

RISK PREDICTION MODELS FOR ENDOMETRIAL CANCER

DEVELOPMENT AND VALIDATION IN THE EPIDEMIOLOGY OF ENDOMETRIAL CANCER CONSORTIUM

June 27, 2023

Joy Shi

Instructor of Epidemiology

CAUSALab and Department of Epidemiology

Harvard T.H. Chan School of Public Health

1

MOTIVATION.



4th most commonly diagnosed cancer
among women in US



Increasing incidence and mortality in the
past decade



Increasing prevalence of major endometrial
cancer risk factors (e.g., nulliparity)



Can we develop a risk prediction model to
identify high-risk?

2

Shi – Risk prediction models for endometrial cancer

||

EXISTING PREDICTION MODELS FOR ENDOMETRIAL CANCER: PLCO TRIAL & NIH-AARP STUDY.


TRAINING DATA

- 146,679 women
- 1,559 incident cases


VALIDATION DATA

- Nurses Health Study
- 37,241 women
- 532 incident cases


MODEL PREDICTORS

BMI, menopausal hormone therapy (MHT) use, parity, menopausal status, age at menopause, smoking status, oral contraceptive (OC) use, HMT × BMI interaction


RESULTS

AUC: 0.67
E/O ratio: 1.20

Pfeiffer et al., (PLOS Medicine, 2013)

3

Shi – Risk prediction models for endometrial cancer

||

EXISTING PREDICTION MODELS FOR ENDOMETRIAL CANCER: EPIC STUDY.


TRAINING DATA

- 201,811 women
- 855 incident cases


VALIDATION DATA

- Internally validated
- Five-fold cross validation


MODEL PREDICTORS

BMI, menopausal status, age at menarche and menopause, OC use, parity, age at first full-term pregnancy, duration of MHT use, smoking status, OC × BMI interaction


RESULTS

AUC: 0.77 (0.71 for age-only model)
E/O ratio: 0.99

Hüsing et al., (EJE, 2016)

4

Shi – Risk prediction models for endometrial cancer

CURRENT GAPS.



- (1) Models were trained on selective study populations
- Limited generalizability



- (2) Contributions of genetic factors have yet to be assessed

5

Shi – Risk prediction models for endometrial cancer

OBJECTIVES.



Develop a model that will predict an individual's 10-year risk for endometrial cancer based on epidemiologic questionnaire data.



Evaluate the model's performance

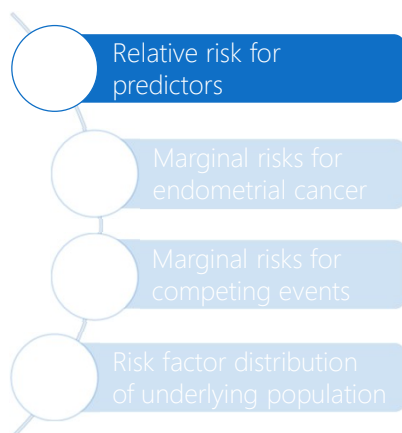


Assess the additive contribution of genetic factors to the model.

6

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL DEVELOPMENT.



DATA:
E2C2

E2C2 Consortium



PREDICTORS



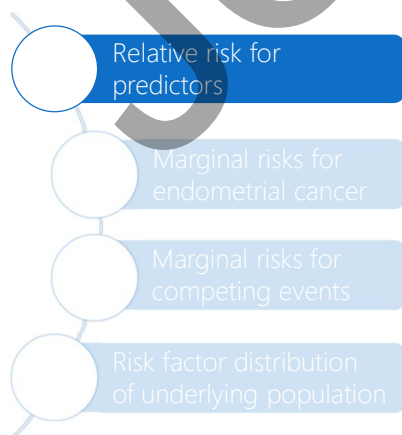
MODEL

- 19 case-control studies
- Postmenopausal, white, aged 45-85
- >6,000 cases and >9,000 controls

7

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL DEVELOPMENT.



