Epidemiology of Cancer Survival: Methodological Considerations

Patrick T. Bradshaw, Ph.D.
Assistant Professor
School of Public Health, Division of Epidemiology
University of California, Berkeley

PB HLTH 258/EPI 252 Cancer Epidemiology
13 April 2020

References

- Chubak J, Boudreau DM, Wirtz HS, McKnight B, Weiss NS. Threats to validity of nonrandomized studies of post diagnosis exposures on cancer recurrence and survival. *Journal of the National Cancer Institute*. 2013. 105(19): 1456-1462.
- de Gonzalez AB, Morton LM. Converting epidemiologic studies of cancer epidemiology to survivorship studies: approaches and challenges. *Cancer Epidemiology Biomarkers and Prevention*. 2012. 21(7): 875-880.

Outline

- **Epidemiology of Cancer Survival**
- **Study Design Considerations**
- **Threats to Validity**
- **Examples**

Epidemiology of Cancer Survival

Cancer Survivors

- Cancer survivor: anyone with history of cancer diagnosis.
- Currently **>15.5 million cancer survivors** in the United States.
  - Includes children & adults.
  - By 2030: estimated to increase to >22 million.

Trends in Five-year Relative Survival Rates (%), 1975-2015

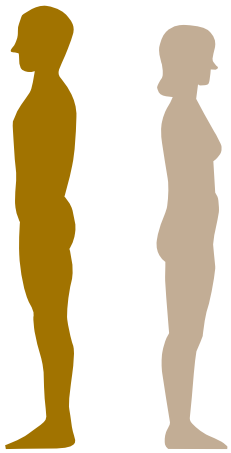
Site	1975-1977	1987-1989	2009-2015
All sites	49	55	69
Breast (female)	75	84	91
Colorectum	50	60	66
Leukemia	34	43	66
Lung & bronchus	12	13	21
Melanoma of the skin	82	88	94
Non-Hodgkin lymphoma	47	51	75
Ovary	36	38	48
Pancreas	3	4	10
Prostate	68	83	99
Urinary bladder	72	79	78

5-year relative survival rates based on patients diagnosed in the 9 oldest SEER registries from 1975-1977, 1987-1989, and 2009-2015, all followed through 2016.
Source: Surveillance, Epidemiology, and End Results (SEER) Program, National Cancer Institute, 2019.

Figure 1. Estimated Number of US Cancer Survivors by Site

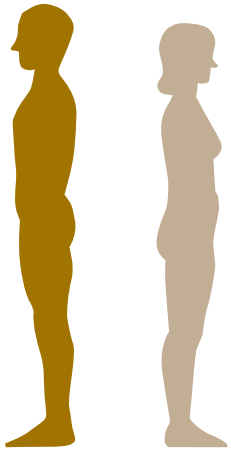
As of January 1, 2019

Male		Female	
Prostate	3,650,030	Breast	3,861,520
Colon & rectum	776,120	Uterine corpus	807,860
Melanoma of the skin	684,470	Colon & rectum	768,650
Urinary bladder	624,490	Thyroid	705,050
Non-Hodgkin lymphoma	400,070	Melanoma of the skin	672,140
Kidney & renal pelvis	342,060	Non-Hodgkin lymphoma	357,650
Testis	287,780	Lung & bronchus	313,140
Lung & bronchus	258,200	Uterine cervix	283,120
Leukemia	256,790	Ovary	249,230
Oral cavity & pharynx	249,330	Kidney & renal pelvis	227,510
All sites	**8,138,790**	**All sites**	**8,781,580**



As of January 1, 2030

Male		Female	
Prostate	5,017,810	Breast	4,957,960
Colon & rectum	994,210	Uterine corpus	1,023,290
Melanoma of the skin	936,980	Thyroid	989,340
Urinary bladder	832,910	Colon & rectum	965,590
Non-Hodgkin lymphoma	535,870	Melanoma of the skin	888,740
Kidney & renal pelvis	476,910	Non-Hodgkin lymphoma	480,690
Testis	361,690	Lung & bronchus	398,930
Leukemia	352,900	Kidney & renal pelvis	316,620
Lung & bronchus	325,680	Ovary	297,580
Oral cavity & pharynx	315,750	Uterine cervix	288,710
All sites	**10,995,610**	**All sites**	**11,174,200**

