Behavior within a Clinical Trial and Implications for Mammography Guidelines

Amanda E. Kowalski

Gail Wilensky Professor of Applied Economics and Public Policy

Department of Economics, University of Michigan

February 2022

U.S. Preventive Services Task Force (USPSTF) 2016 Guidelines:

"The decision to start screening mammography in women prior to age 50 years should be an individual one. Women who place a higher value on the potential benefit than the potential harms may choose to begin biennial screening between the ages of 40 and 49 years"

Do current guidelines target mammograms to women most likely to benefit from them?

"The most important harm is the diagnosis and treatment of noninvasive breast cancer that would otherwise not have become a threat to a woman's health, or even apparent, during her lifetime (that is, overdiagnosis and overtreatment)." (Siu, 2016).

Overdiagnosis

Clinical trials provide the best evidence

Kowalski, Amanda. "Mammograms and Mortality: How Has the Evidence Evolved?" *Journal of Economic Perspectives. 2021.*

I Examine Behavior within a Clinical Trial

- Canadian National Breast Screening Study
  - 89,835 patients enrolled
  - Intervention: access to mammograms for 4 to 5 years during active study period
  - Recorded mammograms, even in control group
  - Examined breast cancer and mortality outcomes from 1980 to 2005 (at least 20 years for all participants) through cancer registry and death records (no attrition)
  - Collected risk factors and demographic data

I Examine Behavior within a Clinical Trial

1. Heterogeneous selection: are women who are more likely to receive mammograms different from other women?

2. Treatment effect heterogeneity: are women who are more likely to receive mammograms more likely to experience better or worse health outcomes because of them?

I Examine Behavior within a Clinical Trial

- Clinical trial literature says little about mammography behavior
  - See Nelson (2016)
- Natural experiment literature examines mammography behavior
  - Many papers focus on mammography as an outcome but do not examine selection and treatment effect heterogeneity
    - Kelaher and Stellman (2000); Habermann, Virnig, Riley, Baxter (2007); Kadiyala and Strumpf (2011, 2016); Finkelstein, Taubman, Wright, Bernstein, Gurber, Newhouse, Allen, Baicker, Oregon Health Study Group (2012); Kolstad and Kowalski (2012); Bitler and Carpenter (2016, 2019); Fedewa, Goodman, Flanders, Han, Smith, Ward, Doubeni, Sauer, Jemal (2015); Mehta, Polsky, Zhu, Lewis, Kolstad, Loewenstein, Volpp (2015); Ong and Mandl (2015); Lu and Slusky (2016); Zanella and Banerjee (2016); Cooper, Kou, Dor, Koroukian, Schluchter (2017); Jacobson and Kadiyala (2017); Buchmueller and Goldzahl (2018); Myerson, Tucker-Seeley, Goldman, Lakdawalla (2020)
  - Two papers corroborate the selection heterogeneity that I find but do not examine treatment effect heterogeneity
    - Kim and Lee (2017); Einav, Finkelstein, Oostrom, Ostriker, Williams (2020)

I Examine Behavior within a Clinical Trial

- I build on LATE and MTE literatures from economics
  - Bjorklund and Moffitt (1987)
  - Imbens and Angrist (1994)
  - Heckman and Vytlacil (1999, 2005, 2007)
  - Vytlacil (2002)
  - Brinch, Mogstad, Wiswall (2015)
  - Cornelissen, Dustmann, Raute, Schoenberg (2018)

I Examine Behavior within a Clinical Trial

"Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments." *NBER WP 22363.*
Divided into two papers immediately below.

"Reconciling Seemingly Contradictory Results from the Oregon Health Insurance Experiment and the Massachusetts Health Reform."
Accepted, Review of Economics and Statistics.

"Behavior within a Clinical Trial and Implications for Mammography Guidelines."
Accepted, Review of Economic Studies.

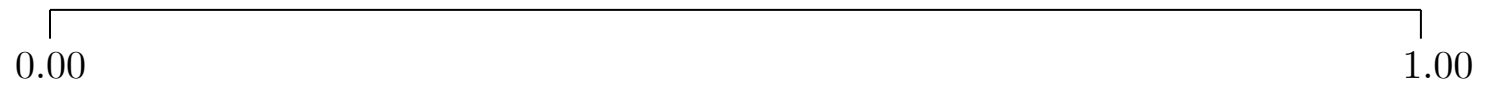"How to Examine External Validity Within an Experiment."
Accepted, Journal of Economics and Management Strategy.