DATA:
E2C2



PREDICTORS



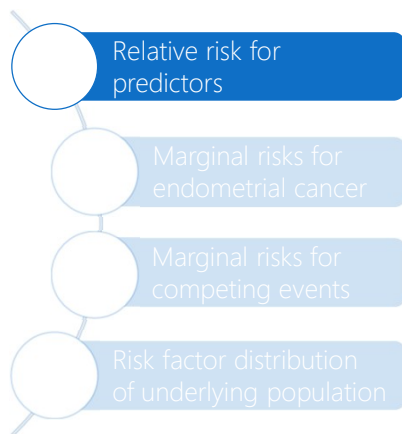
MODEL

- **Reproductive factors**
(e.g., age at menarche, age at first birth)
- **Lifestyle factors**
(e.g., smoking, body mass index)
- **Exogenous hormone-related factors**
(e.g., hormone therapy use)
- **Medical history**
(e.g., diabetes, hypertension)
- **Interaction terms**
(e.g., BMI×HT use, BMI×OC use)

8

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL DEVELOPMENT.



DATA:
E2C2



PREDICTORS



MODEL

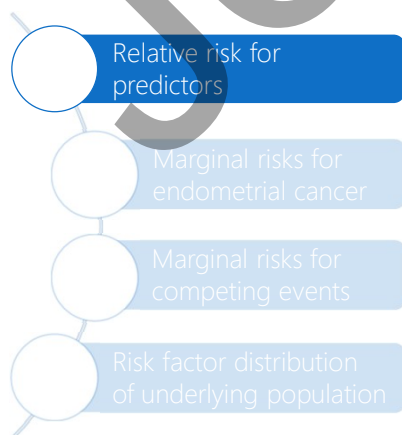
Genetic variants

- 18 genome-wide significant single nucleotide polymorphisms (SNPs)
- From O'Mara et al. (Nature Communications, 2018)

9

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL DEVELOPMENT.



DATA:
E2C2



PREDICTORS



MODEL

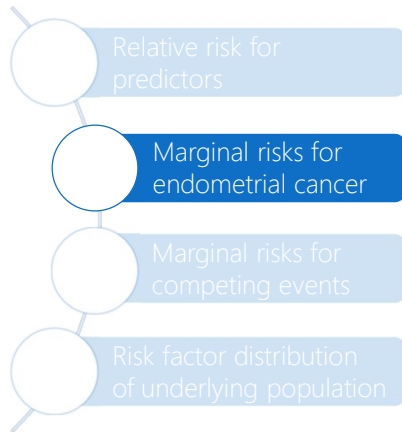
Logistic group LASSO model for variable selection and regularization

- Age and study site forced into the model
- Remaining model parameters were subject to penalization
 - Larger penalty = fewer variables retained
- Leave-one-study-out cross-validation to select tuning parameter

10

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL DEVELOPMENT.

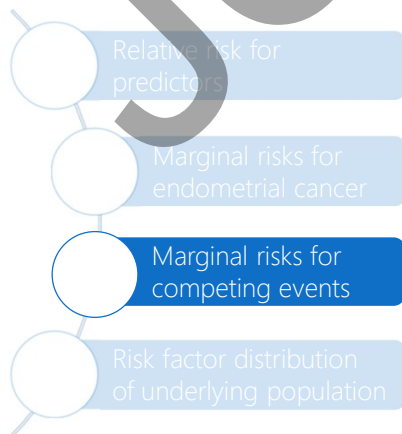


- Corrected for prevalence of hysterectomy
- Data source for endometrial cancer incidence: SEER
 - NHS: 1989-1993
 - PLCO: 1996-2000
 - NHS II: 2003-2007
- Data source for hysterectomy prevalence: BRFSS
 - NHS: 1988
 - PLCO: 1996-1998
 - NHS II: 2006 and 2008

11

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL DEVELOPMENT.



- 1) Hysterectomy: BRFSS
- 2) Other cancers: SEER
 - NHS: 1989-1993
 - PLCO: 1996-2000
 - NHS II: 2003-2007
- 3) Death: CDC WONDER
 - NHS: 1988
 - PLCO: 1997
 - NHS II: 2004