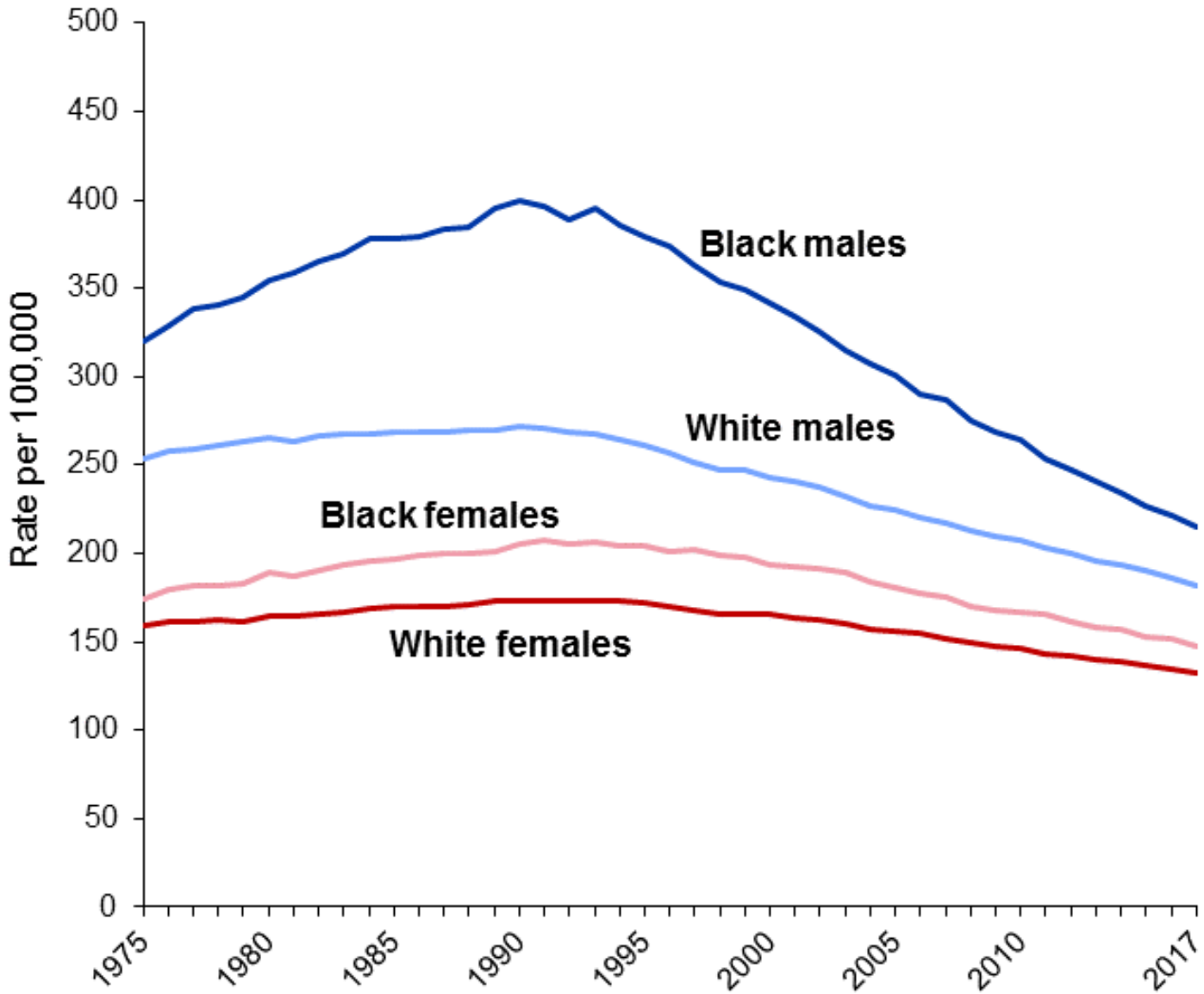


Estimates for specific cancers account for the fact that some individuals have a history of multiple different cancer types. See Sources of Statistics, page 36, for more information.

Source: Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute.

*American Cancer Society. *Cancer Treatment & Survivorship Facts & Figures 2019-2021*. Atlanta: American Cancer Society; 2019.

Trends in Cancer Death Rates* by Sex and Race, US, 1975-2017



*Age-adjusted to the 2000 US standard population.
Source: National Center for Health Statistics, Centers for Disease Control and Prevention, 2019.

Epidemiologic Research in Cancer Survival

- Much of cancer survivorship research involves **clinical trials or other interventions**.
- **Epidemiological/observational studies** of cancer survivors have number of unique challenges.
  - Study design considerations.
  - Threats to validity.
  - Analytical approaches.

Study Design Issues

What is the question?

- **Decide the objective of the study.**
  - Describe burden.
    - *e.g.* disparities; incident comorbidities.
  - Elucidate **etiologic relationship**.
  - **Inform intervention**.
    - Population-level.
    - Clinical/individual.
  - Mortality among cancer survivors vs. cancer-related mortality among healthy population?