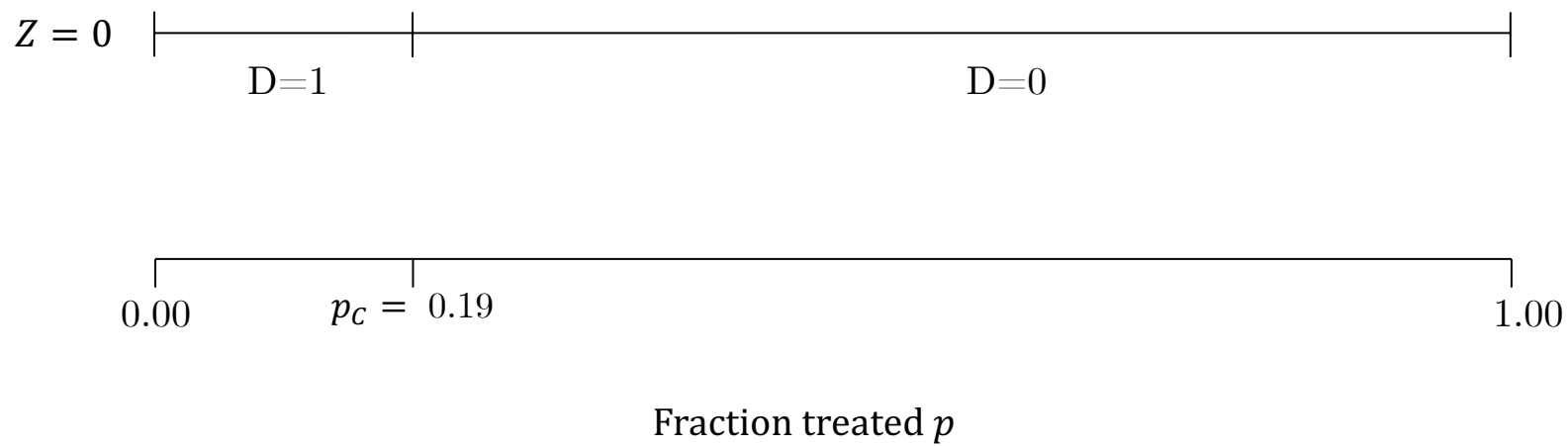"Counting Defiers: Examples from Health Care." arXiv:1912.06739.
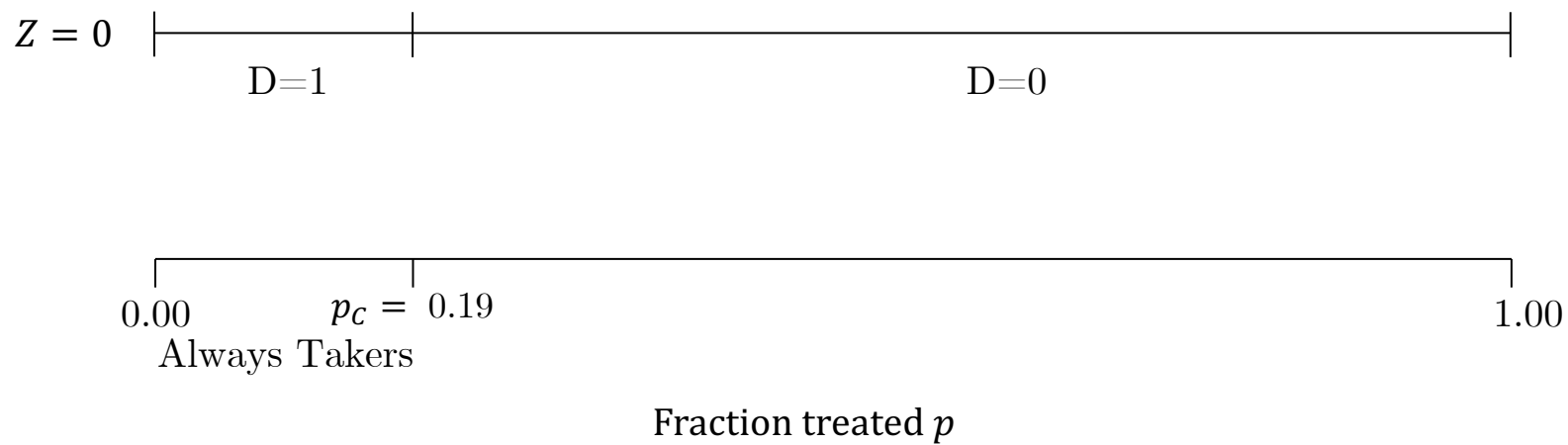*Extending with Jann Spiess.*