12

Shi – Risk prediction models for endometrial cancer

11



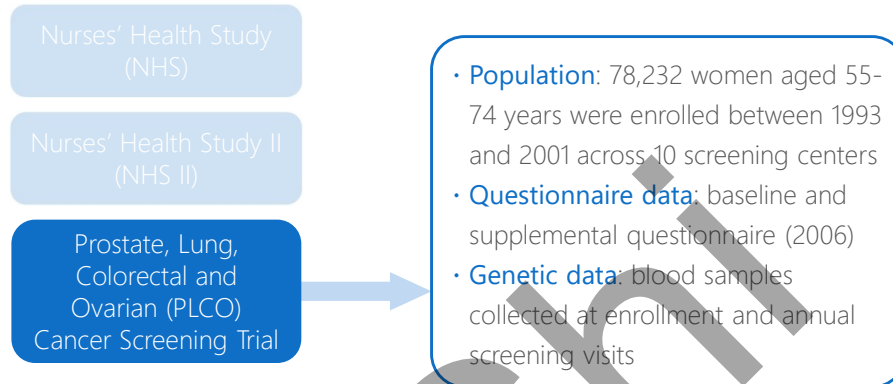
- 13

11



Shi – Risk prediction models for endometrial cancer

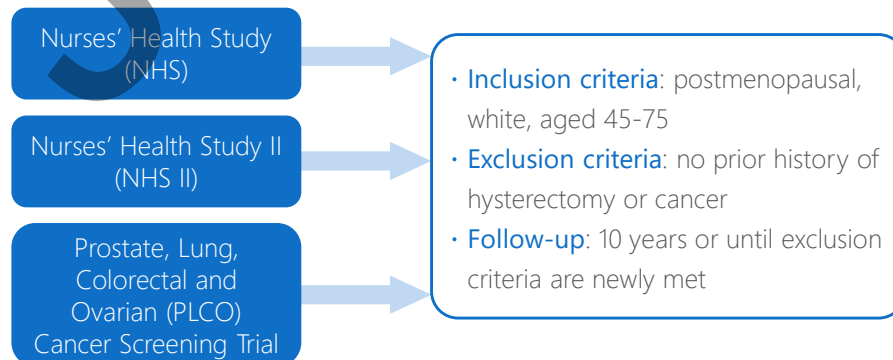
METHODS: MODEL VALIDATION DATA.



15

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL VALIDATION DATA.



16

Shi – Risk prediction models for endometrial cancer

METHODS: MODEL VALIDATION METRICS.



DISCRIMINATION

- Area under the receiver operating characteristic curve (AUC) based on 10-year risk



ABSOLUTE RISK CALIBRATION

- Expected-to-observed (E/O) ratio of 10-year absolute risk across deciles of risk
- Hosmer-Lemeshow χ^2 test



RELATIVE RISK CALIBRATION

- Goodness-of-fit test for predicted versus observed relative 10-year risk across deciles of risk

17

Shi – Risk prediction models for endometrial cancer

METHODS: ABSOLUTE RISK ESTIMATES IN MORE CURRENT POPULATION.

Combined:

- Relative risk estimates from group LASSO model
- Endometrial cancer incidence rates (SEER 2013-2017)
- Hysterectomy prevalence (BRFSS 2016 and 2018)
- Incidence rates for competing risks (2017 CDC WONDER data for mortality; 2013-2017 SEER data for other cancers)
- Risk factor distributions (NHANES 2017-2018)

18

Shi – Risk prediction models for endometrial cancer

RESULTS: MODEL PREDICTORS.

Characteristics	RR	Characteristics	RR
Demographic factors		Reproductive and hormonal factors	
Education, %		Parity, %	
High school or below	(ref)	0	(ref)
Some college or equivalent	0.97	1	1.10
College or above	0.96	2	0.91
Lifestyle factors		3	0.77
Smoking status, %		≥4	0.60
Never smoker	(ref)	Age at first birth, %	
Former smoker	0.80	<20	(ref)
Current smoker	0.64	20 to <25	0.96
Body mass index (kg/m ²), %		25 to <30	0.85
<18.5	0.74	30 to <35	0.83
18.5 to <25	(ref)	≥35	0.84
25 to <30	1.41	Never given birth	1.28
30 to <35	2.49	(cont.)	
≥35	5.57		

