DATA INTEGRATION TO IMPROVE PROSTATE CANCER MORTALITY PREDICTION

ELIZABETH CHASE
BIOSTAT 803 SEMINAR
OCT. 22, 2021

ACKNOWLEDGEMENTS

- Biostatistics: Dr. Matthew Schipper, Elise Covert, Dr. Krithika Suresh
- Radiation Oncology: Dr. Bob Dess, Dr. Will Jackson, Dr. Daniel Spratt
- Cancer Data Access System staff

MOTIVATION

- Most patients diagnosed with prostate cancer will not die of their cancer, instead dying of old age or other comorbidities.

  → ***competing risk***

- As a result, in order to predict a patient's risk of dying of prostate cancer, we should model the competing risks carefully.

- It can be difficult to quantify the risk of ***prostate cancer specific mortality (PCSM)*** relative to OCM and make informed choices about which is "higher priority" for any particular patient.

- We should also consider the competing risks in making treatment decisions: because ***other-cause mortality (OCM)*** makes up the bulk of total mortality, giving a patient a treatment that helps their cancer while hurting their overall health may have an outsize negative effect.

MOTIVATION

- Current NCCN treatment guidelines recommend that clinicians combine life expectancy predictions from the Social Security Administration (SSA) with their clinical gestalt to obtain estimates of other-cause mortality, and then use that to inform treatment decisions.
- Previous research suggests that the SSA life tables are not accurate for prostate cancer patients, and that clinicians are not particularly good at estimating other-cause life expectancy.
- Goal: To more rigorously model other-cause mortality (OCM) in prostate cancer patients, and to combine it with estimates of prostate-cause specific mortality (PCSM) to guide treatment decisions.

BACKGROUND: STAR-CAP STAGING SYSTEM

- STAR-CAP is a new prostate cancer staging system developed by a team at U-M Radiation Oncology.
- It features 6 key predictors: age, Gleason score, pre-treatment PSA, T stage, N stage, and percent of biopsy cores that were positive.
- Fine and Gray regression was used to fit the model and obtain predictions.
- Groups men into 9 STAR-CAP stages, each of which provides a predicted risk of PCSM over time, up to 15 years out from diagnosis.
- In two external validation sets, the STAR-CAP model performed very well, with an AUC of about **0.8**.

OCCAM MODEL: TRAINING DATA

- National Health and Nutrition Examination Survey (NHANES) 1999-2010.
- Mortality follow-up through Dec. 31, 2014.
- NOT a prostate cancer patient population.
- Restrict to men older than 40, free of non-prostate cancer, with complete data for training predictors.
- Final sample consisted of 2,420 men with 459 deaths over a mean follow-up of 103.7 months (8.6 years).
- For more information: https://www.cdc.gov/nchs/nhanes/index.htm

OCCAM MODEL: TRAINING DATA CHARACTERISTICS

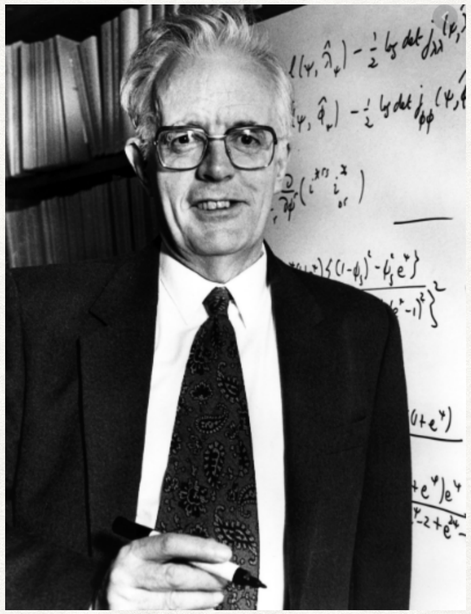
- Mean age: 59.4 years
- 17.0% non-Hispanic Black; 26.7% other race; 56.3% non-Hispanic White
- 24.8% of respondents had not completed high school
- 78.2% of respondents had a BMI over 25
- 22.2% were current smokers; 18.7% had diabetes; 48.3% had hypertension
- 127 patients (5.2%) had a previous prostate cancer diagnosis

OCCAM MODEL: MODEL BUILDING

- Considered the covariates:
    - Demographics: age, race, educational attainment, marital status
    - Comorbidities: arthritis, chronic bronchitis, diabetes, emphysema, hypertension, previous heart attack, liver disease, previous stroke, prostate cancer
    - Other risk factors: smoking status, overweight/obese

OCCAM MODEL: MODEL BUILDING

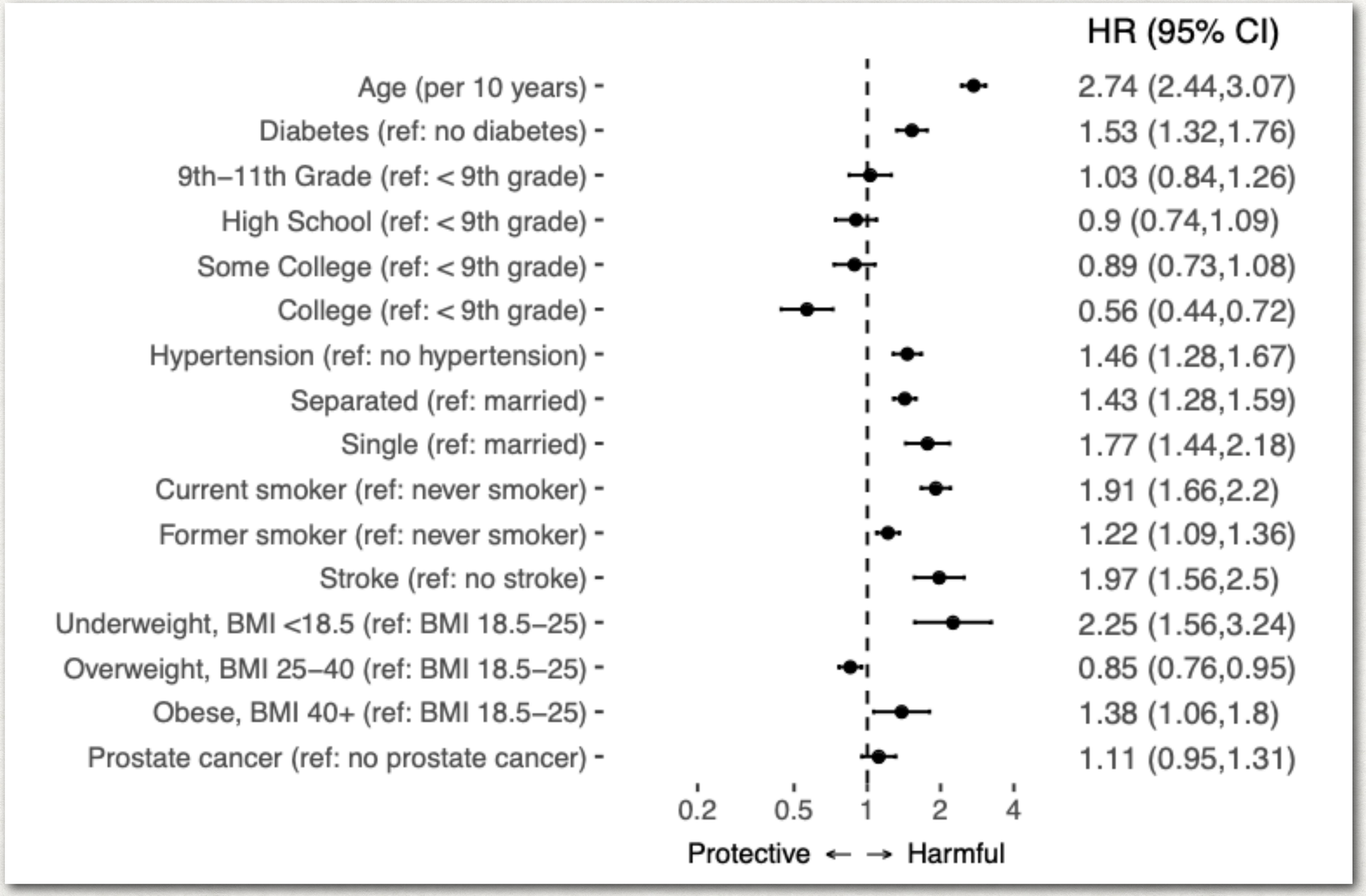
- Considered three modeling strategies:
  - Cox proportional hazards modeling
  - Survival random forest modeling
  - Parametric spline survival modeling



OCCAM MODEL: SENSITIVITY ANALYSES

- Recall that NHANES is not a prostate cancer population—is this model remotely correct?
- Looked at interactions between having prostate cancer and all other predictors.
- Looked at length of time from diagnosis as a predictor, and interactions.
- Outputted linear predictors from our final model and used that as a predictor along with prostate cancer, length of time from diagnosis, and interactions.
- Still included prostate cancer as a predictor in all candidate models.

