Recurring challenges in studying cancer recurrence

Jessica Chubak, PhD

February 12, 2019

UNC Lineberger Cancer Outcomes Research Program Breakfast Seminar

Acknowledgments

Denise Boudreau	R01CA1205621 (COMBO)
Erin Bowles	R01CA172973 (RECORD)
Diana Buist	CRTG-03–024-01-CCE (ACS)
Ruth Etzioni	R01CA09377 (BOW)
Monica Fujii	UG3CA218909-01 (ReCAPSE)
Rebecca Hubbard	R21CA143242 (BRAVA)
Thomas Lumley	T32CA09168
Jennifer Lund	
Barbara McKnight	
Gaia Pocobelli	
Onchee Yu	
Noel Weiss	

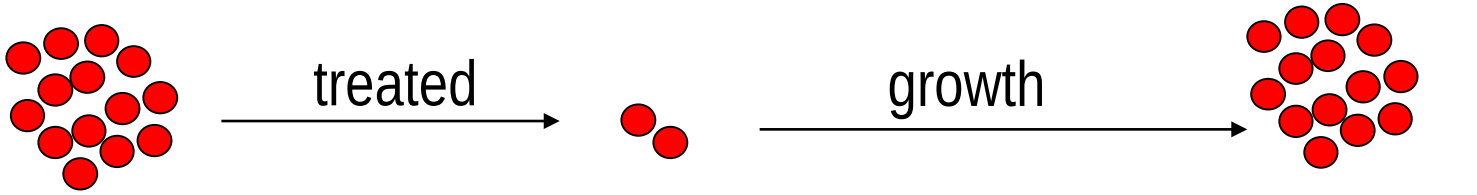
KAISER PERMANENTE®

Outline

- Definitions
- Measurement
- Analysis
- Ongoing and future research

Cancer recurrence

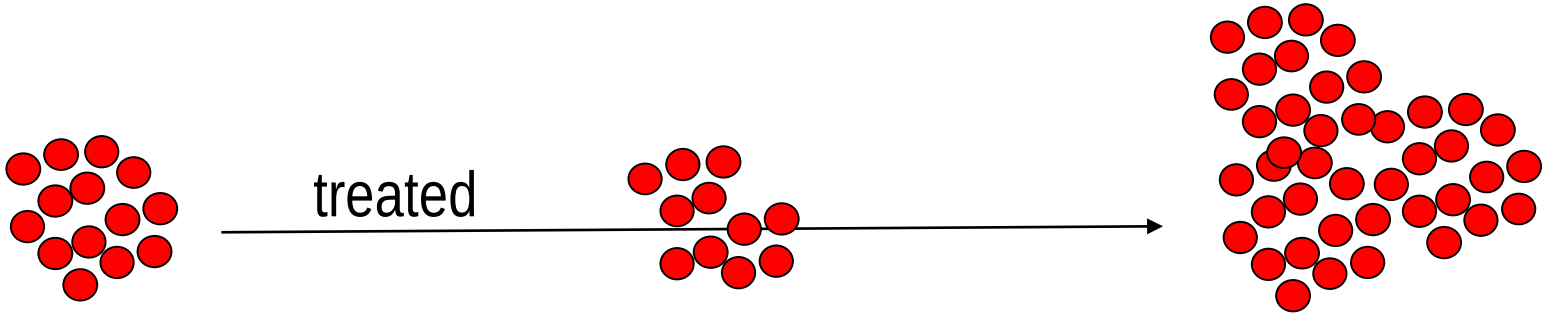
Re-emergence of a cancer in patients who were clinically disease-free



KAISER PERMANENTE®

Progression

Cancer growth (even on treatment)

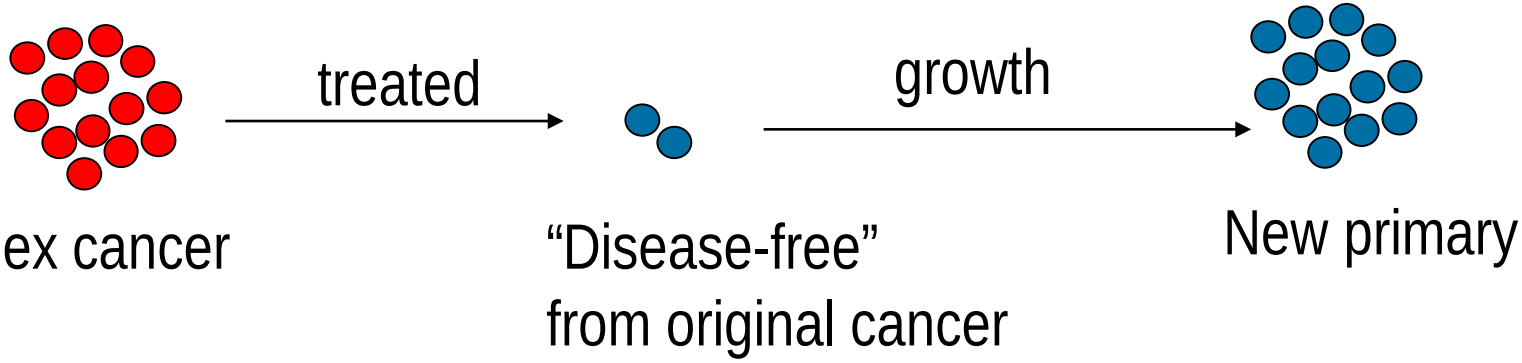


Index cancer

KAISER PERMANENTE®

New primary

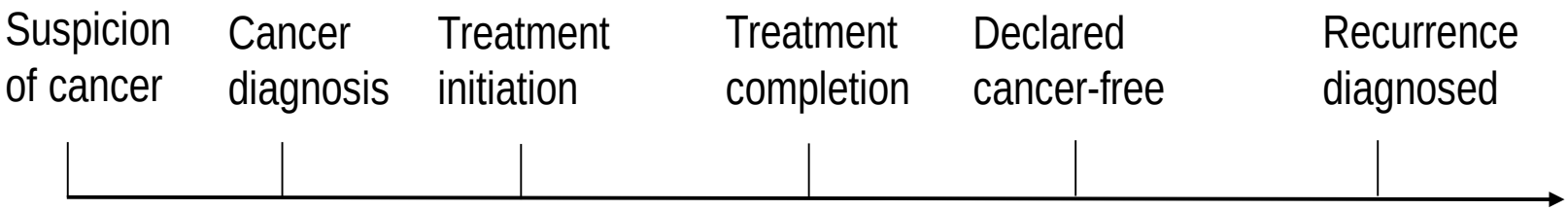
A new cancer in a patient with a prior cancer