Behavior within a Clinical Trial and Implications for Mammography Guidelines
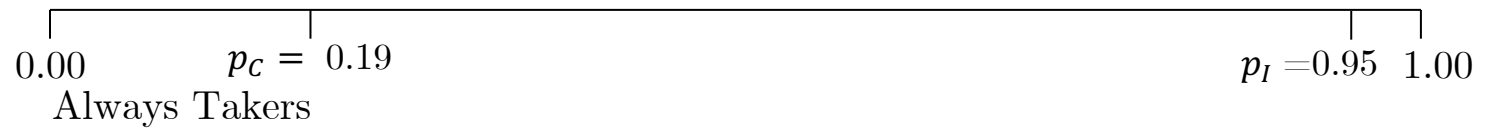
- **Model of Behavior within a Clinical Trial**
  - **First Stage: Mammography**
  - Second Stage: Breast Cancer Incidence
- Results
  1. Selection Heterogeneity
     - *Women more likely to receive mammograms are healthier*
  2. Treatment Effect Heterogeneity
     - *Women more likely to receive mammograms are more likely to be overdiagnosed by them*
- Robustness
- Implications for Mammography Guidelines

0.00

1.00

Fraction treated p

$Z = 0$ |————————————|————————————————————————————————————|

D=1   D=0

|————————————|————————————————————————————————————|

0.00 $p_c = $ 0.19   1.00

Fraction treated p

$Z = 0$ ├─────────────┬──┤

D=1   D=0

0.00 $p_c =$ 0.19   1.00

Always Takers

Fraction treated p

$Z = 1$

D=1   D=0

$Z = 0$

D=1   D=0

0.00 $p_C =$ 0.19 $p_I =$0.95 1.00
Always Takers

Fraction treated p

$Z = 1$

D=1   D=0

$Z = 0$

D=1   D=0

0.00 $p_C = $ 0.19 $p_I = $0.95   1.00

Always Takers   Never Takers

Fraction treated p

$Z = 1$ D=1 D=0

$Z = 0$ D=1 D=0

0.00 $p_c = 0.19$ Compliers $p_I = 0.95$ 1.00
Always Takers Never Takers
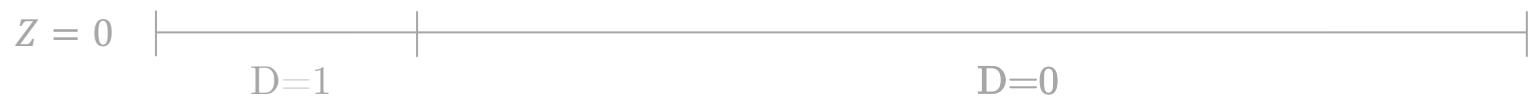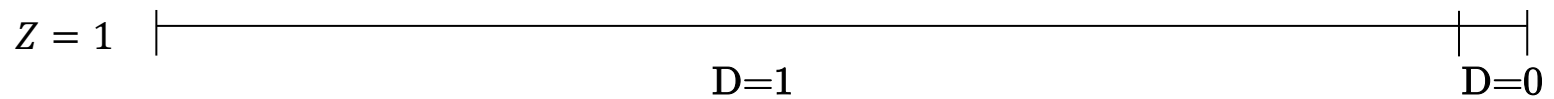
Fraction treated p