OCCAM MODEL


HR (95% CI)
Age (per 10 years)
2.74 (2.44,3.07)
Diabetes (ref: no diabetes)
1.53 (1.32,1.76)
9th–11th Grade (ref: < 9th grade)
1.03 (0.84,1.26)
High School (ref: < 9th grade)
0.9 (0.74,1.09)
Some College (ref: < 9th grade)
0.89 (0.73,1.08)
College (ref: < 9th grade)
0.56 (0.44,0.72)
Hypertension (ref: no hypertension)
1.46 (1.28,1.67)
Separated (ref: married)
1.43 (1.28,1.59)
Single (ref: married)
1.77 (1.44,2.18)
Current smoker (ref: never smoker)
1.91 (1.66,2.2)
Former smoker (ref: never smoker)
1.22 (1.09,1.36)
Stroke (ref: no stroke)
1.97 (1.56,2.5)
Underweight, BMI <18.5 (ref: BMI 18.5–25)
2.25 (1.56,3.24)
Overweight, BMI 25–40 (ref: BMI 18.5–25)
0.85 (0.76,0.95)
Obese, BMI 40+ (ref: BMI 18.5–25)
1.38 (1.06,1.8)
Prostate cancer (ref: no prostate cancer)
1.11 (0.95,1.31)
0.2
0.5
1
2
4
Protective ← → Harmful

OCCAM MODEL: VALIDATION DATA

NATIONAL CANCER INSTITUTE
Division of Cancer Prevention

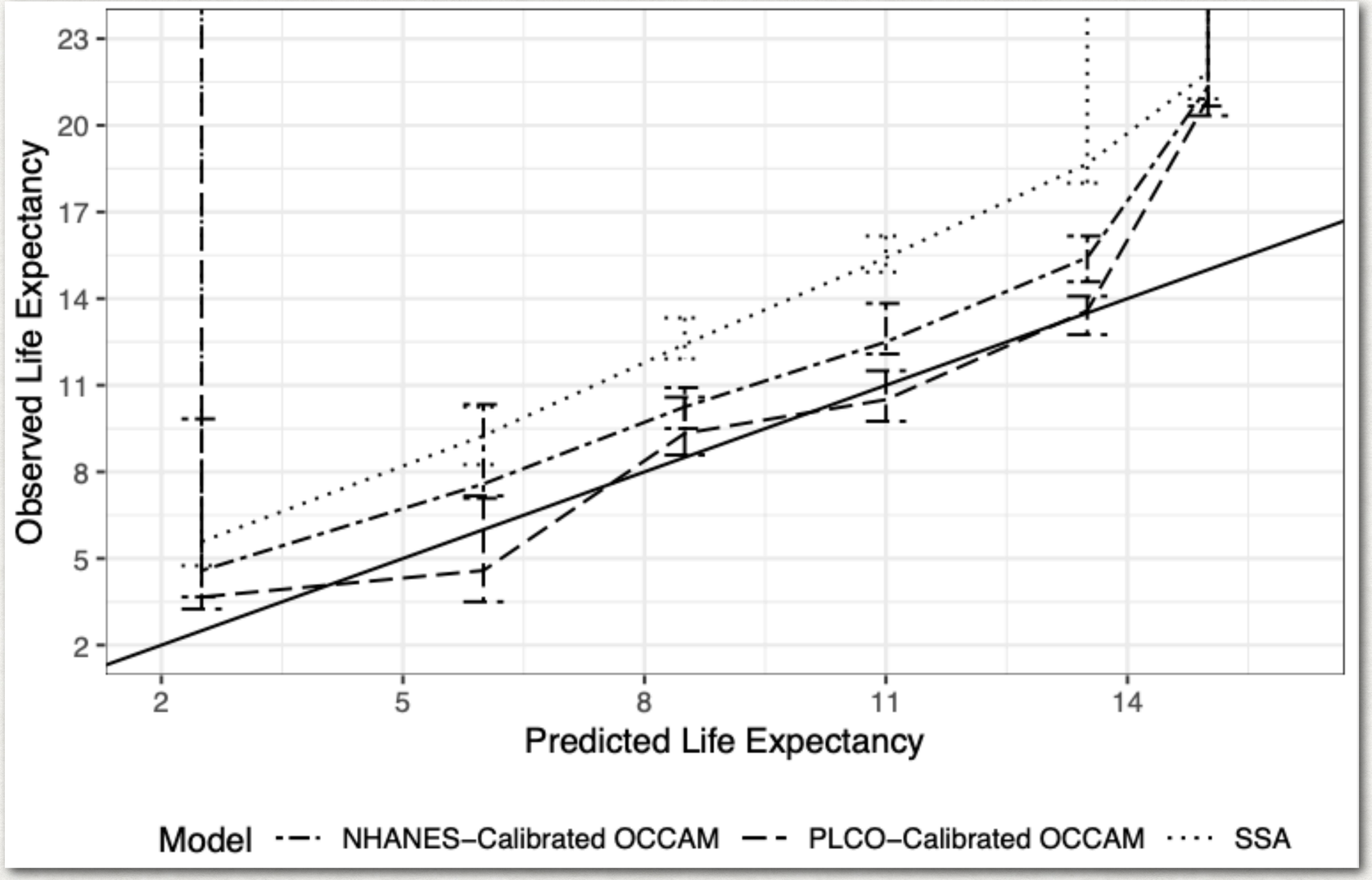
- Extremely different from NHANES training sample
- Mean age: 69.5 years
- 5.8% non-Hispanic Black, 5.3% other race, 89.0% non-Hispanic White
- 7.4% of the sample had not completed a high school degree
- 71.5% of respondents had a BMI over 25
- 9.0% were current smokers; 6.4% had diabetes; 33.4% had hypertension
- Only characteristic on which they weren't significantly different was having a previous heart attack; about ~12% in both samples

OCCAM MODEL

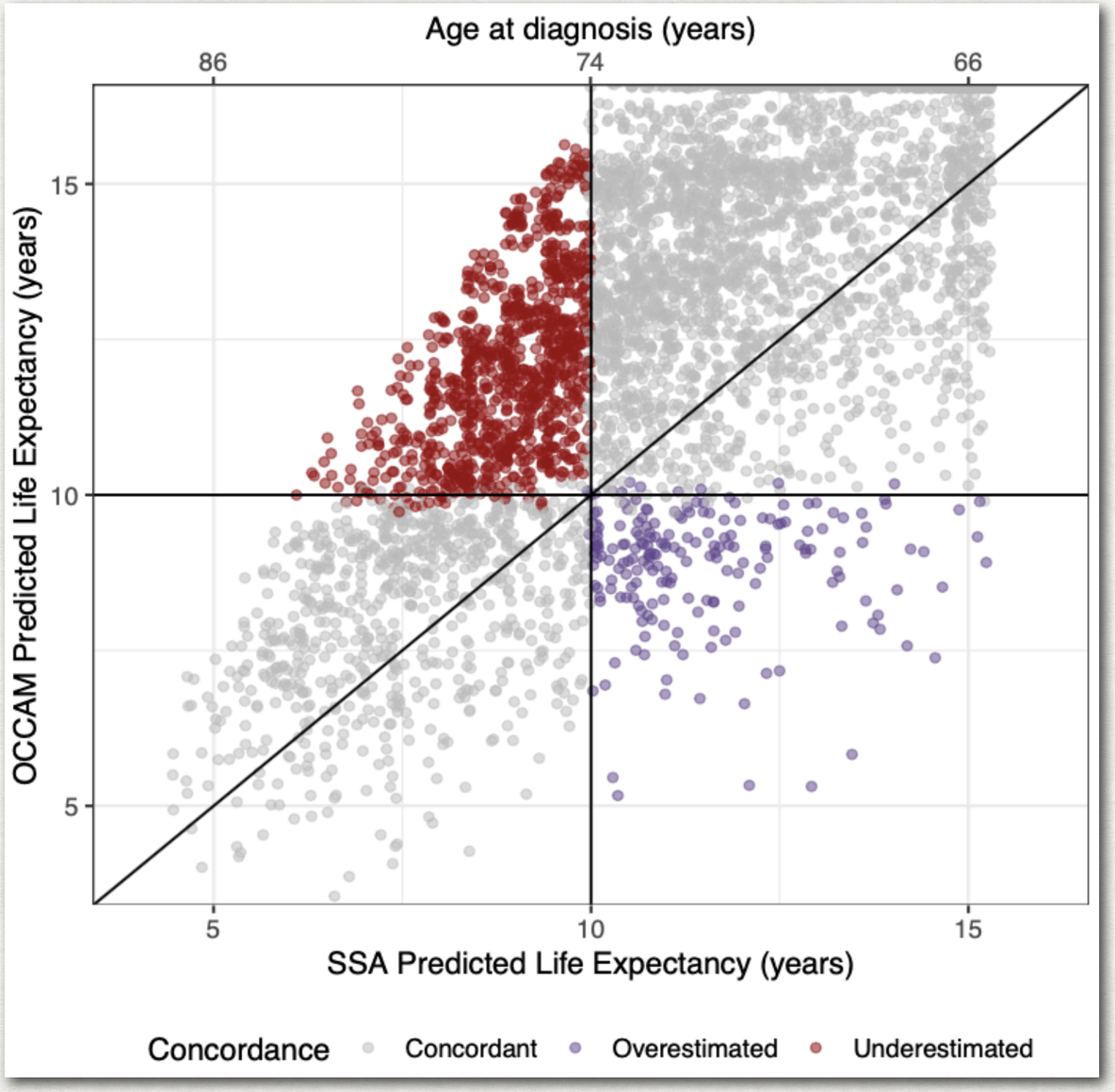
Metric	5 Years	10 Years	15 Years
Time-Dependent AUC	0.70	0.75	0.78