KAISER PERMANENTE®

Cancer recurrence sequence

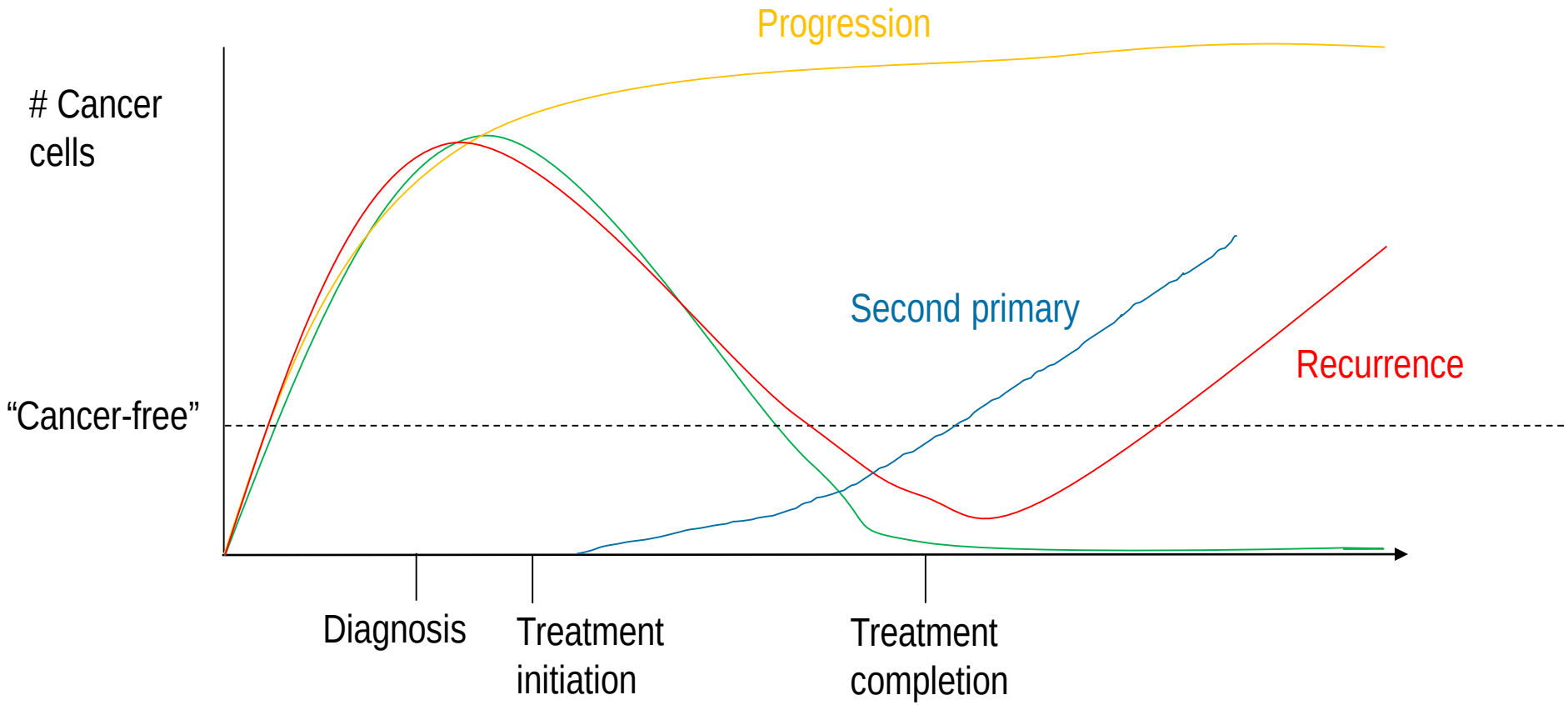
Clinical



KAISER PERMANENTE®

Cancer recurrence sequence

Clinical and biological



KAISER PERMANENTE®

MEASUREMENT

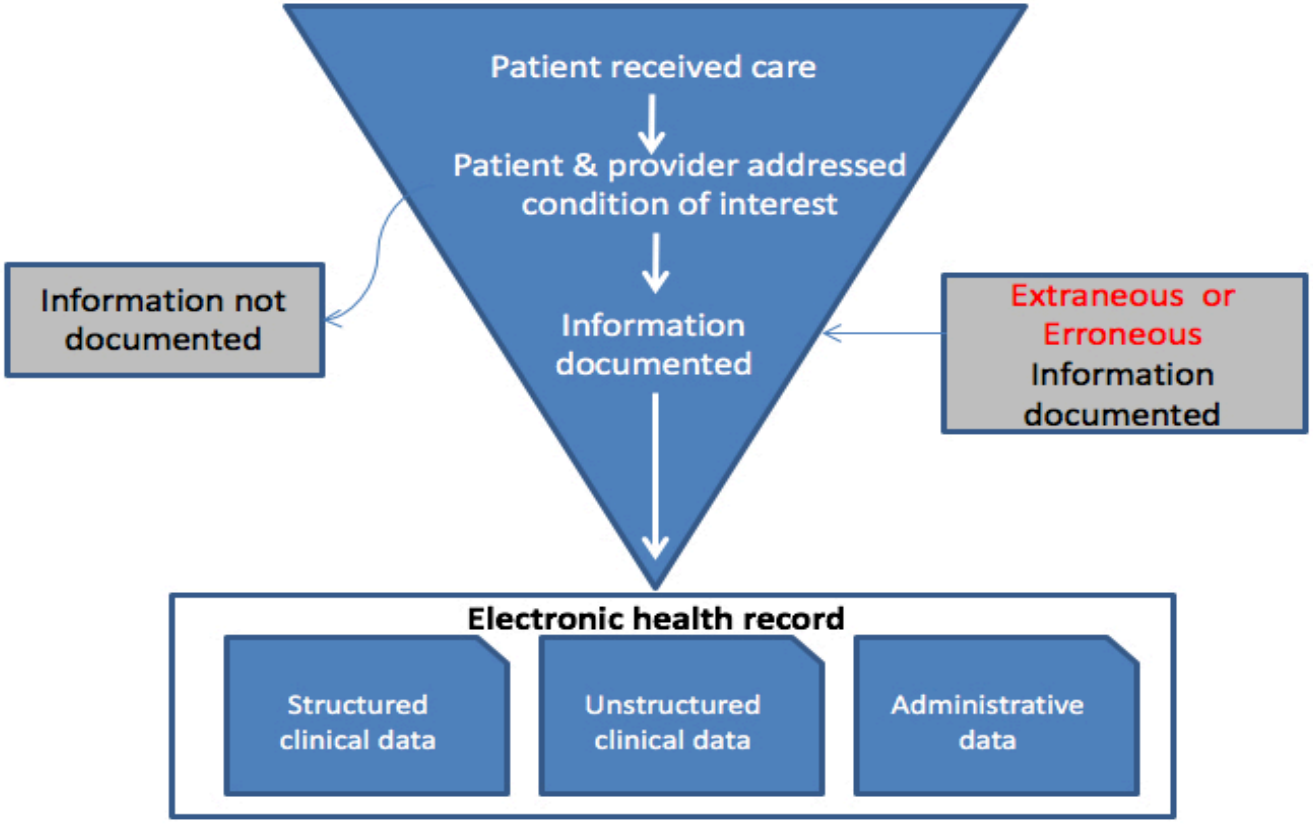
Challenges in identifying recurrences

- Differentiating recurrence from other outcomes
- Timing of recurrence may be ill-defined
- Data sources do not record recurrence well or at all

Data sources

Administrative	• Procedure codes • Diagnosis codes
Structured clinical data	• PRO forms • Laboratory results
Unstructured clinical data	• Clinical notes • Pathology reports • Imaging reports
Tumor registries	• Population-based • Hospital-based

EHR and administrative data capture



KAISER PERMANENTE®

Administrative data

Pros

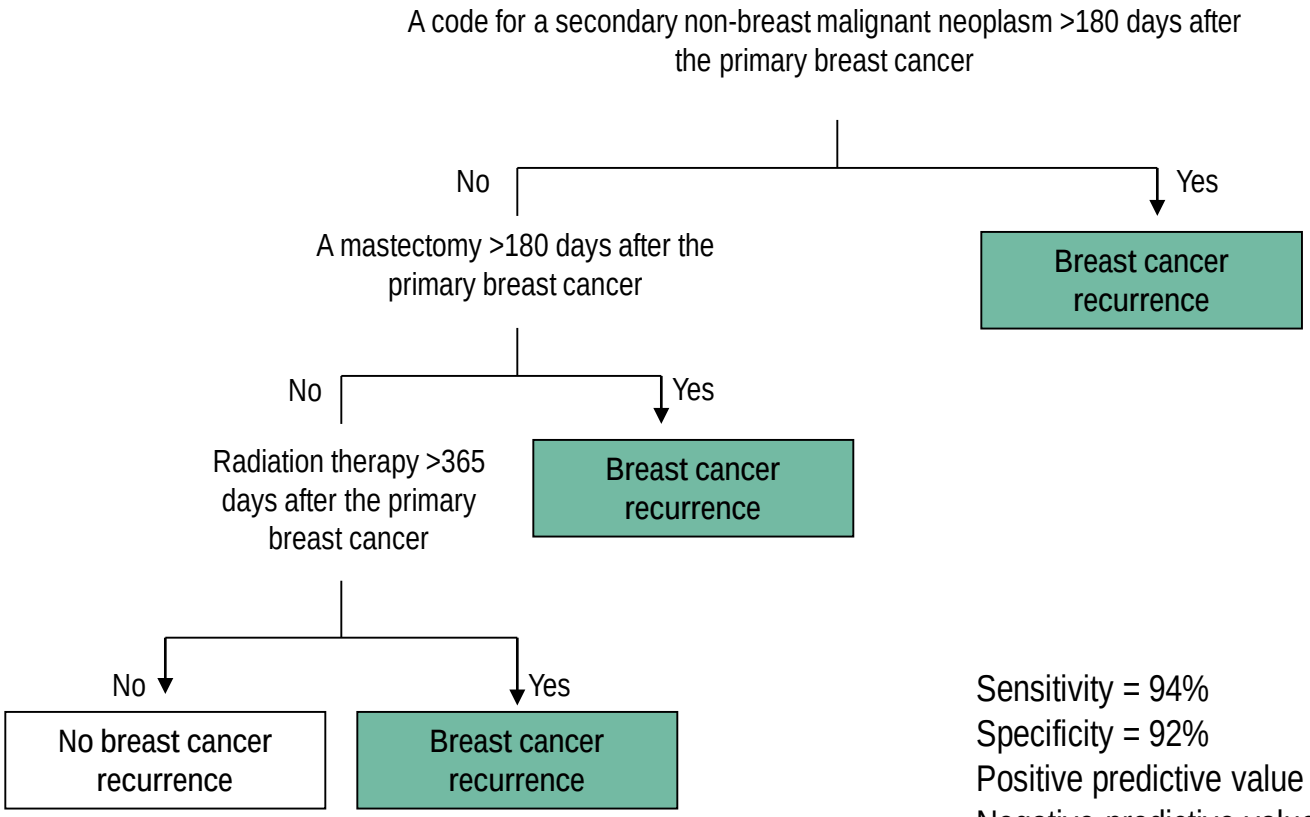
- Large
- Linkages may be possible
- Often “representative” populations

Cons

- Algorithms usually required
- Misclassification
- May not be portable
- Coding changes over time

KAISER PERMANENTE®

Example: high sensitivity algorithm for breast cancer recurrence



Sensitivity = 94%
Specificity = 92%
Positive predictive value = 58%
Negative predictive value = 99%

Chubak et al, 2012. JNCI:104(12): 931-940.