Behavior within a Clinical Trial and Implications for Mammography Guidelines

- **Model of Behavior within a Clinical Trial**
  - First Stage: Mammography
  - **Second Stage: Breast Cancer Incidence**
- Results
  1. Selection Heterogeneity
     - *Women more likely to receive mammograms are healthier*
  2. Treatment Effect Heterogeneity
     - *Women more likely to receive mammograms are more likely to be overdiagnosed by them*
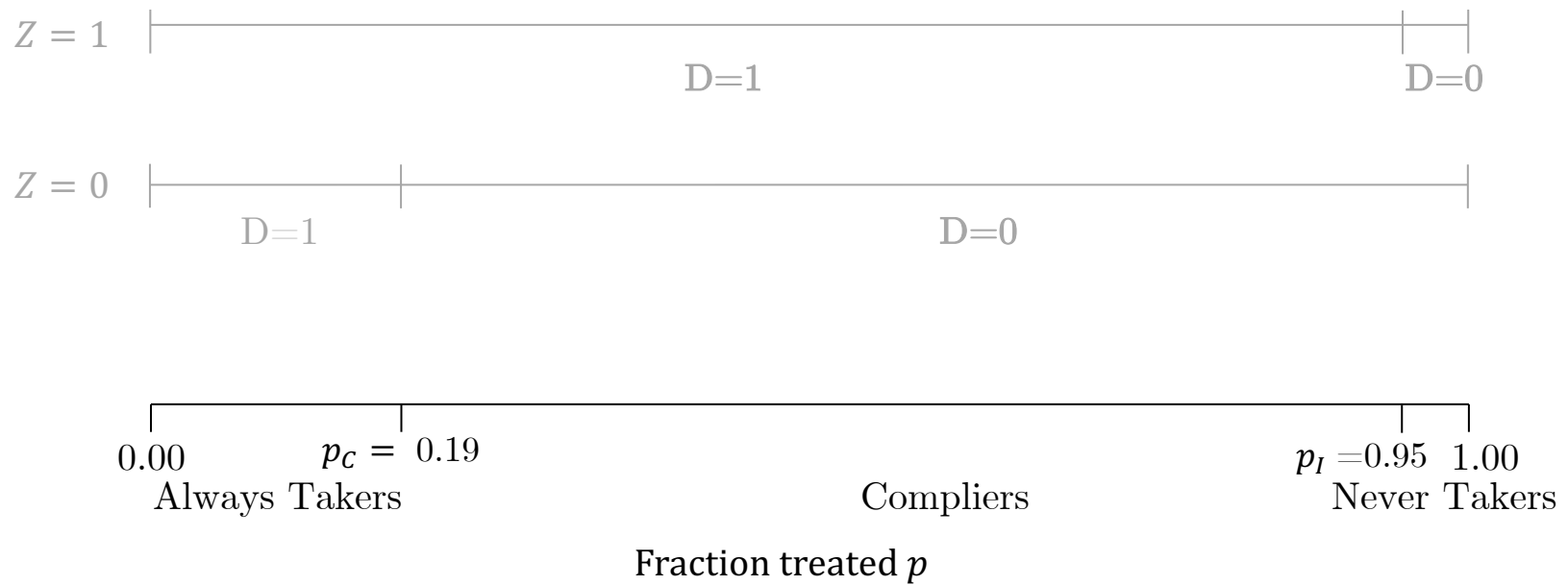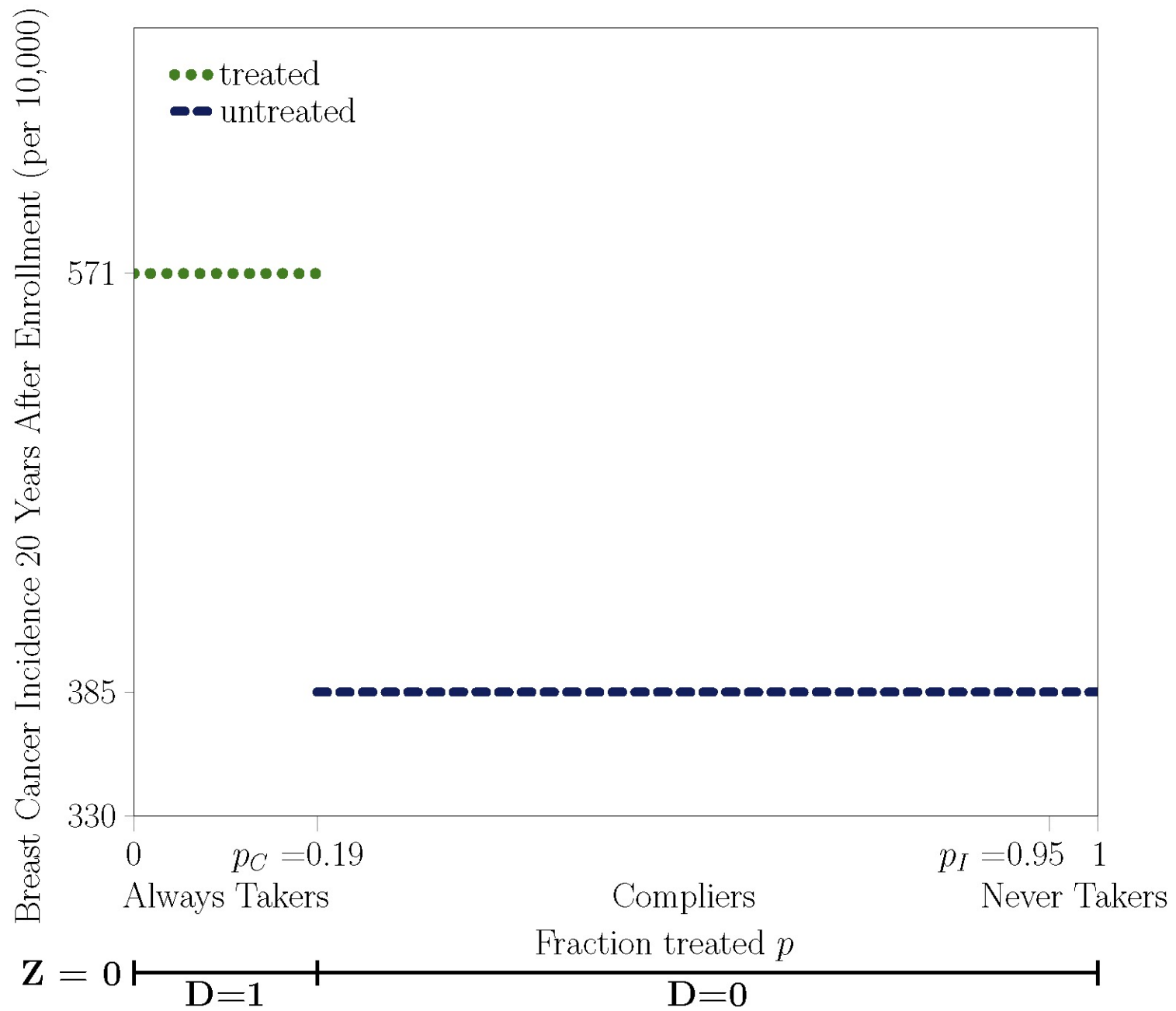- Robustness