Exposures

Pre-diagnosis vs. post-diagnosis?

- Many studies use **pre-diagnosis** exposures?
  - May be logistically attractive.
  - Several limitations/methodological concerns.
- **Post-diagnosis exposures** may be more relevant for intervention.
  - Single or repeated measurements?
  - Are levels of exposure meaningful?
  - Changes in exposure over time?

Outcome

- **What's the relevant endpoint?**
  - **Mortality.**
    - All-cause?
    - Cancer-specific?
    - Other causes of death? (Increasingly important!)
  - Disease **progression/recurrence.**
  - **Co-morbidities.**
  - **Others:** quality of life; DALY
- Consider **feedback between exposure-outcome decision.**
  - Certain outcomes may require longitudinal exposure assessment.

Study Design

- **De novo** cohort studies.
  - Identify/enroll newly diagnosed subjects.
- **Converting existing study** can be efficient:
  - Converting case-control studies to cohort.
    - Incident cases already enumerated.
    - Return to the field?
  - Nested designs:
    - Case-cohort design.
    - Nested case-control.
    - Can be very efficient options for many resources.

Data sources for existing studies

- Administrative databases.
  - Electronic health records.
  - Claims data.
- Existing cohort studies.
  - May be efficient, especially for biological samples.
- Existing case-control studies.
  - Potential for interesting comparisons of some risk factors.
  - May need to quickly return to fieldwork.

Sample size considerations

- Ability to **accrue sufficient events**?
  - e.g. 90% of women with BrCa will survive past 5 years.
    - Recruit 3000 subjects to accrue 300 deaths in 5 years (no dropout).
- **Solutions**:
  - **Pool resources** across multiple studies.
    - Logistical issues/harmonization.
  - **Recruit high-risk subjects**.
    - Generalizability?
  - **Intermediate endpoints**.
    - Biomarkers related to outcome of interest.

Threats to Validity

Threats to valid inference

Frequent methodological issues in cancer survival research:*

1. **Reverse causation.**
2. **Confounding.**
3. **Selection bias.**
4. **Information bias.**

These are sources of **systematic error** in epidemiologic/ observational research.

* Chubak J, et al. Threats to validity of nonrandomized studies of post diagnosis exposures on cancer recurrence and survival. Journal of the National Cancer Institute. 2013. 105(19): 1456-1462.

Reverse causation

Common concern: **disease progression affects exposure**.

- Post-dx weight change and cancer mortality.
- Progression influences medication adherence.
- How do you disentangle?

Reverse causation

- **Analytical attempts** to remediate:
  - **Restriction** (only include those known to be recurrence-free at exposure).
    - Also common to exclude “early deaths.”
  - Exposure lag.
    - Effectively restriction.
- **Caveats** to these approaches:
  - Generalizability?
  - **Selection bias** (only include those surviving >1yr).*
- Probably better to **stratify on stage/subtype.**

Hernan MA. The hazards of hazard ratios. Epidemiology. 2010. 21(1):13-15;
Hernan MA, et al. A structural approach to selection bias. Epidemiology. 2004. 15(5): 615-625.

Confounding

- **Pre-diagnosis confounders** significant concern in de novo cohort studies.
  - Data collection may not include these.
  - Less of an issue in “converted case-control” studies.
- Exposure assessment issues in case-control studies can persist.
  - **Misclassification**.
- **Treatment/stage** tend to be very contentious.
  - Stratification usually interesting question.
  - Proceed cautiously for pre-diagnosis exposures.
    - Treatment could be on causal pathway.

Selection bias

- **Selection bias**: selection/retention into study dependent on variables related to exposure and/or outcome.
  - **Self-selection**:
    - Study participants often healthiest ones.
    - Best to try to achieve most representative sample; maximum response/retention.