- Significantly outperformed the SSA life tables and the National Vital Statistics System's life expectancy estimates.
- Comparable or better performance to other OCM prediction tools, but with far fewer predictors (8).

OCCAM MODEL


Observed Life Expectancy
23
20
17
14
11
8
5
2
2
5
8
11
14
Predicted Life Expectancy
Model
NHANES–Calibrated OCCAM
PLCO–Calibrated OCCAM
SSA

OCCAM MODEL


Age at diagnosis (years)
86
74
66
15
10
5
OCCAM Predicted Life Expectancy (years)
5
10
15
SSA Predicted Life Expectancy (years)
Concordance
Concordant
Overestimated
Underestimated

GOALS, REVISITED

- Rigorously model OCM in prostate cancer patients.
- Integrate these OCM predictions with PCSM predictions to help clinicians better understand the role of competing risks in prostate cancer patients and the impact of their treatment decisions.

CHALLENGES

- We have two models (STAR-CAP and OCCAM), each predicting one component of mortality in prostate cancer patients.
- The data sources in which they were built/validated are very different:
    - NHANES isn’t a prostate cancer population, and it covers a less educated, less healthy, and less affluent population than what STAR-CAP studies. It only has comorbidity/demographic predictors.
    - PLCO is a prostate cancer population, but it’s an extraordinarily healthy and affluent population of prostate cancer patients. It has information on comorbidities and some prostate cancer predictors, but not all of the ones we need for STAR-CAP.
    - STAR-CAP is fairly representative of the general prostate cancer population, but it has no comorbidity/demographic predictors other than age and race.

APPROACH

- Step 0: Re-fit STAR-CAP to be cause-specific model rather than Fine and Gray.
- Step 1: Stack STAR-CAP and PLCO data into a combined dataset.

APPROACH

Patient ID	Dataset	Age	Time to Death	Event	T stage	N stage	% cores	PSA	Gleason	Race	Smoking	Marital	Educ	Diabetic	Stroke	Hypertension	BMI
1	STARCAP	65	180	2	T1a	N0	50	4	6	NHB							
2	STARCAP	72	95	1	T3b	N0	75	45	4+3	NHW							
3	STARCAP	58	180	0	T2	N0	81	0.3	3+4	Other							
4	STARCAP	83	24	2	T1c	N0	20	12	6	NHW							

APPROACH

Patient ID	Dataset	Age	Time to Death	Event	T stage	N stage	% cores	PSA	Gleason	Race	Smoking	Marital	Educ	Diabetic	Stroke	Hypertension	BMI
1	PLCO	55	200	0	T1a	N0		4	6	NHB	Former	Married	HS	No	No	No	24
2	PLCO	84	12	2	T1c	N0		10	6	Other	Current	Married	BS	Yes	No	Yes	26
3	PLCO	59	200	0	T2	N0		5	7	NHW	Former	Single	Pro	No	No	Yes	30
4	PLCO	75	60	1	T4	N1		195	10	NHB	Never	Married	BS	No	No	Yes	19.5

APPROACH

Patient ID	Dataset	Age	Time to Death	Event	T stage	N stage	% cores	PSA	Gleason	Race	Smoking	Marital	Educ	Diabetic	Stroke	Hypert ension	BMI
1	STARCAP	65	180	2	T1a	N0	50	4	6	NHB							
2	STARCAP	72	95	1	T3b	N0	75	45	4+3	NHW							
3	STARCAP	58	180	0	T2	N0	81	0.3	3+4	Other							
4	STARCAP	83	24	2	T1c	N0	20	12	6	NHW							
1	PLCO	55	200	0	T1a	N0		4	6	NHB	Former	Married	HS	No	No	No	24
2	PLCO	84	12	2	T1c	N0		10	6	Other	Current	Married	BS	Yes	No	Yes	26
3	PLCO	59	200	0	T2	N0		5	7	NHW	Former	Single	Pro	No	No	Yes	30
4	PLCO	75	60	1	T4	N1		195	10	NHB	Never	Married	BS	No	No	Yes	19.5