KAISER PERMANENTE®

Electronic health record data

Pros

- Often includes claims or administrative data
- Rich information
- Cancer specific modules

Cons

- Data often unstructured
- Externally received care may not be document

SEER registry

SEER collects multiple primaries, not recurrences

- Registrar determines the primary site for a case
- Use site-specific rules for determining number of primaries
- Recurrences are part of a single primary

Hospital-based cancer registries

Pros

- Hospital-based registries collect recurrence (e.g., FORDS)
- NAACCR provides structure for collecting recurrence

Cons

- FORDS manual based on physician diagnosis
- Denominator might not be clear

Alternative data sources for studying cancer recurrence

- Ongoing clinical trials
- Cohort studies with patient report
- Data consortia
  - ASCO CancerLinQ
- International datasets
  - Public Health England's National Cancer Intelligence Network
- Others?

Tradeoffs

Population size

Selection bias

Detection bias

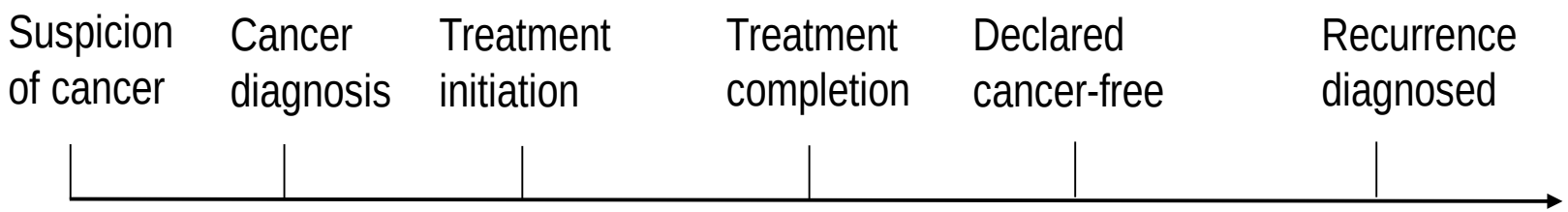
ANALYSIS

Analytic decisions to make

- When does at risk time begin?
- What exposure periods are of interest?
- What time scale should we use in survival analysis?
- How well can confounding by pre-diagnosis exposures be controlled?
- Should exposure be lagged?
- How should we handle people who die of the cancer of interest without a recurrence?
- How should we account for different surveillance regimens?
- When should we combine recurrence with other outcomes?
- How should we handle competing events?

When does “at risk” time for recurrence begin?

- After confirmation (or assumption) of disease-free status?
- After treatment completion?
- At diagnosis of primary?
- Before diagnosis?



What exposure periods are of interest?

- Relative to clinical events
  - After confirmation (or assumption) of disease-free status?
  - After treatment completion?
  - At diagnosis?
  - Before diagnosis?
  - Between any of these intervals?
- Relative to start of follow-up
  - Before follow-up?
  - At start of follow-up?
  - During follow-up?

What time scale should we use in survival analysis?

- Time since diagnosis?
- Time since treatment initiation?
- Time since treatment completion?
- Time since disease-free confirmation/assumption?

My original approach

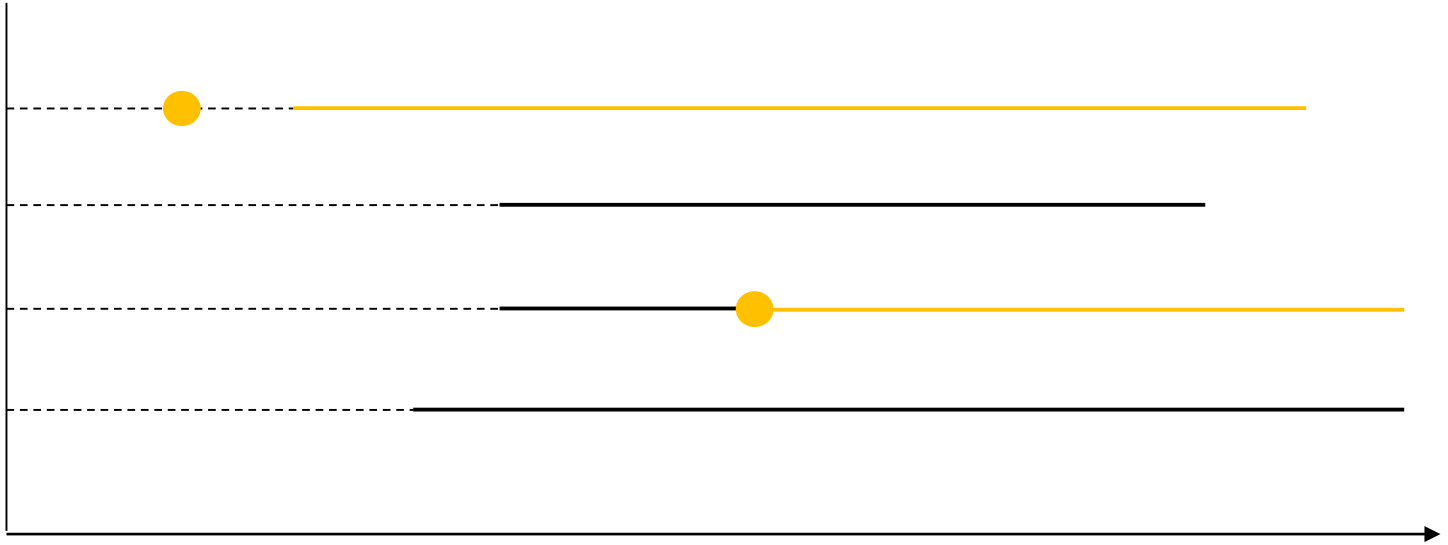
- Focus on people presumed to be disease-free
- Start follow-up when eligible to recur
- Use a meaningful time scale in survival analyses
- Consider hypothesized exposure-outcome mechanisms
- Avoid immortal time bias with time-varying exposures

Example

Is antidepressant use after early stage breast cancer associated with risk of recurrence?

Antidepressants and risk of breast cancer recurrence

Cohort entry:	Surgery + 120 days
Time scale:	Time since diagnosis
Exposure period:	Any time after diagnosis



Time since diagnosis

KAISER PERMANENTE®

Methods that have influenced me recently

- New user/treatment decision design (Ray, Brookhart, Stürmer, Lund)
- Trial emulation (Hernán, Robins, Danaei)

New user / treatment decision designs

- Ask questions about starting, stopping, switching exposures
- Questions are clinically relevant
- Avoids “depletion of susceptibles”

Examples of new user / treatment decision questions

- Should antidepressant users stop or continue use at cancer diagnosis?
- Should depressed patients who have not used antidepressants before a cancer diagnosis initiate antidepressant use at diagnosis?

Trial emulation

- Clinically relevant questions
- Pre-specify:
  - Eligibility requirements
  - Population
  - “Intervention” and comparator(s)
  - Outcome
  - Time frame

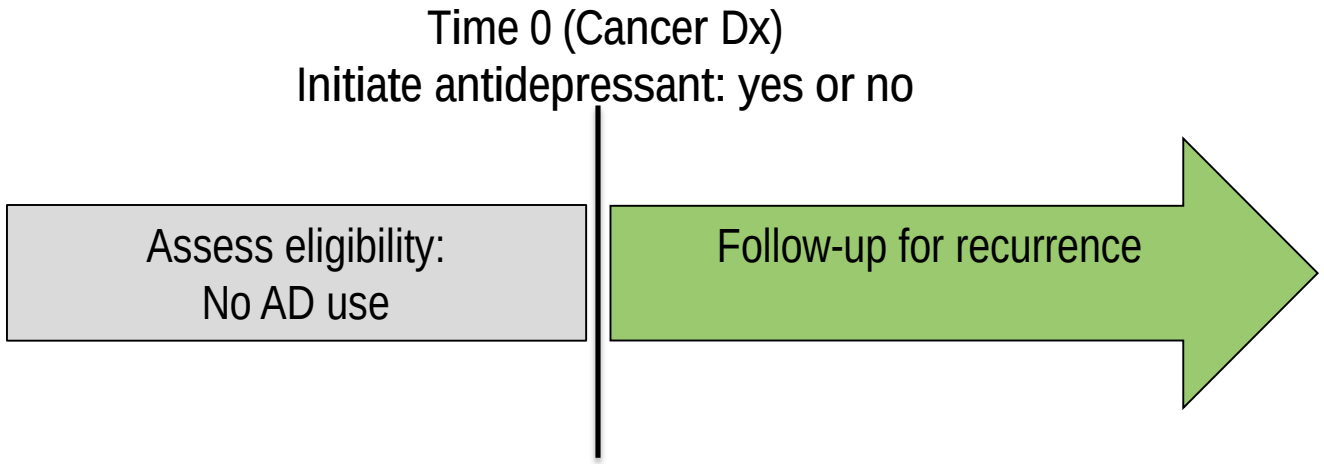
Should depressed patients who have not used antidepressants before a cancer diagnosis initiate antidepressant use at diagnosis?