19

Shi – Risk prediction models for endometrial cancer

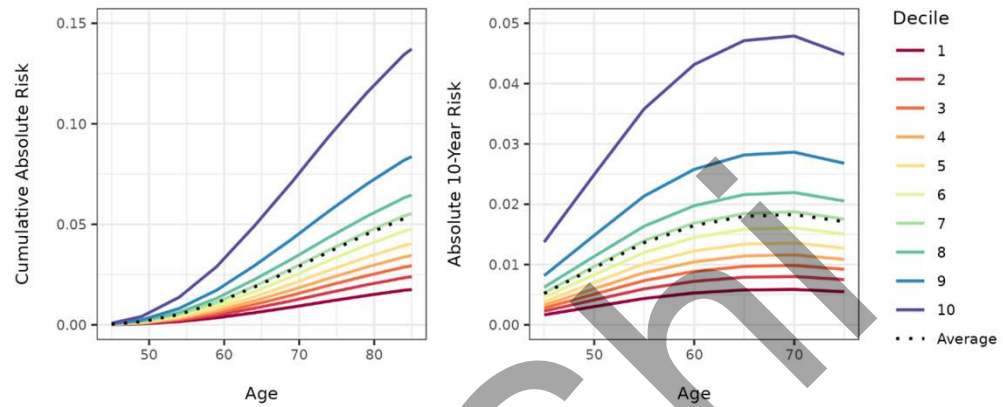
RESULTS: MODEL PREDICTORS.

Characteristics	RR	Characteristics	RR
Reproductive and hormonal factors		Reproductive and hormonal factors	
Age at menarche, %		Any E+P HT use, %	0.82
≤9	(ref)	Duration of E+P HT use (years), %	
10-11	1.04	0	(ref)
12-13	1.04	>0 to 5	1.00
14-15	0.92	>5 to 10	1.00
≥16	0.89	>10	1.00
Any HT use, %	1.61	Any OC use, %	0.79
Any E-only HT use, %	1.06	Duration of OC use (years), %	
Duration of E-only HT use (years), %		0	(ref)
0	(ref)	>0 to 5	1.05
>0 to 5	0.84	>5 to 10	0.94
>5 to 10	1.42	>10	0.69
>10	2.55	Clinical factors	
(cont.)		History of diabetes, n (%)	1.39
		History of hypertension, n (%)	1.22

20

Shi – Risk prediction models for endometrial cancer

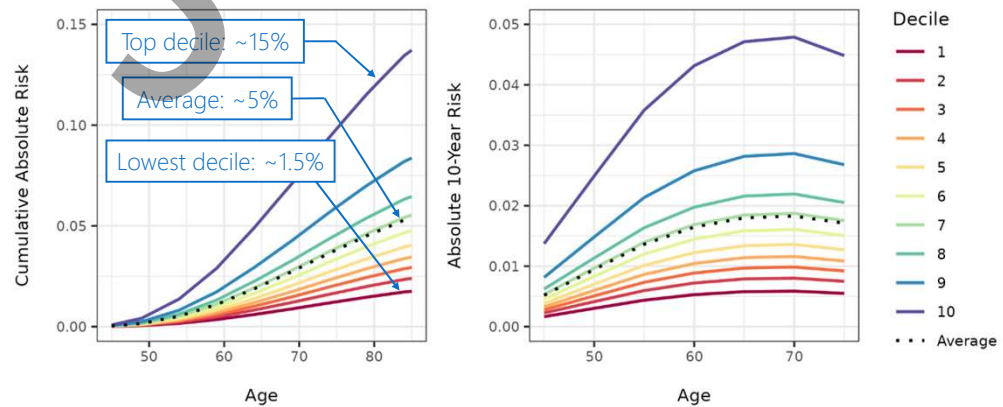
RESULTS: ESTIMATED CUMULATIVE AND 10-YEAR RISKS.



21

Shi – Risk prediction models for endometrial cancer

RESULTS: ESTIMATED CUMULATIVE AND 10-YEAR RISKS.



22

Shi – Risk prediction models for endometrial cancer

RESULTS: MODEL DISCRIMINATION (EPIDEMIOLOGIC MODEL).