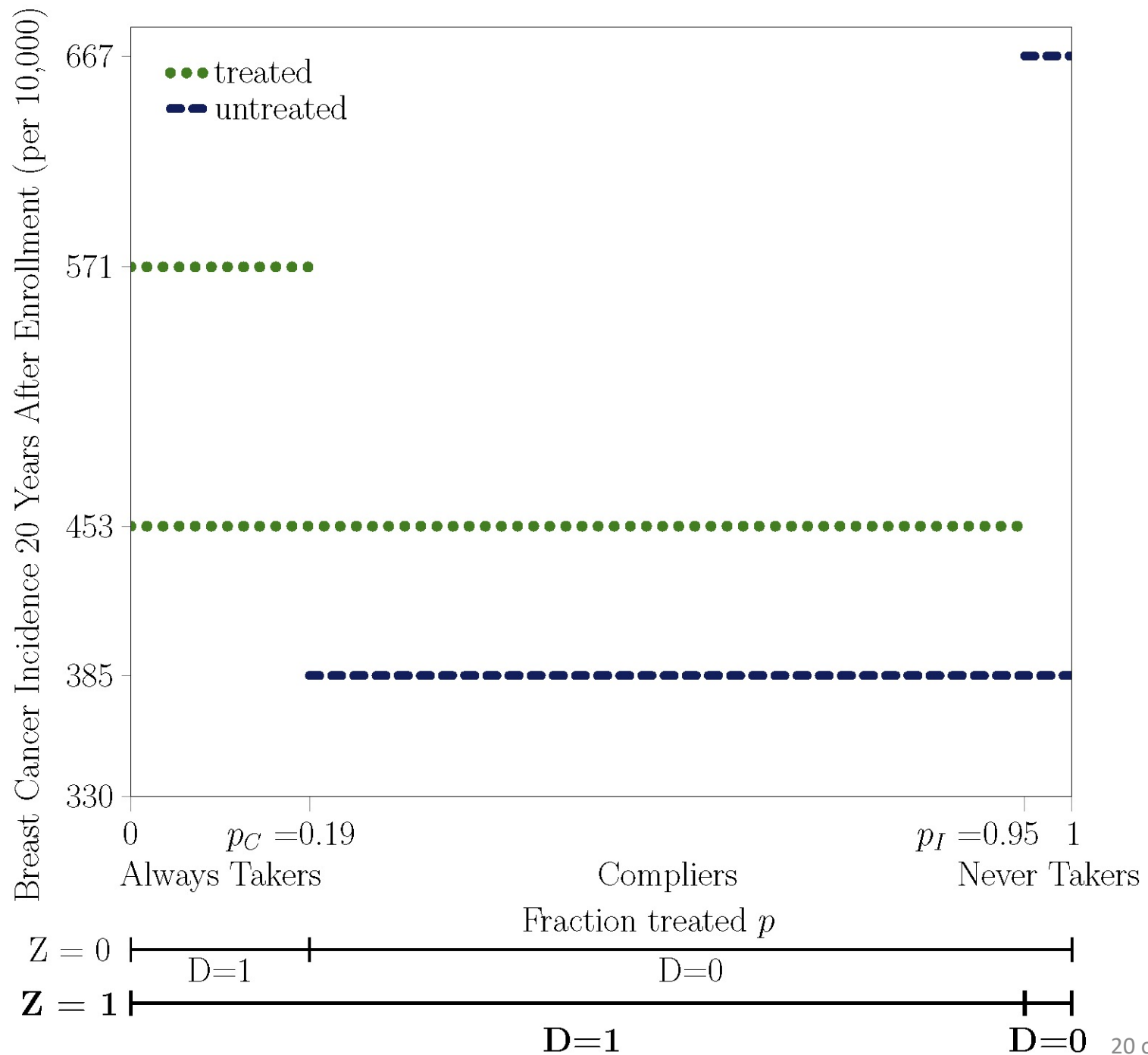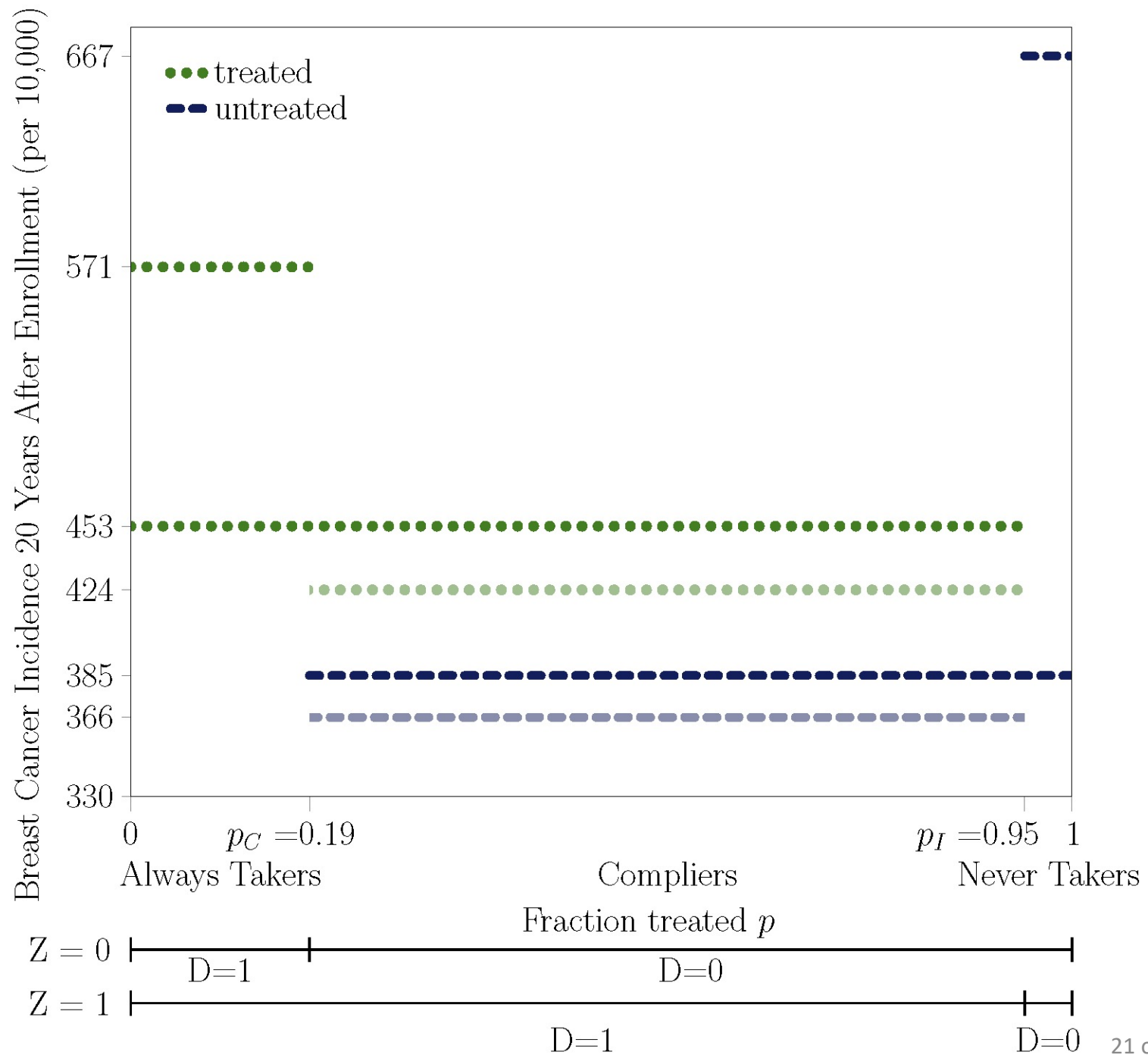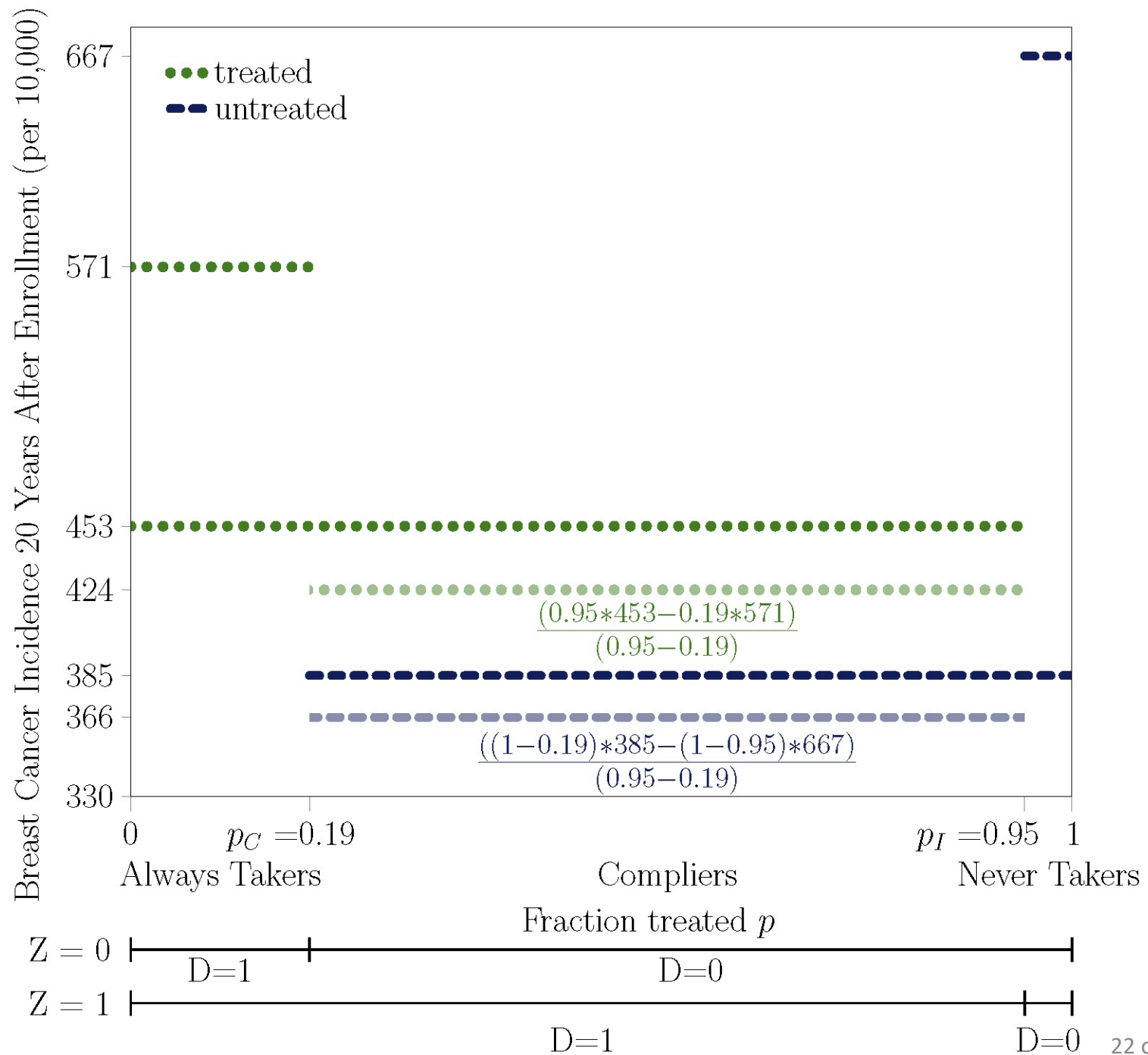- Implications for Mammography Guidelines