Population	Early stage breast cancer patients with depression who have not filled an antidepressant rx in the year before cancer diagnosis
Intervention	Fill antidepressant at diagnosis
Comparator	Do not fill antidepressant at diagnosis
Outcome	Recurrence
Time frame	Maximum of 5 years follow-up between 1995 and 2010

Time 0

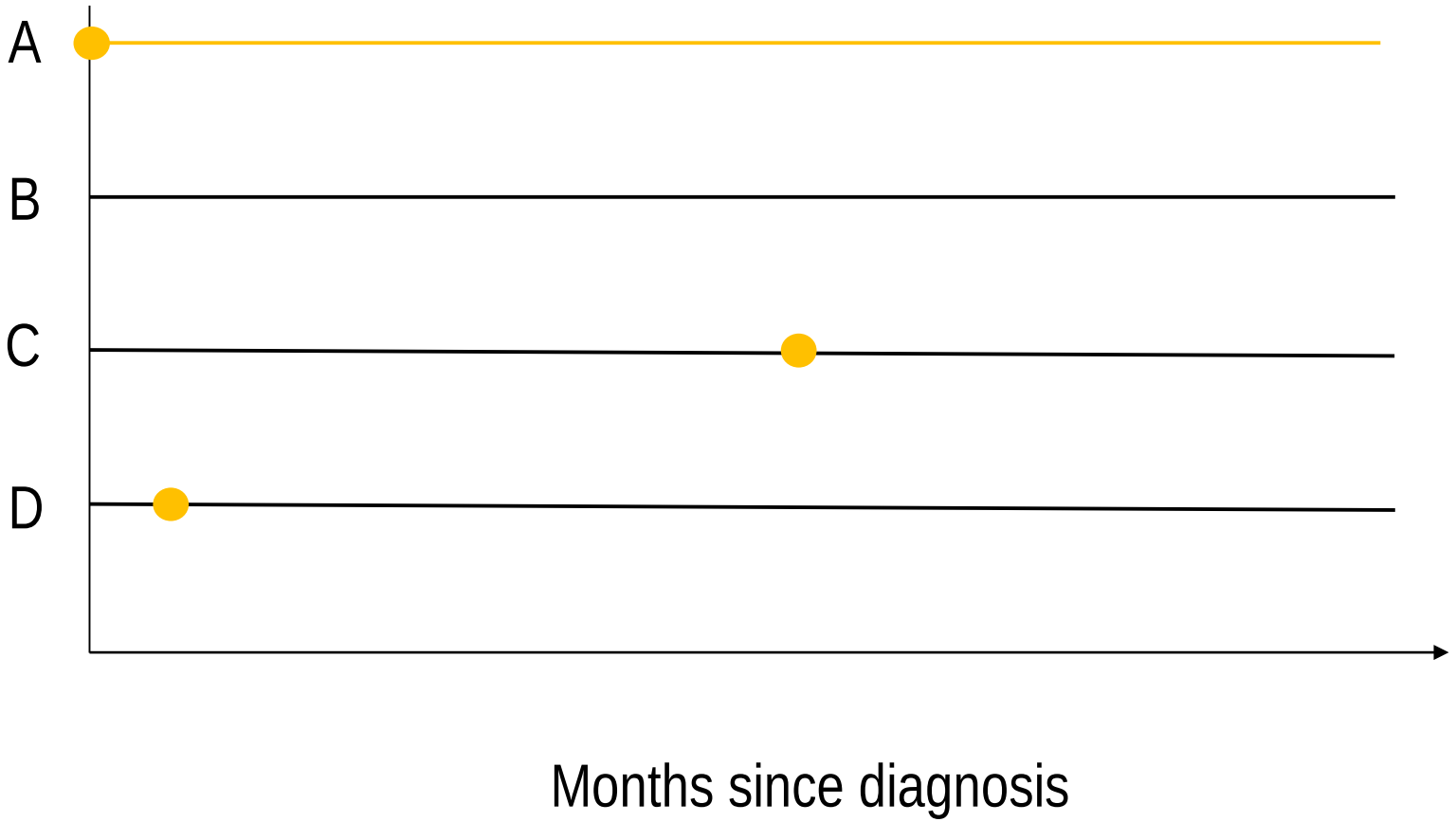
At time 0:

