mode of diagnosis.

Clinical Stage	Estimate
I	0.000 (0.000-0.000)
II	4.053 (3.117-4.998)
III	3.420 (2.759-4.502)
IV	1.889 (1.399-2.465)

Table 11: Delay in cause-specific mortality due to screening-induced stage shift, by stage at clinical incidence.

- [2] Haesook T Kim. Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 13(2 Pt 1):559–565, January 2007. PMID: 17255278.