VALIDATION COHORT	NUMBER OF PARTICIPANTS	NUMBER OF EVENTS	AUC (95% CI)
NHS	68,150	700	0.647 (0.626, 0.667)
NHS II	56,076	304	0.693 (0.664, 0.723)
PLCO	39,996	511	0.640 (0.615, 0.665)

23

Shi – Risk prediction models for endometrial cancer

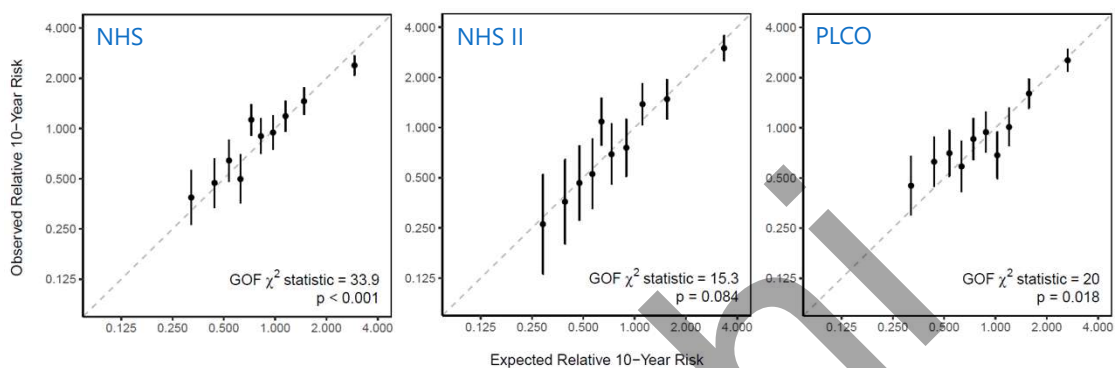
RESULTS: MODEL DISCRIMINATION (EPIDEMIOLOGIC + GENETIC MODEL).

VALIDATION COHORT	NUMBER OF PARTICIPANTS	NUMBER OF EVENTS	AUC (95% CI)	
			EPIDEMIOLOGIC MODEL	EPIDEMIOLOGIC + GENETIC MODEL
NHS (Genetic cohort)	11,365	166	0.613 (0.570, 0.656)	0.613 (0.570, 0.656)
PLCO (Genetic cohort)	30,102	401	0.635 (0.606, 0.664)	0.665 (0.636, 0.693)

24

Shi – Risk prediction models for endometrial cancer

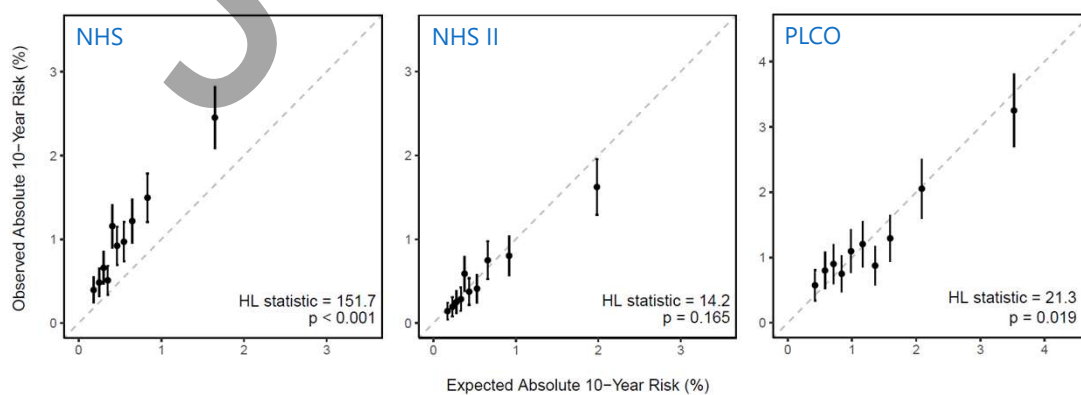
RESULTS: MODEL CALIBRATION OF RELATIVE RISKS.



25

Shi – Risk prediction models for endometrial cancer

RESULTS: MODEL CALIBRATION OF ABSOLUTE RISKS.