1. Evaluate eligibility
2. “Assign” treatment
3. Start follow-up



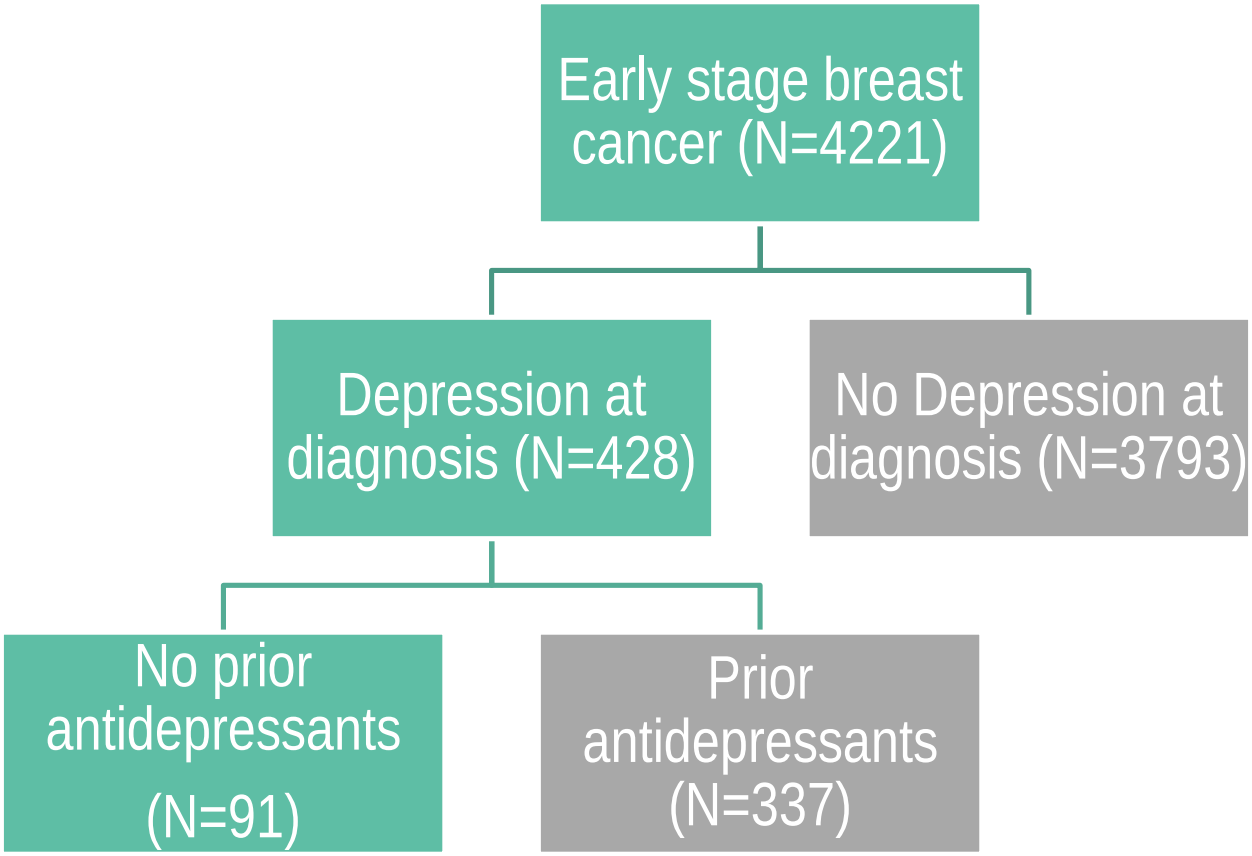
KAISER PERMANENTE®

Example



KAISER PERMANENTE®

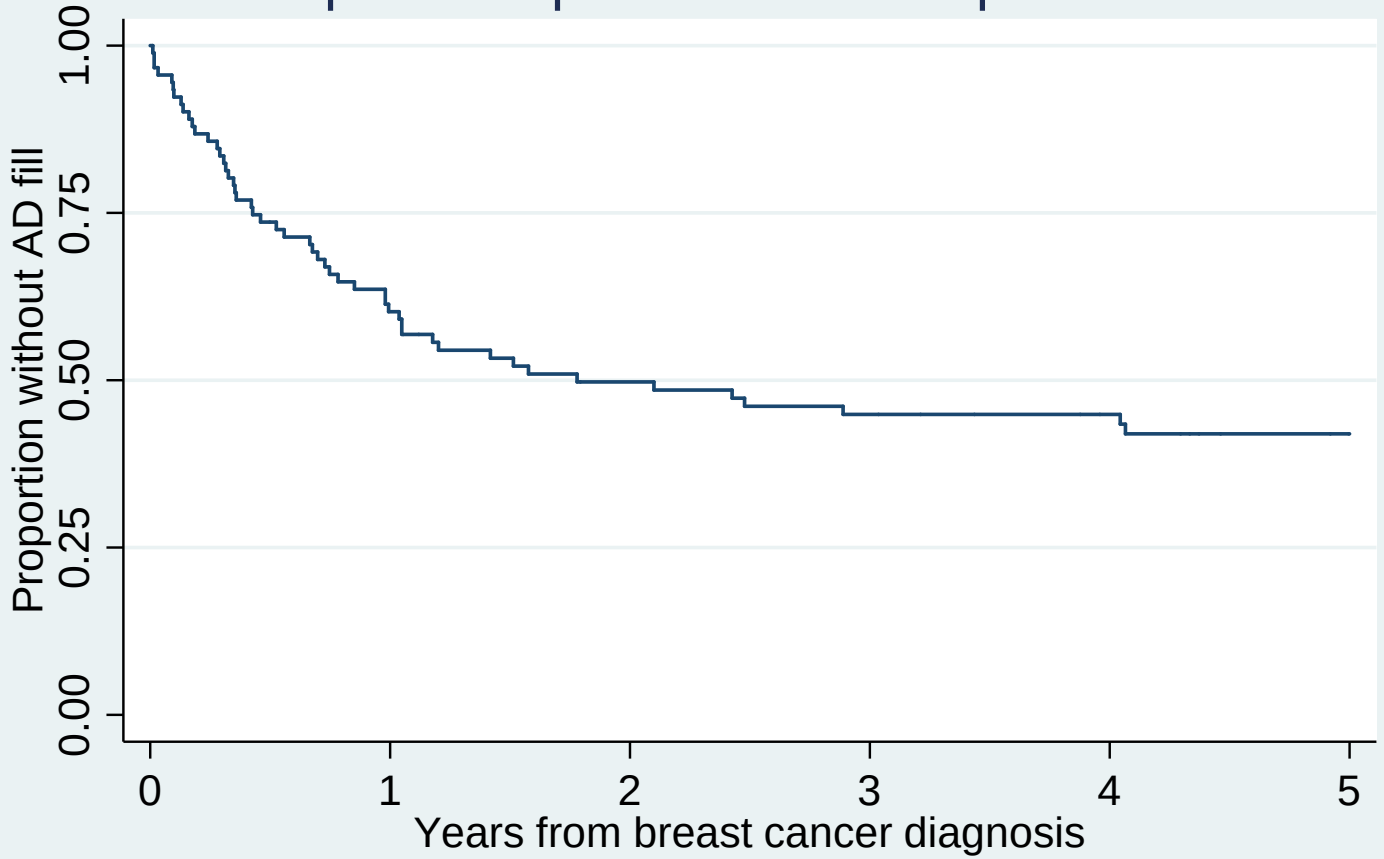
Practical challenge: small numbers



Antidepressant fill on diagnosis date = 0

KAISER PERMANENTE®

Time from breast cancer diagnosis to antidepressant prescription fill in depressed patients with no prior AD use