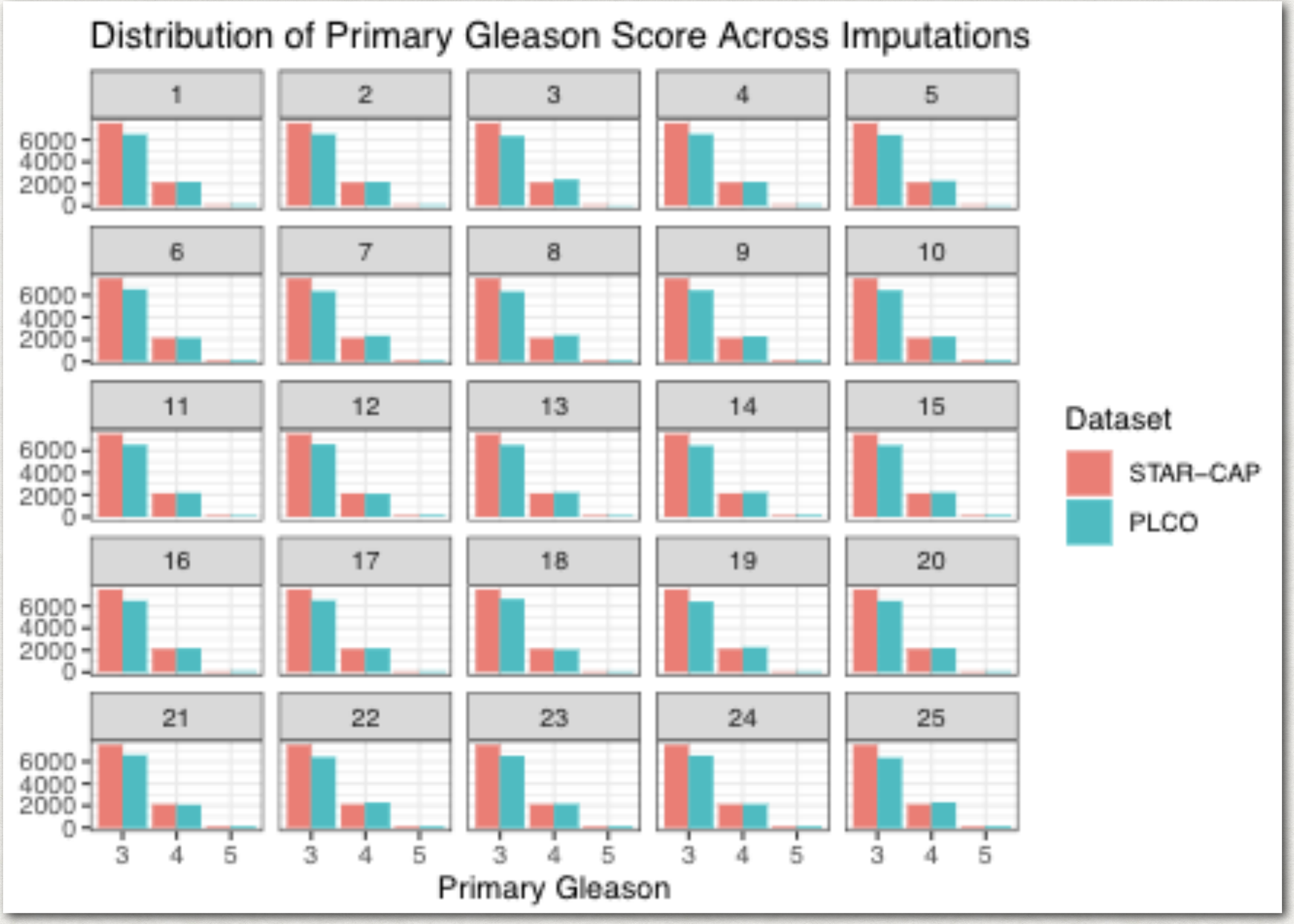
APPROACH

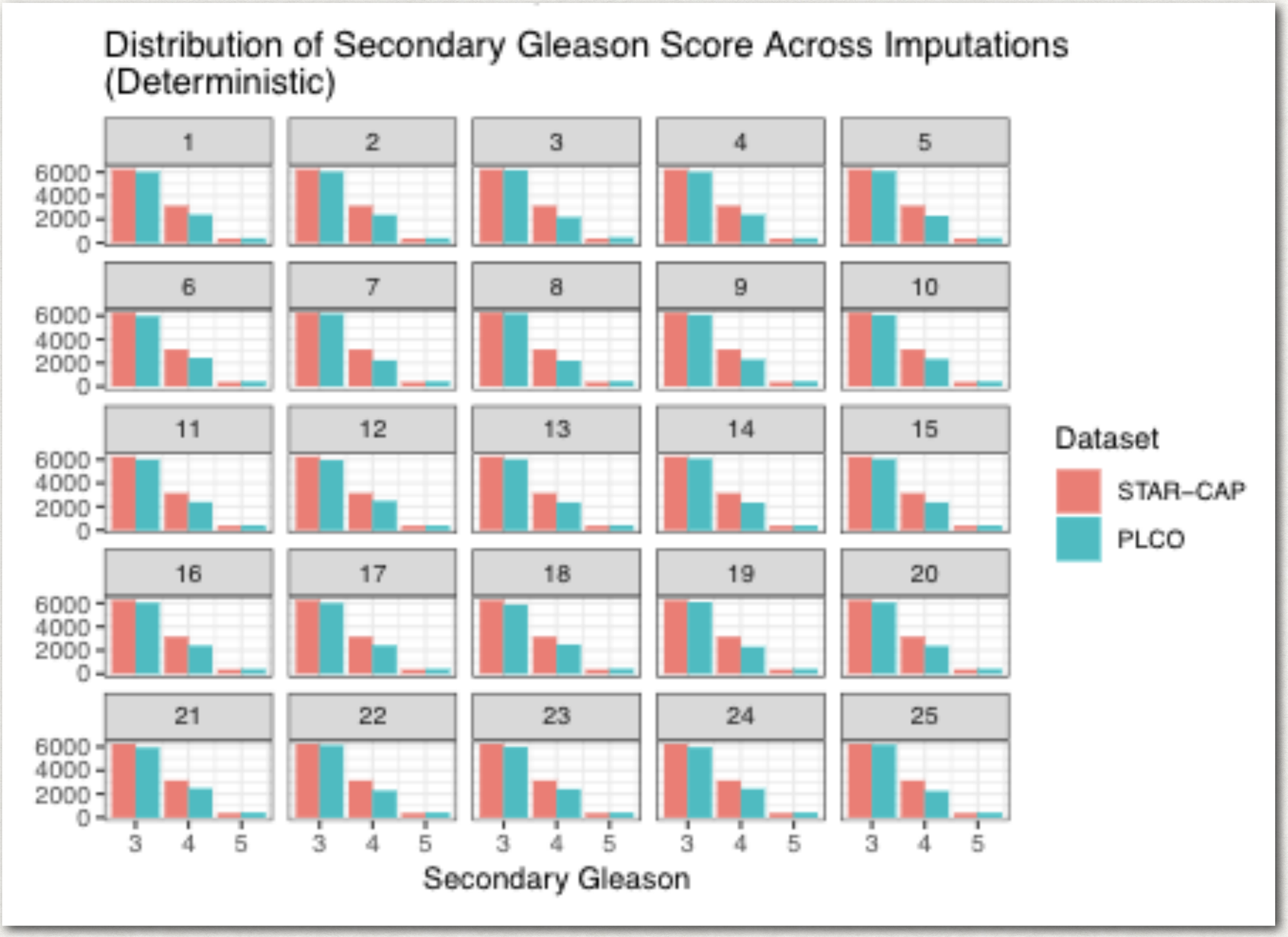
- Step 0: Re-fit STAR-CAP to be cause-specific model rather than Fine and Gray.
- Step 1: Stack STAR-CAP and PLCO data into a combined dataset.
- Step 2: Use multiple imputation via chained equations to create 100 imputations, using all predictors/outcomes shown above.

APPROACH

- Step 0: Re-fit STAR-CAP to be cause-specific model rather than Fine and Gray.
- Step 1: Stack STAR-CAP and PLCO data into a combined dataset.
- Step 2: Use multiple imputation via chained equations to create 100 imputations, using all predictors/outcomes shown above.
- Note that this approach assumes that the associations between prostate cancer predictors (pre-treatment PSA, N stage, T stage, Gleason score, percent positive cores, age) are the same in STAR-CAP as they are in PLCO, because we are using the interdependency in STAR-CAP to inform the imputation of percent positive cores and Gleason score in PLCO.


Distribution of Percent Positive Cores Across Imputations
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Percent Positive Cores
Dataset
STAR−CAP
PLCO


Distribution of Primary Gleason Score Across Imputations
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
6000
4000
2000
0
3
4
5
Primary Gleason
Dataset
STAR−CAP
PLCO

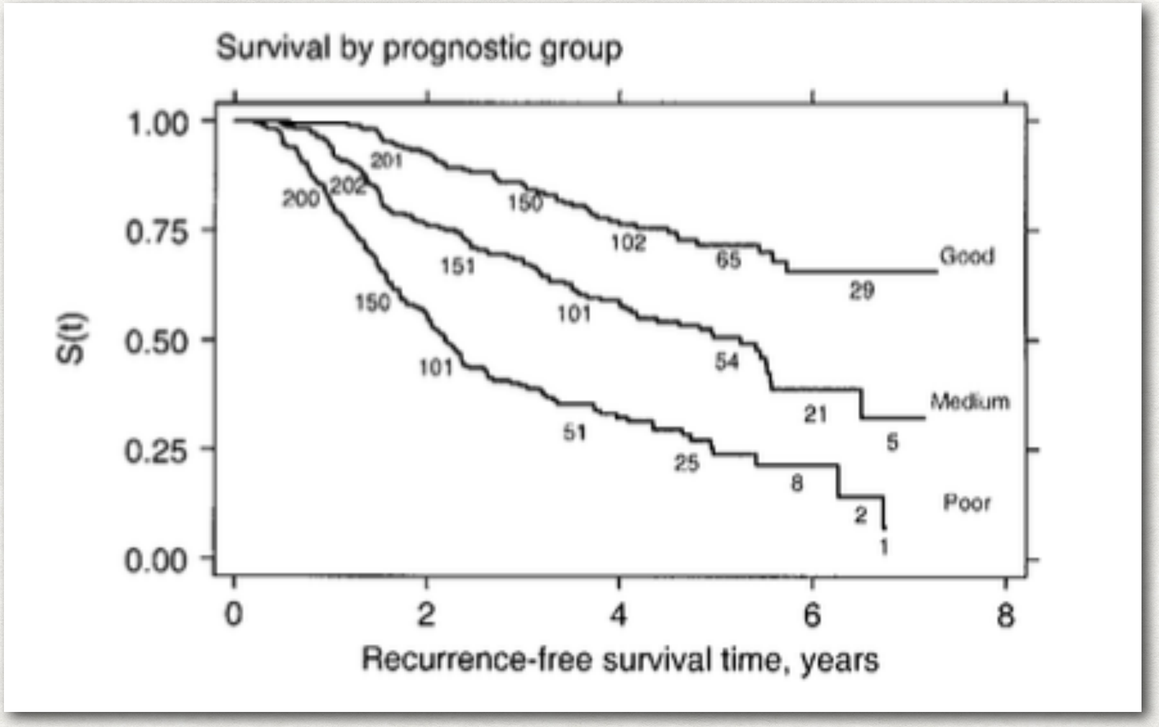

Distribution of Secondary Gleason Score Across Imputations (Deterministic)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
6000
4000
2000
0
3
4
5
Secondary Gleason
Dataset
STAR−CAP
PLCO