Delayed entry into cohort

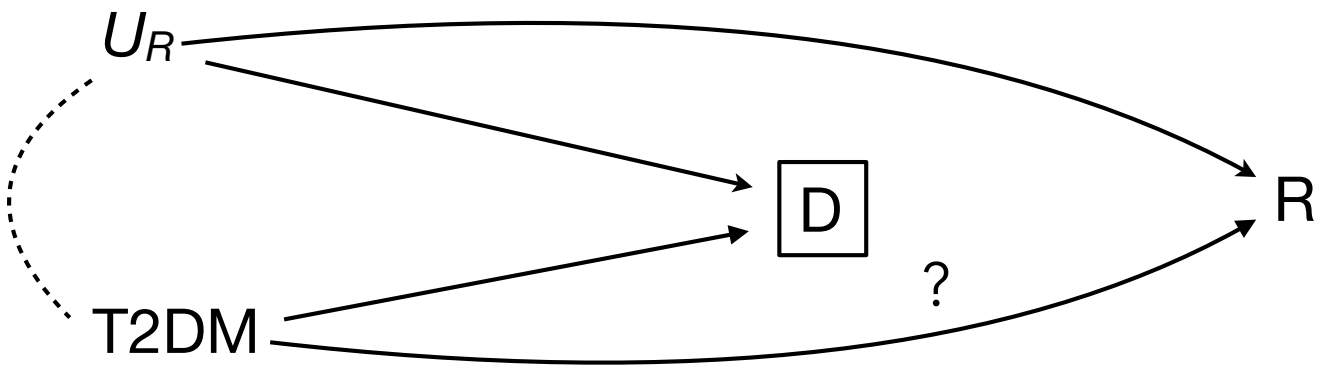
- **Event times (e.g. deaths) before a certain point not observed.**
  - Many survivorship studies only recruit individuals after a certain point (post-tx; several years after diagnosis).
  - Entry points could be different for each subject.
    - Need a consistent definition of time=0.
  - Follow-up time doesn't correspond to same scale for everyone.
- Necessitates use of delayed-entry time-to-event methods.
  - Inferences become limited: effect of exposure on death near diagnosis would be unclear.

Loss to follow-up

- **Loss to follow-up:** subjects fail to come for study visits.
  - Best case (non-differential): reduces your sample size.
  - Worst case (differential): biases results.
  - Sickest patients may not respond to follow-up questionnaire.
- Participation is linked to outcome (survival).
  - Form of **informative censoring**.
- Other similar mechanisms:
  - Competing risks: death from other causes.

Loss to follow-up

Goal: association between diabetes (*T2DM*) and diagnosed recurrence (*R*):*

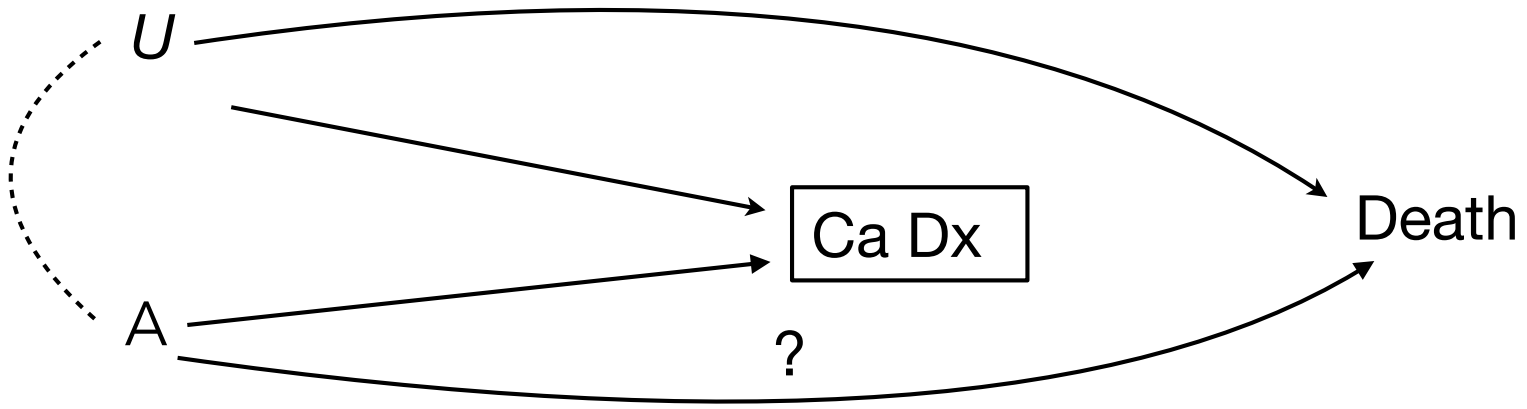


- Undiagnosed recurrence (U_R) increases likelihood of dropout (*D*).
- Diabetes (*T2DM*) also independently related to dropout (*D*).
  - Including only those who did not drop out conditions on collider:
    - Creates artifactual relationship between *T2DM* and *R* (dashed line).

* Chubak J, et al. Threats to validity of nonrandomized studies of post diagnosis exposures on cancer recurrence and survival. Journal of the National Cancer Institute. 2013. 105(19): 1456-1462.

Loss to follow-up

- May apply to many pre-diagnosis exposures (A) and mortality:



- Exposure independently related to cancer diagnosis.
- Uncontrolled factor (U) also related to likelihood of diagnosis and death.
- Would not necessarily identify U as confounder here.

Selection bias

Assessing selection bias:

- Compare characteristics of responders to non-responders.
  - Requires you know something about non-participants/ dropouts.
  - Assumes any differences will be reflected in known characteristics.
  - If participation/retention related to unknown characteristic, bias potentially unresolvable.