Number at risk

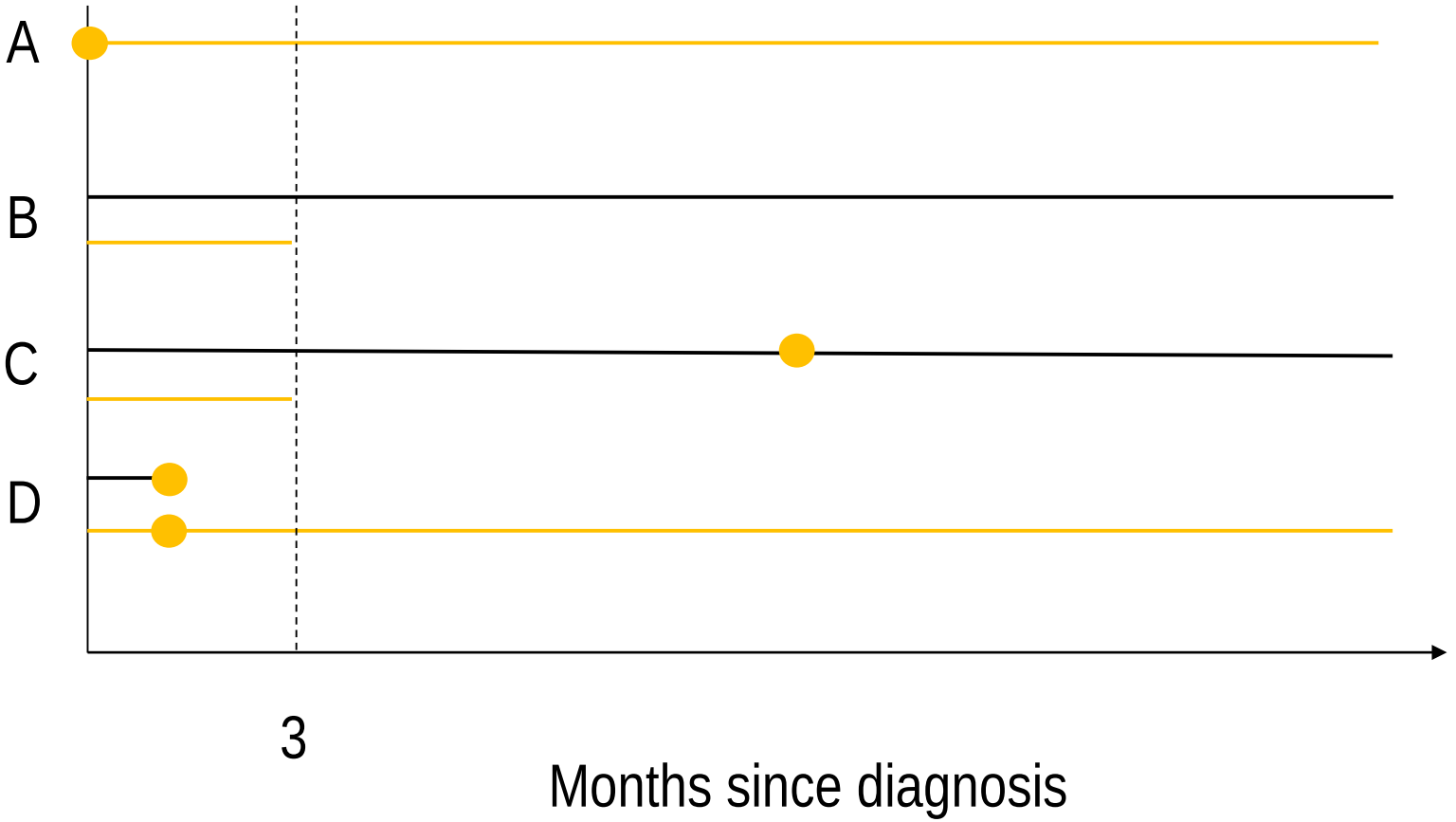
0	1	2	3	4	5
91	54	41	37	31	23

KAISER PERMANENTE®

Possible solutions

- Find large datasets that have recurrence data
- Use grace period

Example – clones to allow for grace periods



KAISER PERMANENTE®

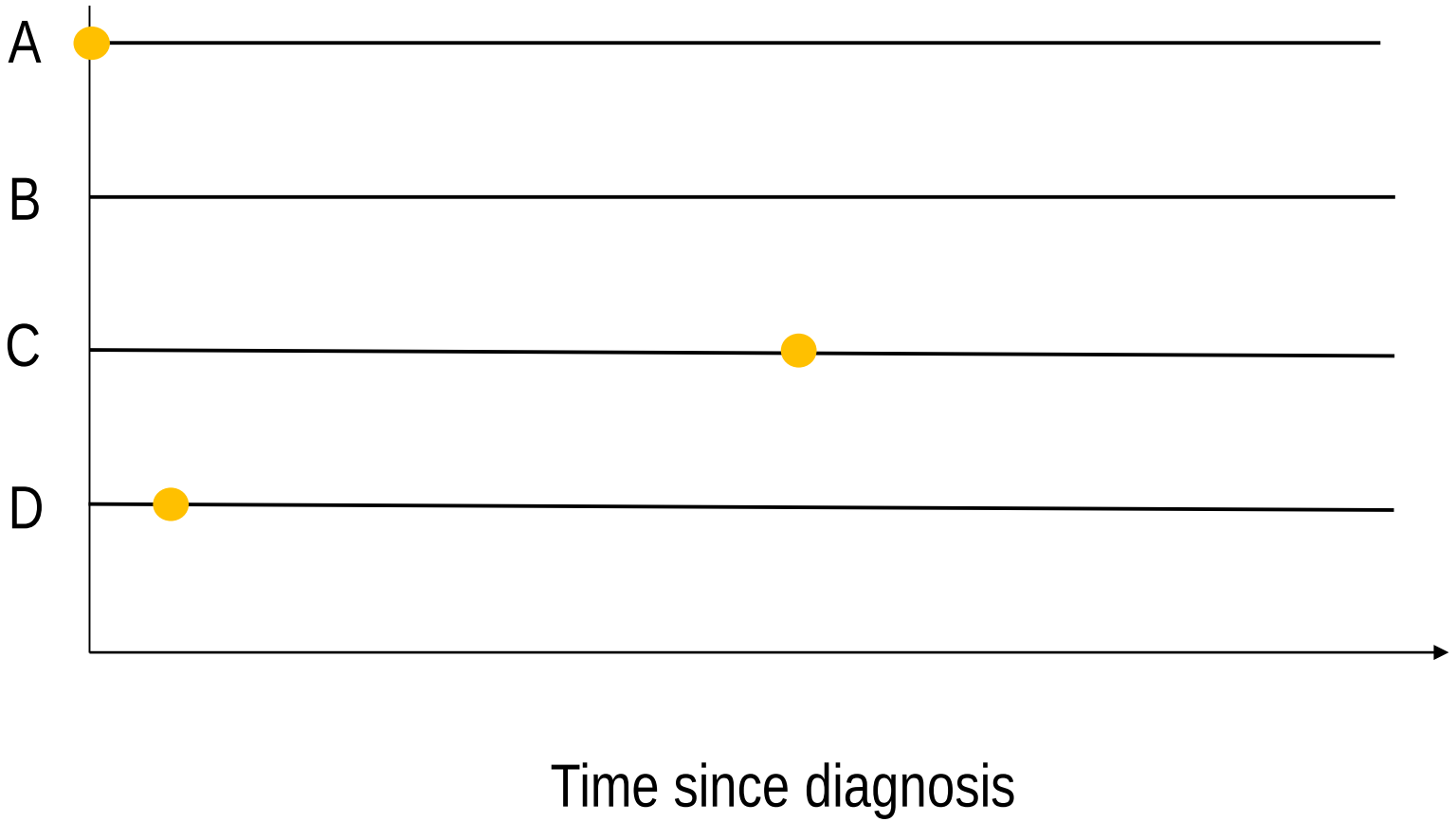
Possible solutions

- Find large datasets that have recurrence data
- Use grace period
  - Analysis is complex, or at least unfamiliar
- Consider other questions, designs, or populations
  - Indirect and potentially biased

Should depressed patients who have not used antidepressants before a cancer diagnosis initiate antidepressant use at or after diagnosis?

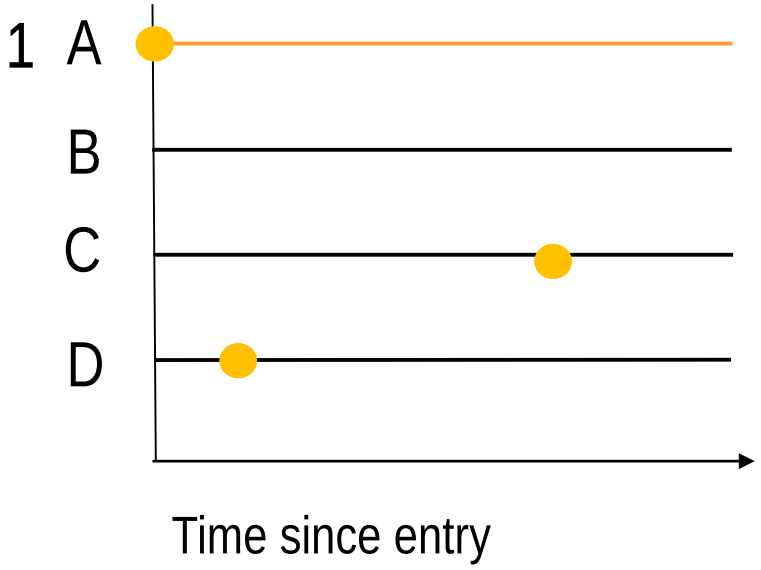
Population	Early stage breast cancer patients with depression who have not filled antidepressants prescriptions in the year before diagnosis or since diagnosis
Intervention	Fill antidepressant at or after diagnosis
Comparator	Do not fill antidepressant at or after diagnosis
Outcome	Recurrence
Time frame	Maximum of 5 years follow-up between 1995 and 2010

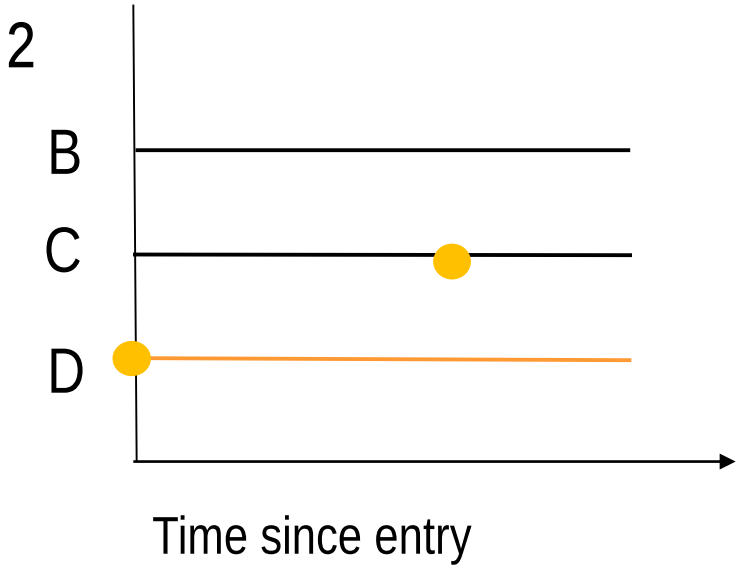
Example

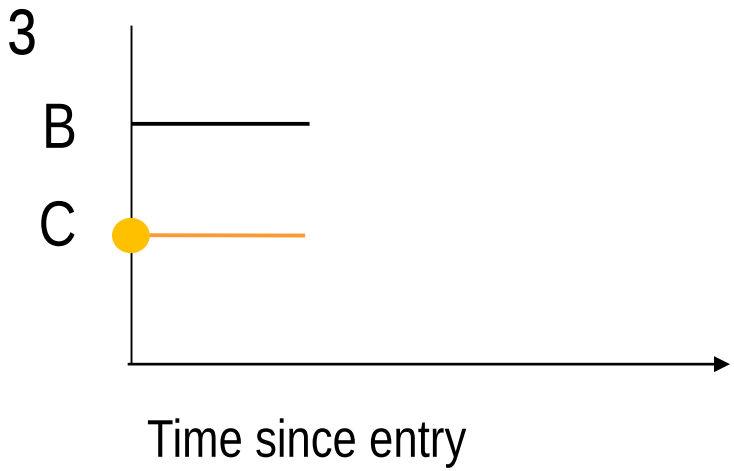


KAISER PERMANENTE®

Example, continued







Remaining analytic decisions

- Granularity of timescale (days or months)
- How to define/handle progression on treatment
- How to handle death (competing risk)
- How to handle different surveillance patterns (detection bias)
- How to handle non-adherence

ONGOING RESEARCH AND FUTURE GOALS

Ongoing research in methods for ascertaining cancer recurrence

- PRISSMM (Schrag)
- ReCAPSE (Etzioni, Chubak, Huang)

PRISSMM Overview

PRISSMM is a **standard taxonomy** for classification and communication of structured information about cancer status and treatment outcomes following specification of baseline information for patients with solid tumors

Each letter in PRISSMM corresponds to a dimension of cancer status or treatment response.