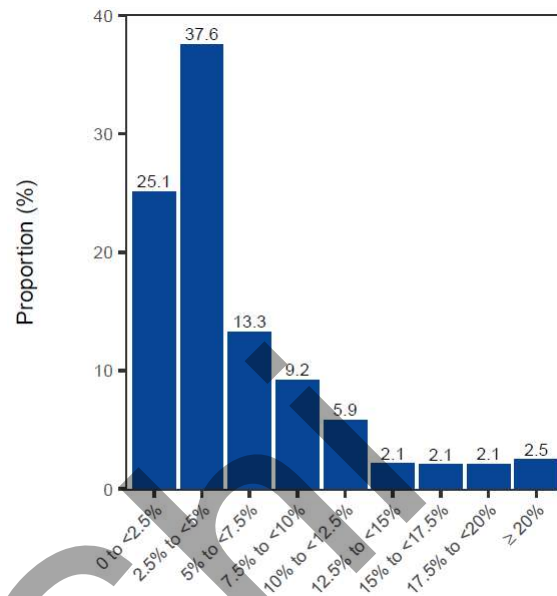


26

Shi – Risk prediction models for endometrial cancer

||

RESULTS: ESTIMATED RISKS IN A MORE CURRENT POPULATION.

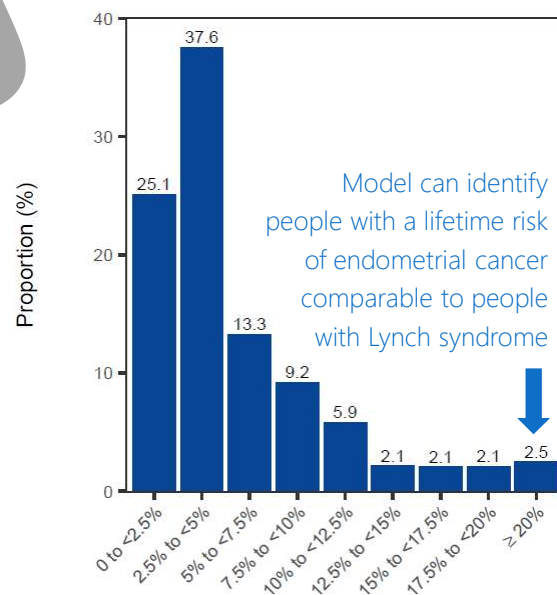


27

Shi – Risk prediction models for endometrial cancer

||

RESULTS: ESTIMATED RISKS IN A MORE CURRENT POPULATION.



28

Shi – Risk prediction models for endometrial cancer

||

STRENGTHS.



SUMMARY

- Prediction model demonstrated moderate discrimination
- Well calibrated in NHS II and PLCO
- Based on clinical factors alone
- Minimal improvements with addition of published genetic factors



IMPLICATIONS

Potential tool for identifying people at high risk of endometrial cancer

- Screening high risk individuals
- Risk-based prevention strategies
- Enrollment in prevention or screening trials

29

Shi – Risk prediction models for endometrial cancer

||

HOW DOES OUR MODEL COMPARE AGAINST PREVIOUS MODELS?



VARIABLE SELECTION

- Group LASSO vs. stepwise approaches
- Our models included more risk factors (e.g., education, E+P HT use, diabetes, hypertension)



DISCRIMINATION

- Smaller AUCs (0.64 to 0.69) in our model than EPIC (0.77)
- EPIC model largely driven by predictive contribution of age (AUC=0.71 in age-only model)



GENERALIZABILITY

- We used external data to estimate risk factor distributions and baseline incidence
- Previous models developed on selective populations

30

Shi – Risk prediction models for endometrial cancer

LIMITATIONS & NEXT STEPS.


AVAILABILITY OF
RISK FACTORS

Could not include family history of endometrial cancer because these data were not collected in NHANES


RECALL BIAS

- Models based on case-control data
- Previous analyses of E2C2 data have reported similar RR estimates between cohort and case-control studies


AVAILABILITY OF
GENETIC DATA IN NHS

- Genetic data pooled from GWAS of different disease outcomes
- Matching on factors may explain lower AUC


NEXT STEPS

Expanding to multi-racial/multi-ethnic populations

ACKNOWLEDGEMENTS.



- Immaculata De Vivo
- Margaret Du

PhD Committee Members

- Peter Kraft
- Bernard Rosner
- Miguel Hernan

Funding

- UCB Fellowship

E2C2 collaborators and study participants

||

You can read more about the study here:

<https://news.harvard.edu/gazette/story/2023/02/new-model-identifies-those-at-high-risk-for-endometrial-cancer/>

<https://pubmed.ncbi.nlm.nih.gov/36688725/>

QUESTIONS?

You can connect with me here:



joyshi@hsph.harvard.edu



@joy_shi1

33

Shi – Risk prediction models for endometrial cancer

Joy Shi