Selection bias

Analytical solutions:

- Properly account for follow-up time in survival analysis.
- **Inverse probability of selection weighting**.
  - Estimate probability of selection/retention.
  - Give greater weight to observations that had a low chance of participating.
- **Selection models**.
  - Simultaneously models outcome and probability of selection/retention.
- Analytical methods useful, but **require strong assumptions**.

Information bias

- **Differential outcome ascertainment:**
  - Outcomes may be monitored more aggressively among exposed.
  - Not as much concern for mortality endpoints (NDI).
- **Infrequent follow-up** for exposure/covariate assessment:
  - Exposures change over time.
  - Large concern for certain designs (converted case-control).
- **Measurement error:**
  - Error in self report.
  - Health records were designed for billing/administrative decision making, not epidemiology.

Information bias

Immortal person-time:*

- Follow-up time included where subject could not have possibly had the event.
- Usually occurs between selection into study (e.g. cancer diagnosis) and exposure assessment.
  - *e.g.* only those who lived long enough after diagnosis to give blood would be included in study.
    - Time between diagnosis and blood draw is immortal person-time.
- **Solution**: delayed entry survival analysis.
  - Could still have bias if time to entry is informative.

* Lash TL and Cole SR. Journal of Clinical Oncology. 2009. 27(3): e55-e56.

Example: Weight Change and Breast Cancer Survival

Study rationale

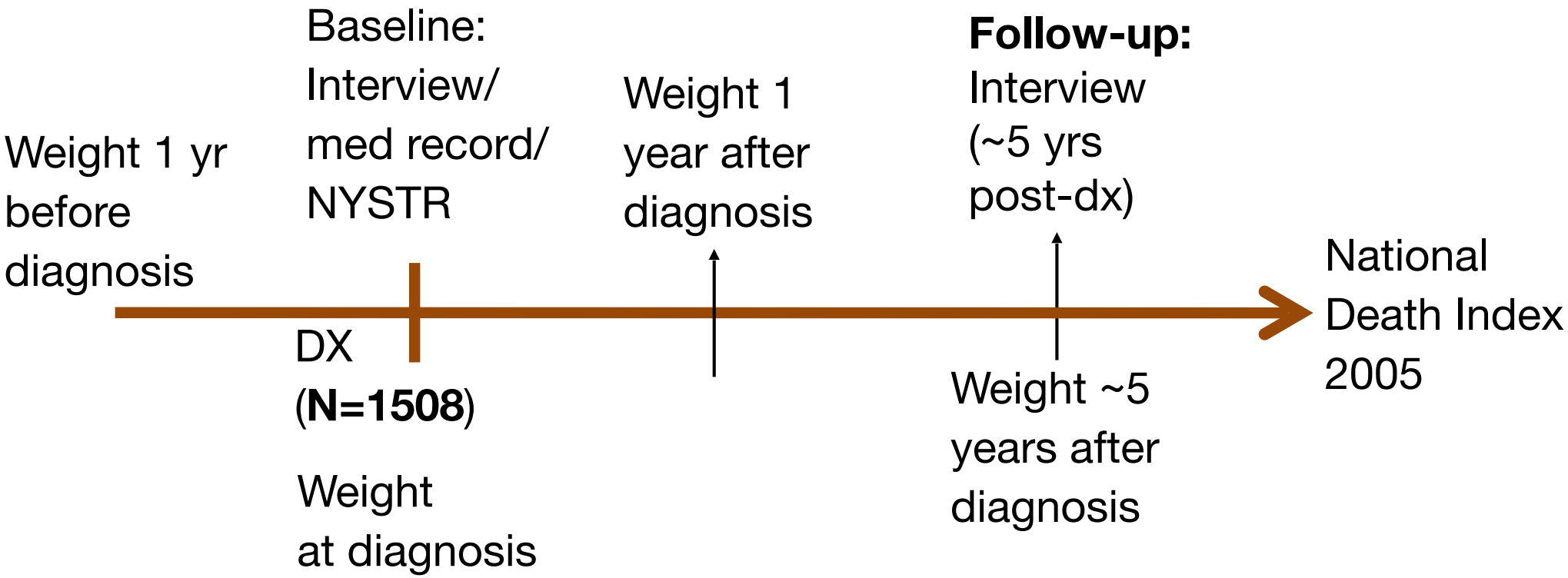
- Few studies at the time examined **post-diagnosis weight changes** and breast cancer survival.
- Previous studies **methodologically limited**:
  - Many used pre-diagnosis weight change.
  - Selection bias/left-truncation: most studies recruited survivors 2+ years after diagnosis.
  - Information bias due to single exposure: previous studies did not use longitudinal weight change over time.
- Reverse causation.