Courtesy of D. Schrag

PRISSMM:
A Taxonomy for Defining Cancer Outcomes

Pathologic evidence of locoregional or distant evidence of tumor

Radiographic evidence of locoregional recurrent or persistent tumor

Imaging evidence of distant/disseminated tumor beyond the primary site

Symptoms of tumor on physical exam or symptoms that can be attributed to tumor

Signs of cancer on physical exam or symptoms that can be attributed to tumor

Tumor **M**arker evidence of persistent or recurrent tumor

Oncology **M**edical Provider assessment

Each curation effort may focus on some or all of the PRISSMM components
Signs may be relevant for melanoma outcomes
Markers may not be relevant for lung outcomes

Courtesy of D. Schrag

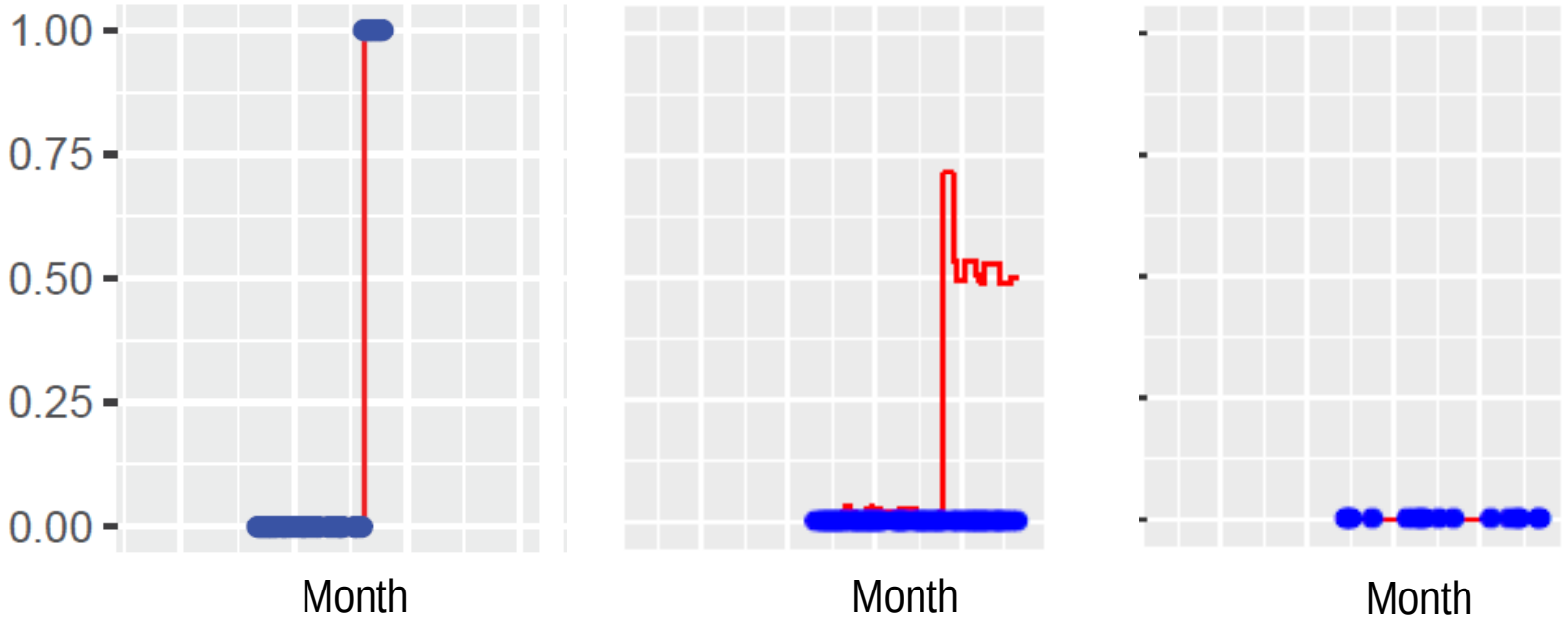
ReCAPSE

Recurrence from Claims And PROs for SEER Enhancement

- Breast cancer second events
- Claims and self-report
- Machine learning approaches to classify periods as:
  - Pre-recurrence or second event
  - Post-recurrence or second event
- Input
  - Patient and index cancer characteristics and treatment
  - Counts of code grouping, time since last and next grouping

Person level predictions vs. gold standard

Probability of second breast cancer event



Actual

Predicted

KAISER PERMANENTE®

Person-level prediction accuracy

Seattle area study: SEER records + claims

Probability threshold	Specificity	Sensitivity
0.1	0.89	0.76
0.2	0.93	0.76
0.3	0.94	0.71
0.4	0.94	0.71
0.5	0.94	0.71

Person-level prediction accuracy

KP Washington study: SEER records + utlization

Probability threshold	Specificity	Sensitivity
0.1	0.96	0.96
0.2	0.97	0.96
0.3	0.97	0.96
0.4	0.98	0.95
0.5	0.98	0.92

My wish list

- Agreement on definition(s) of recurrence
- Large datasets with recurrence
- Inclusion of recurrence in population-based cancer registries
- Structured collection of recurrence in EHRs
- Validation of algorithms in diverse settings
- Guidance on analytic methods for studies of cancer recurrence

Thank you

Jessica.Chubak@kp.org

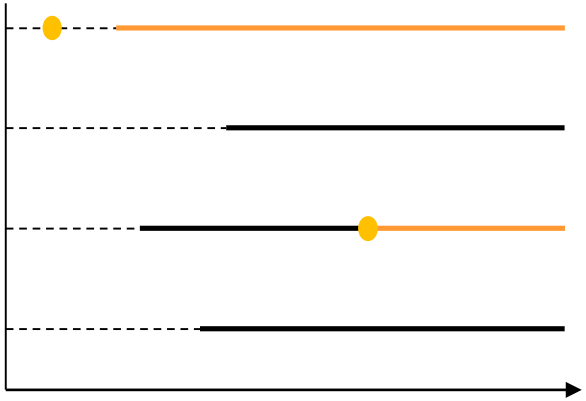
KAISER PERMANENTE®

How to avoid selection bias

- Follow everyone from cancer diagosis forward
- Don’t condition on future events
- Account for left truncation
- Beware of and account for differential loss to follow-up

Alternatives

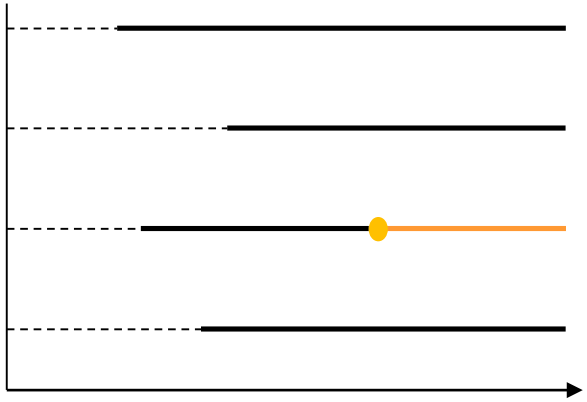
Exposure starting at diagnosis + 120 d



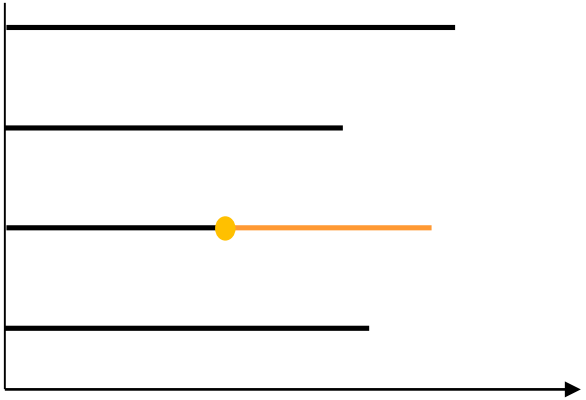
Time since diagnosis

Time since surgery + 120

Exposure starting surgery + 120 d



Time since diagnosis



Time since surgery + 120

Comparison of alternatives

Antidepressant use and breast cancer recurrence

Time scale	Start of exposure	Hazard ratio (95% CI)
Diagnosis	Diagnosis	1.2 (0.8-1.8)
Diagnosis	Surgery + 120 days	1.2 (0.8-1.9)
Surgery + 120 days	Diagnosis	1.2 (0.8-1.9)
Surgery + 120 days	Surgery + 120 days	1.2 (0.8-1.9)

Example: breast cancer

Rule M8

Abstract **multiple primaries**[ii] when the patient has a subsequent tumor after being **clinically disease-free** for **greater than five years** after the original diagnosis or last recurrence.

Note 1: The rules are hierarchical. This rule **only** applies when there is a **subsequent breast tumor.**

Note 2: **Clinically** disease-free means that there was **no evidence** of recurrence on follow-up.