APPROACH

- Step 3: For each imputed dataset:
    - Drop STAR-CAP data and restrict to only PLCO data (with imputed values guided by STAR-CAP).
    - Obtain a Fine and Gray STAR-CAP prediction for absolute risk of PCSM for each patient. Use these predictions to calculate time-dependent AUC of STAR-CAP under this data augmentation/ imputation scheme.
    - Obtain the cause-specific hazard of PCSM, $\lambda_{pcsm}(t)$, from the cause-specific STAR-CAP model.
    - Obtain the cause-specific hazard of OCM, $\lambda_{ocm}(t)$, from OCCAM.
    - Integrate these two predictions using the formula given in Pfeiffer and Gail to obtain an estimate of absolute risk of PCSM.
    - Use the integrated predictions to calculate time-dependent AUC of integrated predictions.

APPROACH

- We will take a brief diversion into survival analysis to explain how we're integrating these two models.
- Let T be a patient's time to death. Then we might be interested in $S(t) = Pr(T > t)$, which we call the *survival function*.
- Key features:
  - Always decreasing
  - Always between 0 and 1
  - Starts at 1; eventually hits 0



APPROACH

- Another important function is the hazard function, $\lambda(t)$. It is interpreted as the instantaneous probability of the patient dying at time t, given that they've survived up to time t.
- Mathematically:

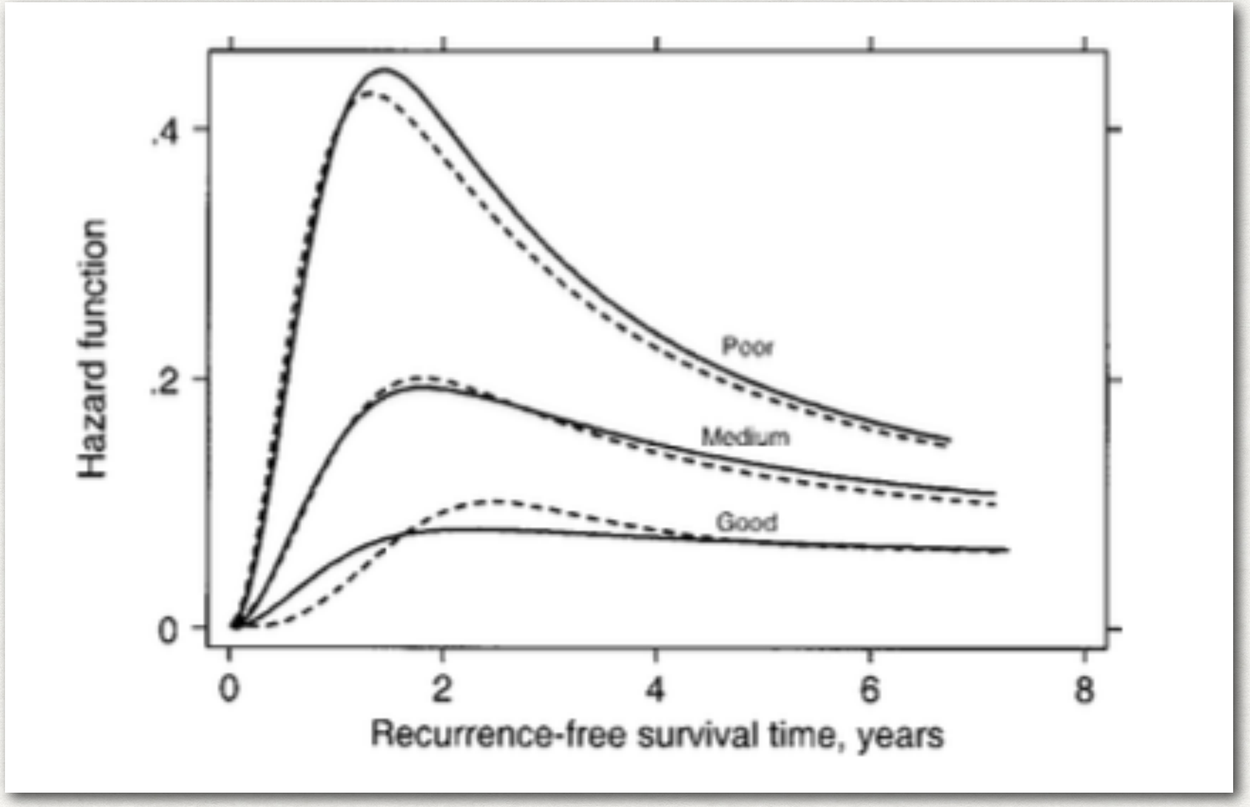
$$\lambda(t) = lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \,|\, T \geq t)}{\Delta t}$$

- To provide slightly more intuition:

$$\lambda(t) \approx Pr(T = t \,|\, T \geq t) = \frac{Pr(T = t, T \geq t)}{Pr(T \geq t)} = \frac{Pr(T = t)}{Pr(T \geq t)} \approx \frac{Pr(T = t)}{Pr(T > t)} = \frac{f(t)}{S(t)}$$

APPROACH

- Features of the hazard:
  - Always positive
  - Not constrained by 1
  - Can increase and decrease



- Also note the existence of the cumulative hazard, $\Lambda(t) = \int_0^t \lambda(u)du$, and a key identity: $S(t) = exp(-\Lambda(t))$, which only holds for continuous T.

APPROACH

- We are working with competing risks, in which a patient could die of two things: prostate cancer or other-causes.
- As a result, we can have *cause-specific* hazard functions, $\lambda_{pcsm}(t), \lambda_{ocm}(t)$: the rate at which subjects who have yet to die are dying of either PCSM or OCM.
- Note that $\lambda_{os}(t) = \lambda_{pcsm}(t) + \lambda_{ocm}(t)$.

APPROACH

- We wish to obtain the absolute risk of PCSM in the presence of competing risks:
  $F_{pcsm}(t) = Pr(T_{pcsm} \leq t)$
- How will we do this?

$$F_{pcsm}(t) = \int_0^t f_{pcsm}(u)du = \int_0^t \lambda_{pcsm}(u)S(u)du$$

$$= \int_0^t \lambda_{pcsm}(u)exp[-\Lambda(u)]du = \int_0^t \lambda_{pcsm}(u)exp[-\int_0^u \lambda(v)dv]du$$

$$= \int_0^t \lambda_{pcsm}(u)exp[-\int_0^u \lambda_{pcsm}(v) + \lambda_{ocm}(v)dv]du$$

APPROACH

- We wish to obtain the absolute risk of PCSM in the presence of competing risks:
  $F_{pcsm}(t) = Pr(T_{pcsm} \leq t)$
- How will we do this?

$$F_{pcsm}(t) = \int_0^t f_{pcsm}(u)du = \int_0^t \lambda_{pcsm}(u)S(u)du$$

$$= \int_0^t \lambda_{pcsm}(u)exp[-\Lambda(u)]du = \int_0^t \lambda_{pcsm}(u)exp[-\int_0^u \lambda(v)dv]du$$

$$= \int_0^t \lambda_{pcsm}(u)exp[-\int_0^u \lambda_{pcsm}(v) + \lambda_{ocm}(v)dv]du$$