Study rationale

- Our analysis: data from **existing case-control study** that was **converted to follow-up study**.
  - Incident cases.
  - Repeated assessment of follow-up (at 5 years).
  - Pre-diagnosis exposures.
  - Death assessed from National Death Index
    - No differential loss to follow-up regarding outcome.
  - Estimated **hazard ratios** (similar to rate ratio) for overall and breast cancer-specific mortality.
  - Considered **follow-up from diagnosis** for everyone.

Study methods

- **Data source**: Long Island Breast Cancer Study Project (PI: Marilie Gammon)
  - 1508 women diagnosed with in situ/invasive breast cancer on Long Island, NY, 1996-1997

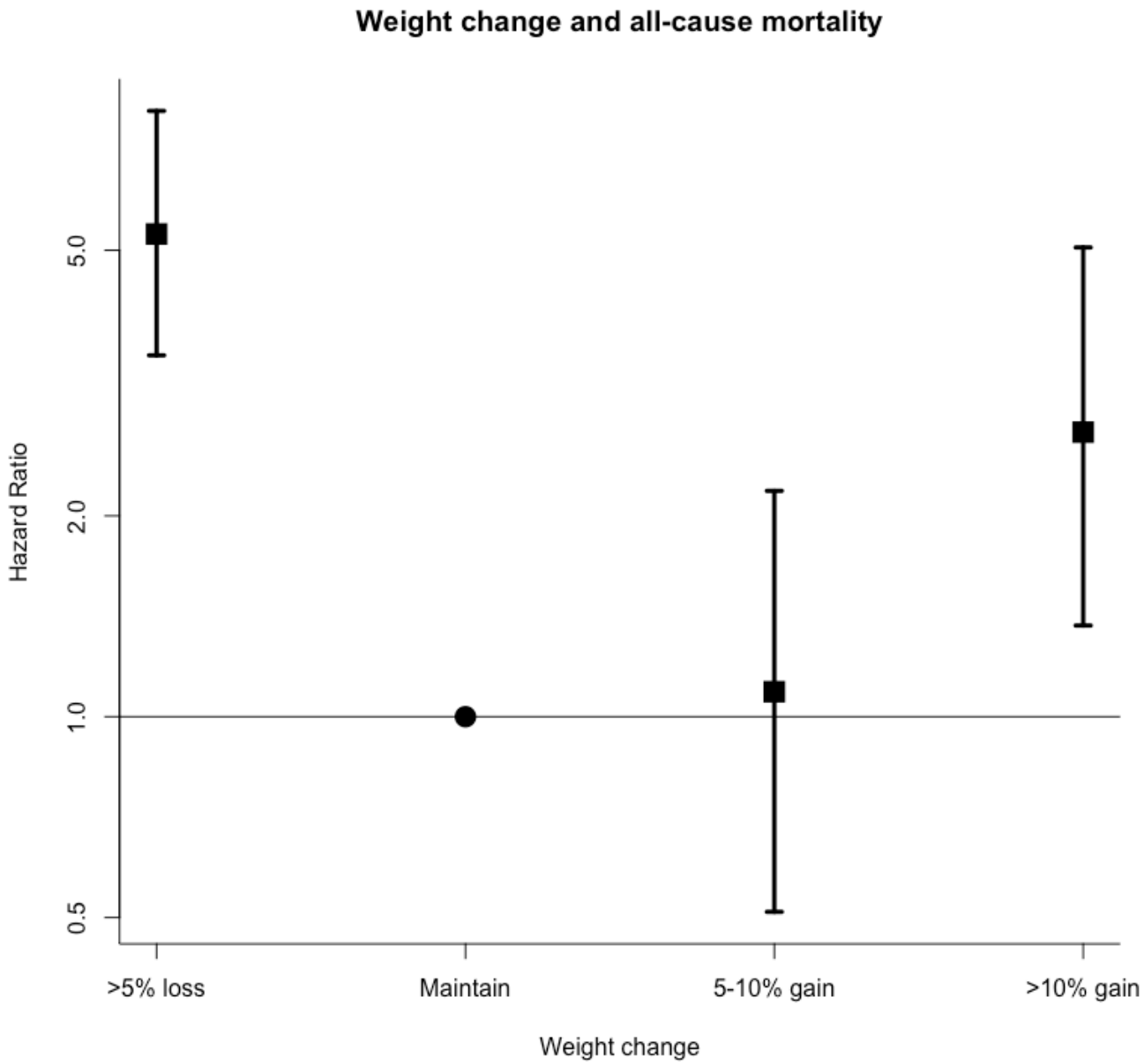


Analysis

- Approx 1/3 did not respond to follow-up questionnaire.
  - Many of these died before re-entered the field.
  - Concerned with selection bias.
  - Used selection model approach: jointly **model probability of remaining in sample** and time-to-death.*

* Bradshaw PT, *et al*. A Bayesian proportional hazards regression model with non-ignorably missing time-varying covariates. *Stat Med*. 2010.