- Mammograms are NED
- Scans are NED
- Tumor biomarkers are NED

Note 3: When there is a recurrence less than or equal to five years of diagnosis, the "**clock**" starts over. The time interval is calculated from the **date of last recurrence.** In other words, the patient must have been disease-free for greater than five years from the date of the last recurrence.

Note 4: When it is **unknown/not documented** whether the patient had a recurrence, use **date of diagnosis** to compute the time interval.

Note 5: The physician may state this is a **recurrence,** meaning the patient had a previous breast tumor and now has another breast tumor. **Follow the rules;** do not attempt to interpret the physician's statement.

Rule M11

Abstract a **single primary**[i] when separate/non-contiguous tumors are on the same row in Table 3 in the Equivalent Terms and Definitions. Timing is irrelevant.

Note 1: The tumors **must be the same behavior**. When one tumor is in situ and the other invasive, continue through the rules.

Note 2: The same row means the tumors are:

- The same histology (same four-digit ICD-O code) **OR**
- One is the preferred term (column 1) and the other is a synonym for the preferred term (column 2) **OR**
- A NOS (column 1/column 2) and the other is a subtype/variant of that NOS (column 3)

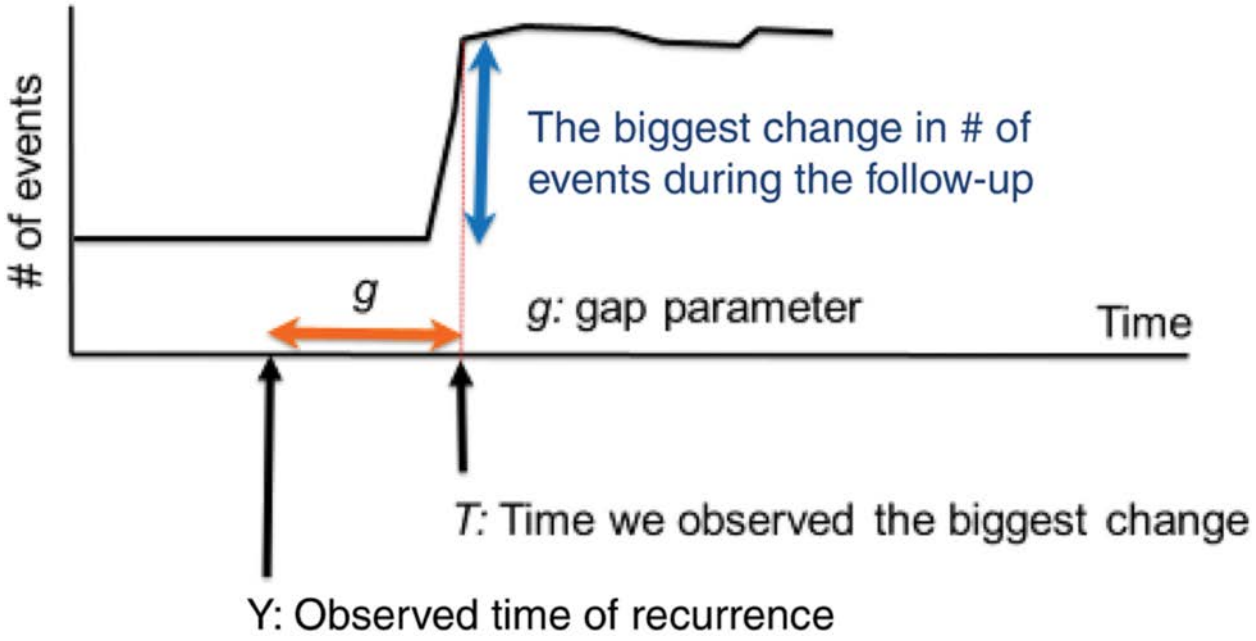
Possible emulation failures

1. Time 0 is set after both eligibility and strategy assignment
2. Time 0 is set at eligibility but after strategy assignment
3. Time 0 is set before eligibility and strategy assignment
4. Time 0 is set at eligibility but strategy is assigned after time 0

Example 2: two phase algorithms for breast cancer recurrence and timing

1. Recurrence
   - If >34 secondary malignancy codes → recurrence
   - If not, logistic regression to estimate probability of recurrence based on number of codes
   - AUC = 95%
2. Timing
   - Based on biggest change in number of events
   - Prediction error: 11.7%

Ritzwoller, et al. *JNCI.* 110(3):273–281

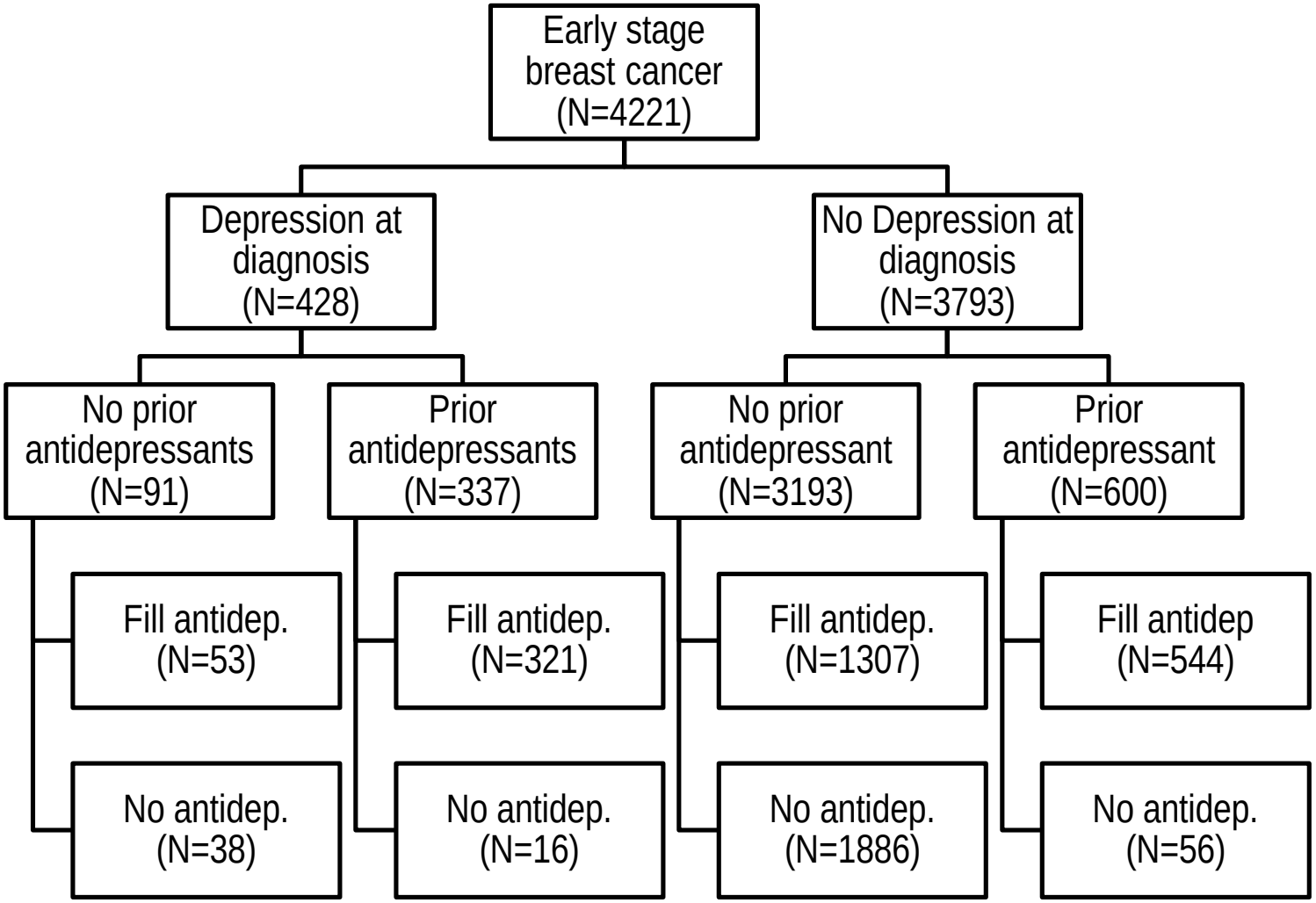


From: Development, Validation, and Dissemination of a Breast Cancer Recurrence Detection and Timing Informatics Algorithm

J Natl Cancer Inst. 2017;110(3):273-281. doi:10.1093/jnci/djx200

KAISER PERMANENTE®

Practical challenge: small numbers



KAISER PERMANENTE®

Decisions to make

- Eligibility
  - How far to look back to exclude AD users
- Exposure
  - How long a grace period to allow for initiation (requires a grace period to handle people filling rx between diagnosis and diagnosis + X months)
- Analysis
  - Granularity of timescale (days or months)
  - How to define/handle progression on treatment
  - How to handle death
  - How to handle different surveillance
  - How to handle non-adherence