STAR-CAP

OCCAM

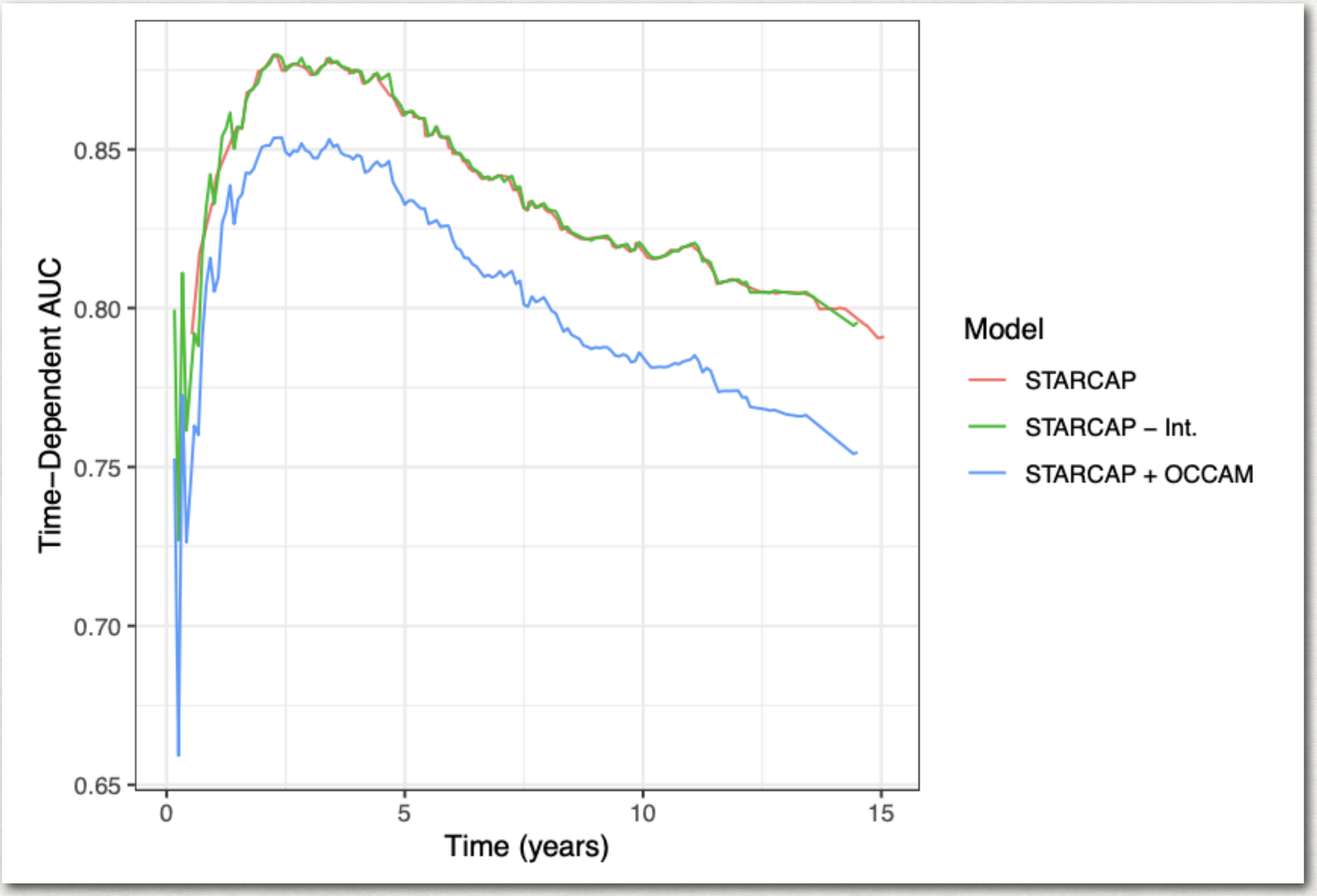
APPROACH

- Step 3: For each imputed dataset:
    - Drop STAR-CAP data and restrict to only PLCO data (with imputed values guided by STAR-CAP).
    - Obtain a Fine and Gray STAR-CAP prediction for absolute risk of PCSM for each patient. Use these predictions to calculate time-dependent AUC of STAR-CAP under this data augmentation/ imputation scheme.
    - Obtain the cause-specific hazard of PCSM, $\lambda_{pcsm}(t)$, from the cause-specific STAR-CAP model.
    - Obtain the cause-specific hazard of OCM, $\lambda_{ocm}(t)$, from OCCAM.
    - Integrate these two predictions using the formula given in Pfeiffer and Gail to obtain an estimate of absolute risk of PCSM.
    - Use the integrated predictions to calculate time-dependent AUC of integrated predictions.

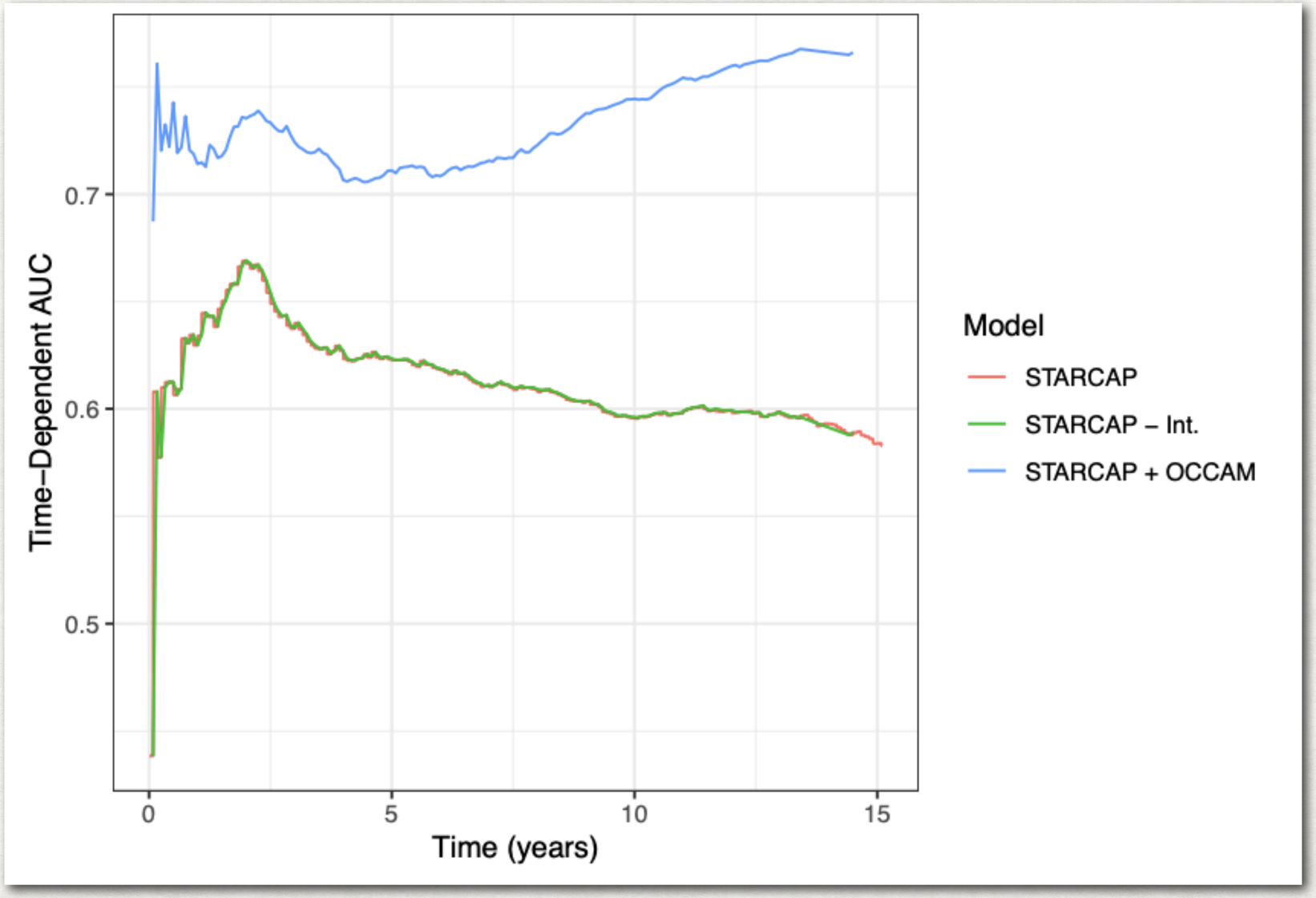
APPROACH

- Step 4: Pool the estimates:
    - Pool individual level predictions using Rubin's Rules (in theory, a variance estimate for these is available using Rubin's Rules, but I haven't finished that yet).
    - Pool AUC by taking both the mean and the median; provided they look fairly similar, use the mean. (No uncertainty available for this.)
- Step 5: Assess:
    - Compare AUC between original STAR-CAP model and the integrated prediction of PCSM.
    - Explore how STAR-CAP and integrated predictions differ.