Study results



Similar for breast cancer-specific deaths.

Study results

- **Weight gain after diagnosis** associated with greater risk of death.
- **Weight loss** also associated with greater risk of death.
  - **Difficult to exclude reverse causality**.
    - Analytical methods unlikely to be able to dig us out.
  - Weight loss intentionality important.

Example: Cardiovascular Disease Mortality among Breast Cancer Survivors

Cardiovascular disease mortality

- **5-year breast cancer survival**: ~90%.
- With longer follow-up, **BrCa survivors dying from other causes**:
  - After 5 years: other causes of death start to dominate.
- **Heart disease** accounts for >35% of non-breast cancer deaths among survivors >50 years of age.
  - Late effects of treatments documented.
- **Magnitude of excess burden** of CVD not clear.

Cardiovascular disease mortality

- **Objective**: compare CVD mortality among women with BrCa compared to the control group from the case-control study.
  - Will estimate **hazard ratios** (similar to rate ratio) for mortality.
  - **Competing risk**: death from breast cancer.

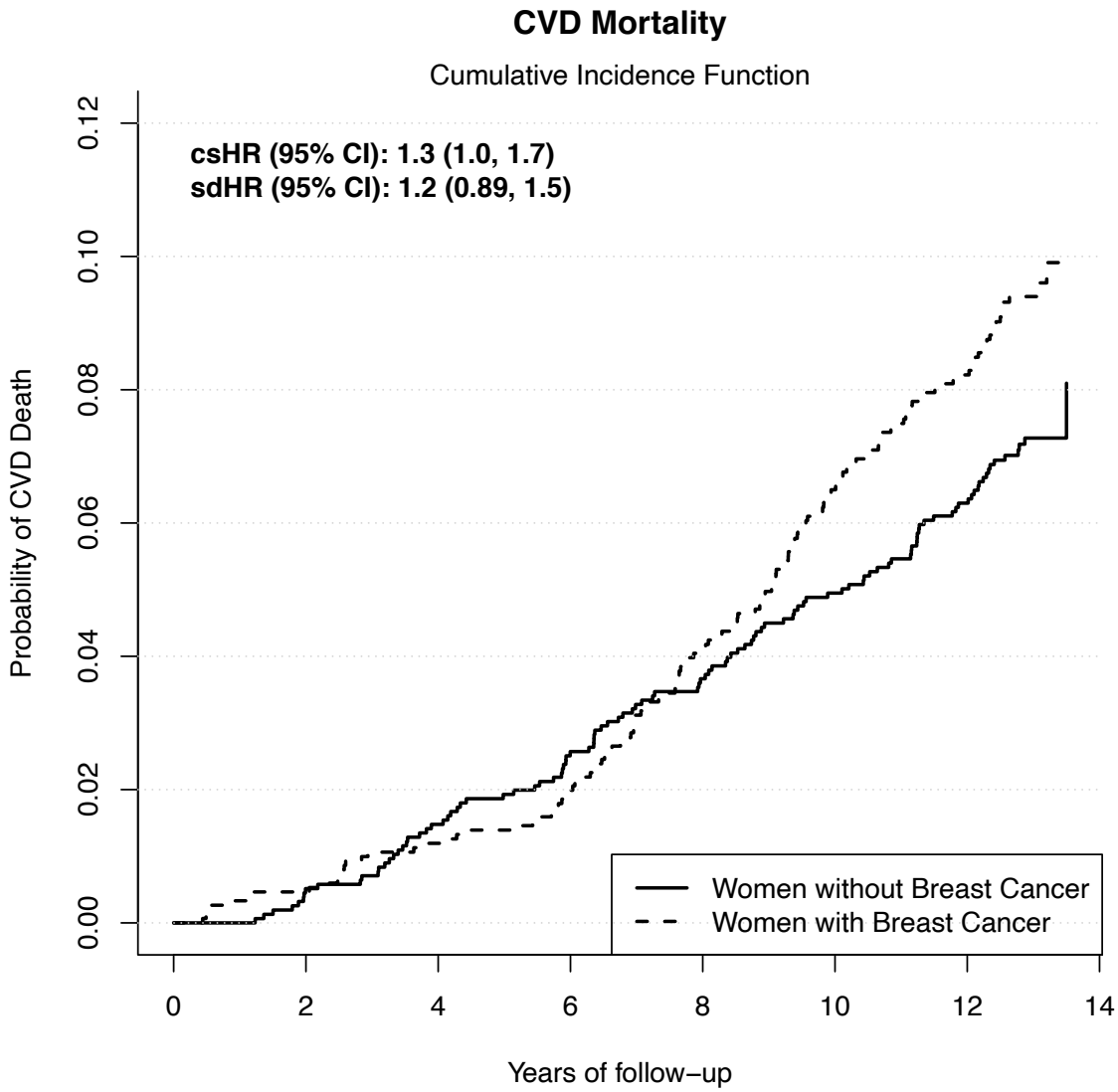
Cardiovascular disease mortality

- **Cause-specific** proportional hazards models for CVD-related mortality:
  - **Censor** (administratively remove from risk set) when competing event occurs (e.g. death from BrCa, accident, etc…
    - **Cause-specific** hazard ratios.
  - **Characterizes rate of CVD death**, but not risk.
  - *Doesn't account for how fast individuals die from other factors*.

Cardiovascular disease mortality

- Need to account for risk of competing event to determine relative **risk of CVD mortality** between groups.
- Estimated **subdistribution hazard ratios** with Fine-Gray model.
  - Increase/decrease in subdistribution HR implies increase/decrease in risk of event in the presence of competing risks.
- Estimated nonparametric **Cumulative Incidence Function** to plot probability of CVD death over time.
- Kaplan-Meier curves inappropriate in presence of competing risks.

*Bradshaw PT, et al. Cardiovascular disease mortality among breast cancer survivors. Epidemiology. (2016): 1311-1321.



a Approximately 13 years follow-up. HR adjusted for age at diagnosis (restricted cubic spline), menopausal status at diagnosis, HRT use (ever/never), smoking (active/passive), alcohol intake (g/day), BMI 1 year before diagnosis, history of hypertension, diabetes, stroke, myocardial infarction.

*Bradshaw PT, *et al.* Cardiovascular disease mortality among breast cancer survivors. *Epidemiology.* (2016): 1311-1321.

Analysis

Stratified by time since diagnosis:

	Cause-Specific Hazard Ratio[a] (95% CI)		
Disease Status	**Years 0-6.9**	**Years 7+**	**P-ixn**
No Breast Cancer	1.	1.	0.001
Breast Cancer	**0.80 (0.52, 1.24)**	**1.76 (1.25, 2.48)**	

	Subdistribution Hazard Ratio[a] (95% CI)		
Disease Status	**Years 0-6.9**	**Years 7+**	**P-ixn**
No Breast Cancer	1.	1.	<0.001
Breast Cancer	**0.59 (0.40, 0.88)**	**1.92 (1.36, 2.72)**	

a Approximately 13 years follow-up. Models adjusted for age at diagnosis, menopausal status at diagnosis, HRT use (ever/never), smoking (active/passive), alcohol intake (g/day), BMI 1 year before diagnosis, history of diabetes, myocardial infarction, hypertension, dyslipidemia or stroke; includes interaction with time and diabetes.

*Bradshaw PT, *et al.* Cardiovascular disease mortality among breast cancer survivors. *Epidemiology.* (2016): 1311-1321.

CVD Mortality

- **Cause-specific hazard ratios:**
  - BrCa survivors have similar (slightly lower) rate of dying from CVD over first 7 years.
  - **After year 7**, women with breast cancer have higher rate of dying from CVD.
- **Subdistribution hazard ratios:**
  - Women with breast cancer have notably lower risk of dying from CVD over first 7 years.
    - More likely to die from BrCa during this period.
  - After year 7, women with breast cancer have much higher risk of dying from CVD.
    - Combination of increased rate of CVD death + the decreased rate of competing event.

Summary

- Carefully consider design issues in the planning phase.
- Keep common threats to validity in mind.
  - Best to avoid up front.
    - Analytical methods can be useful, but most rely on strong assumptions.
  - Consider bias analysis methods* to quantify influence of systematic errors.
    - Misclassification.
    - Uncontrolled confounding.
    - Selection bias.

Acknowledgements

- For LIBCSP studies presented:
  - Marilie D. Gammon, Ph.D. (PI)
  - Joseph Ibrahim, Ph.D.
  - other LIBCSP collaborators.
- Funding sources:
  - NCI R01 LIBCSP parent projects.
  - For Dr. Bradshaw:
    - NCI T32 Cancer Epidemiology Training Grant to Andrew Olshan, Ph.D.
    - Susan G. Komen Postdoctoral Fellowship.
    - Marilyn Gentry Fellowship, American Institute for Cancer Research.