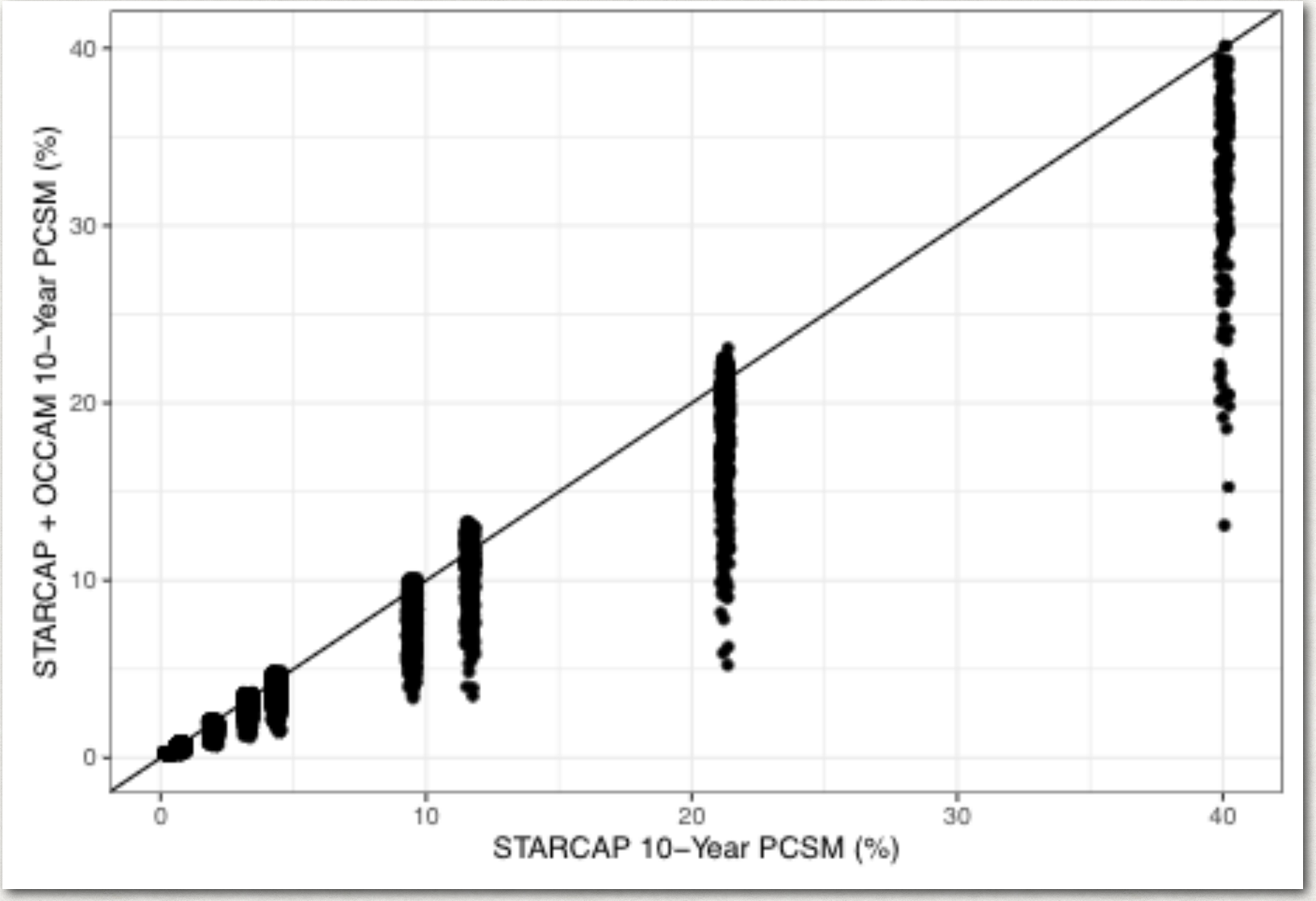
RESULTS: PCSM



RESULTS: OVERALL SURVIVAL

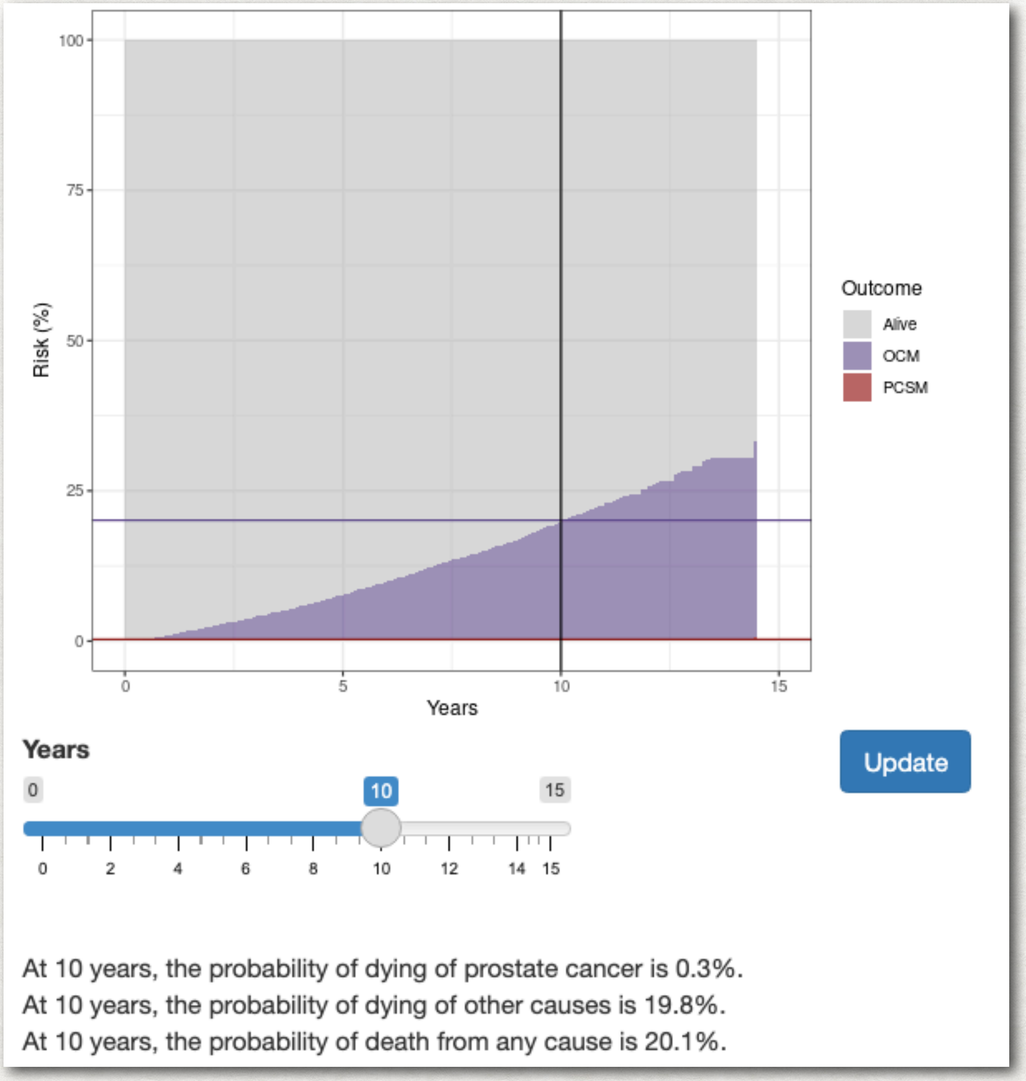


RESULTS: COMPARISON OF PREDICTIONS



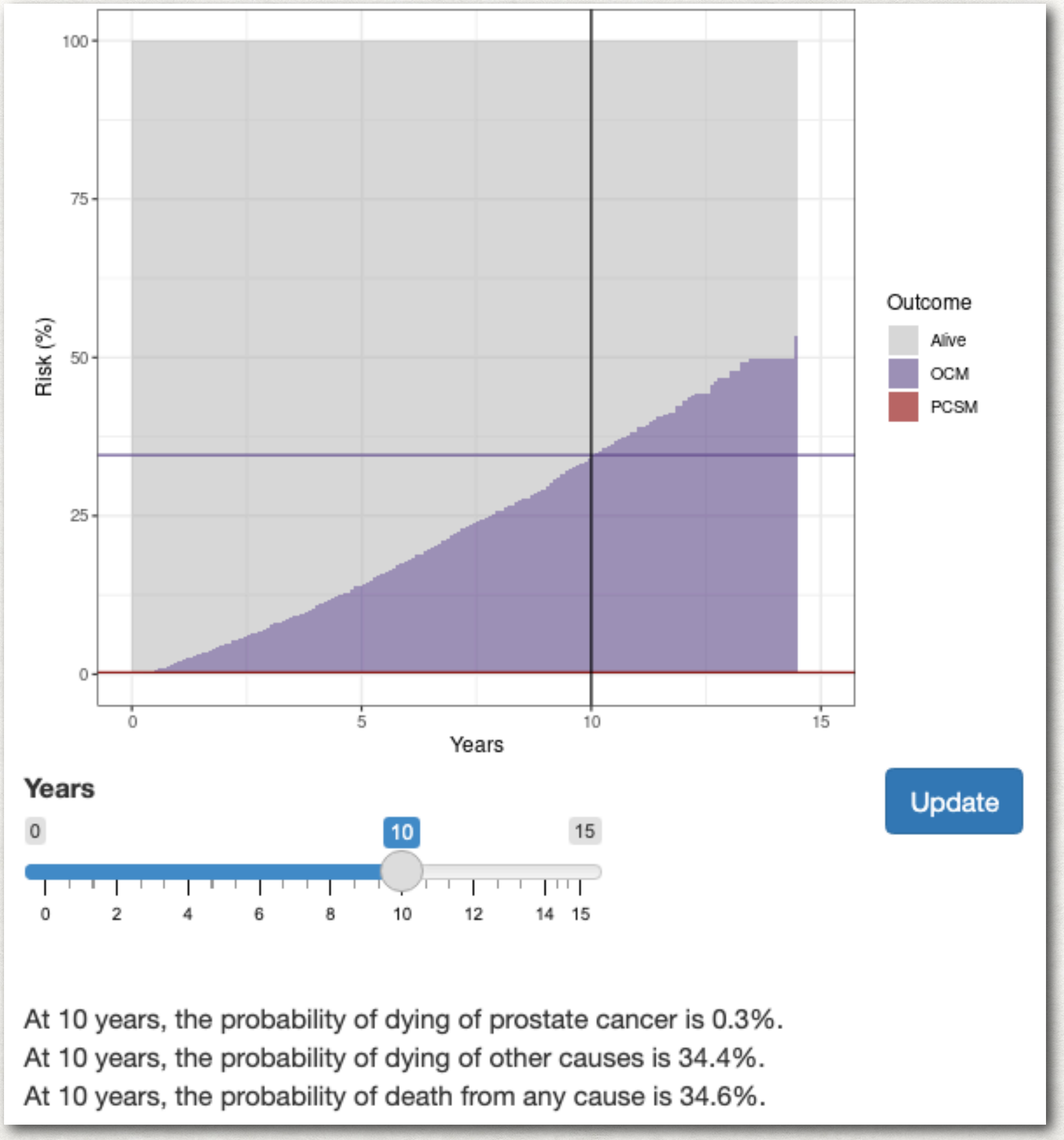
RESULTS: UNDERSTANDING IMPACT

- 65 year old man, 70 inches tall, weighs 200 lbs.
- Has diabetes
- High school graduate
- Currently married
- Never smoker
- Stage T1c, no nodal involvement
- Gleason 3+3
- Pre-treatment PSA of 4 ng/mL
- 25% positive cores



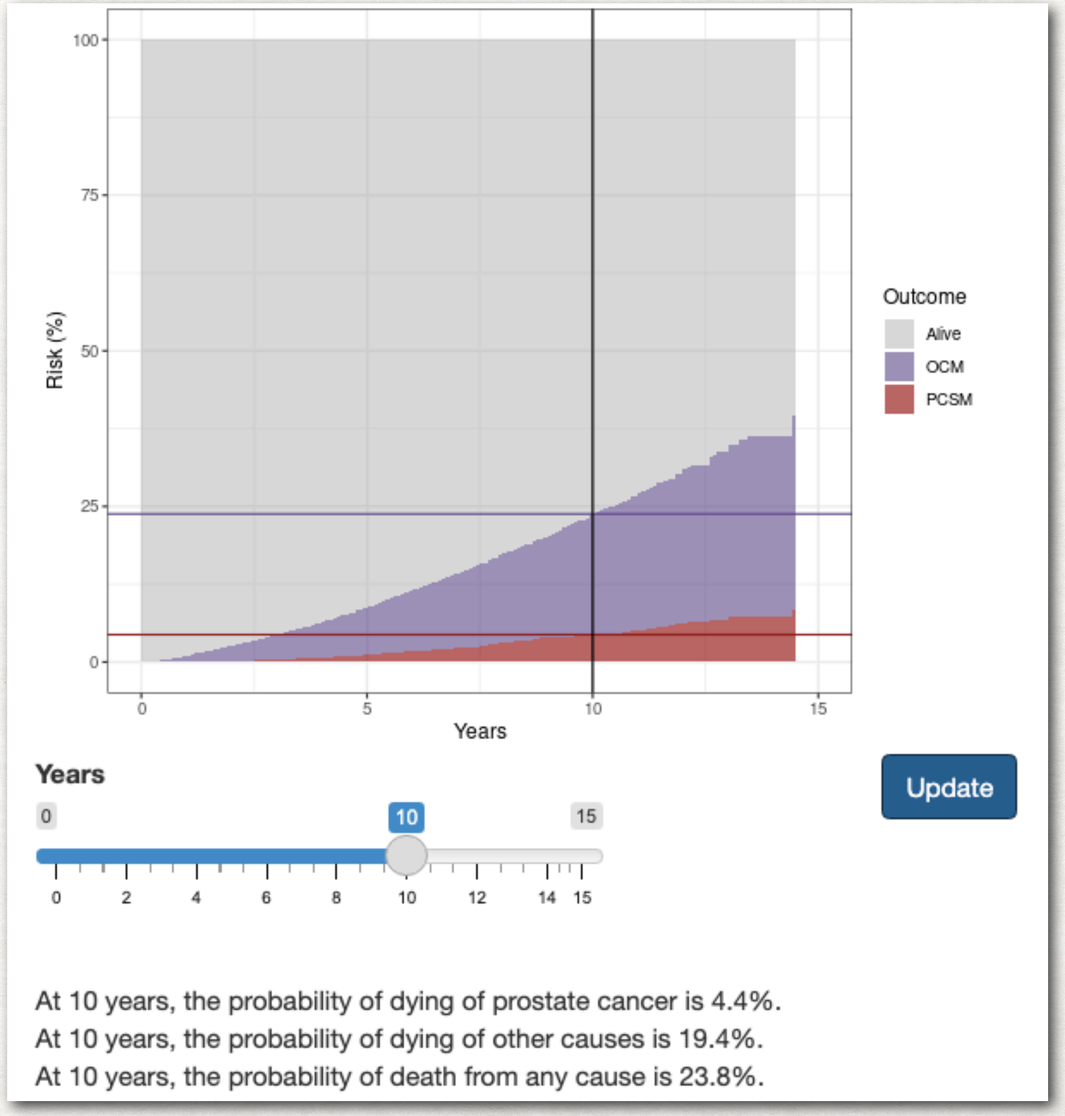
RESULTS: UNDERSTANDING IMPACT

- Exact same, but now a current smoker.



RESULTS: UNDERSTANDING IMPACT

- Exact same as first guy (but now back to being a never smoker), but with stage T3b and Gleason 4+3.



DISCUSSION

- This project provides more insight into how competing risks operate in prostate cancer and the importance of taking other-cause mortality into account.
- Large assumptions were made about:
    1. the dependence structure between prostate-cancer predictors in PLCO
    2. the transportability of a model built in NHANES to be combined with a model built in STAR-CAP
    3. the interplay of competing risks in prostate cancer.
- However, still seems to produce decent results.
- Allows clinicians to think more about what goes into other-cause mortality risk for prostate cancer patients and how that might affect treatment priorities.

REFERENCES

- PC Albertsen, JA Hanley, J Fine. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* 293(17): 2005.
- American Cancer Society. Key Statistics for Prostate Cancer. Cancer.org. https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html. Updated 2019. Accessed Sep. 1, 2020.
- RT Dess, K Suresh, MJ Zelefsky, et al. Development and validation of a clinical prognostic stage group system for nonmetastatic prostate cancer using disease-specific mortality results from the International Staging Collaboration for Cancer of the Prostate. *JAMA Oncology* 6(12): Dec. 2020.
- R Pfeiffer and M Gail. *Absolute Risk: Methods and Applications in Clinical Management and Public Health*, 2017.
- J Walz, A Gallina, et al. Clinicians are poor raters of life-expectancy before radical prostatectomy or definitive radiotherapy for localized prostate cancer. *BJU International* 100(6): 2007.