Behavior within a Clinical Trial
and Implications for Mammography Guidelines
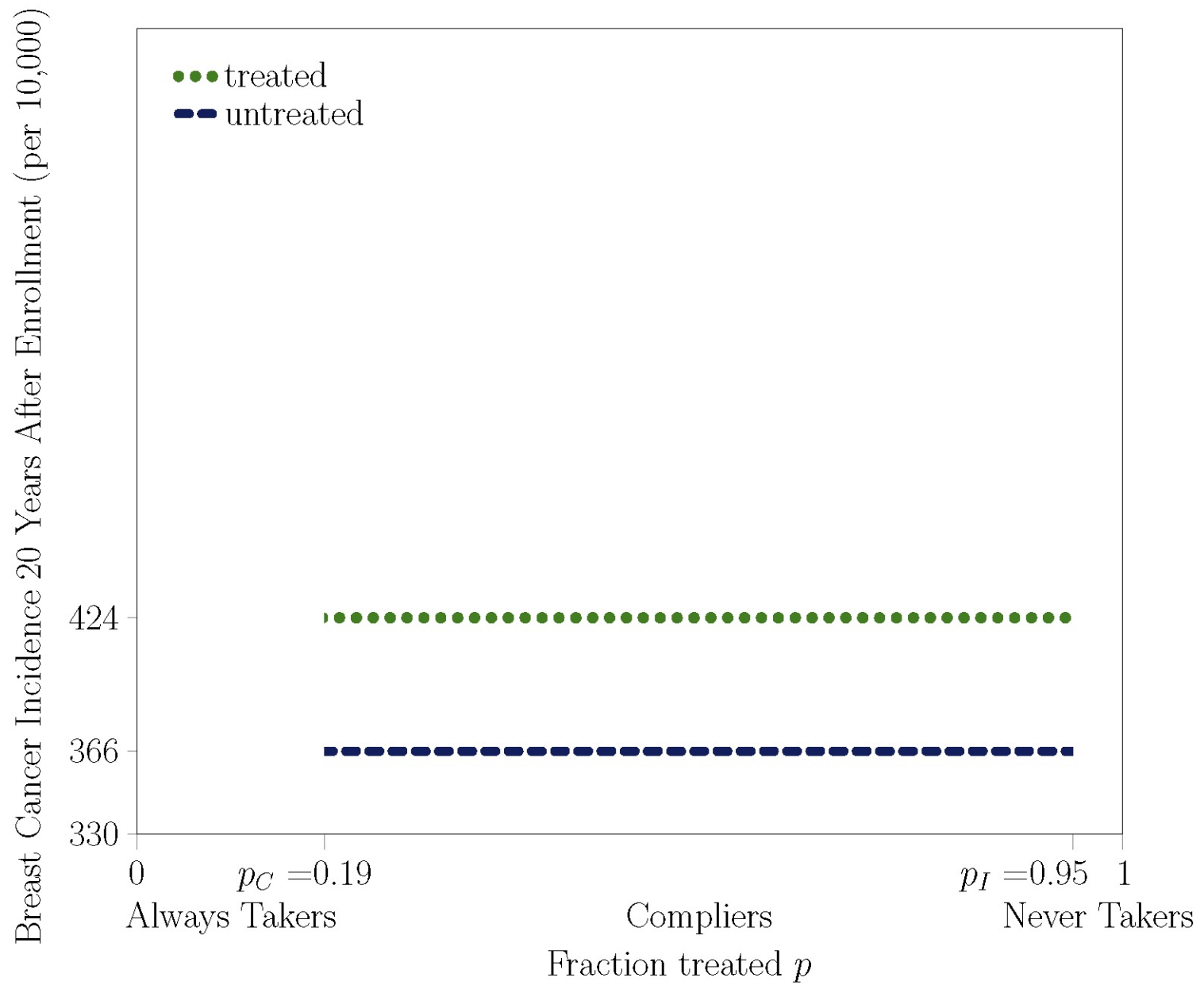
- Model of Behavior within a Clinical Trial
  - First Stage: Mammography
  - Second Stage: Breast Cancer Incidence
- **Results**
  1. **Selection Heterogeneity**
     - *Women more likely to receive mammograms are healthier*
  2. Treatment Effect Heterogeneity
     - *Women more likely to receive mammograms are more likely to be overdiagnosed by them*
- Robustness
- Implications for Mammography Guidelines
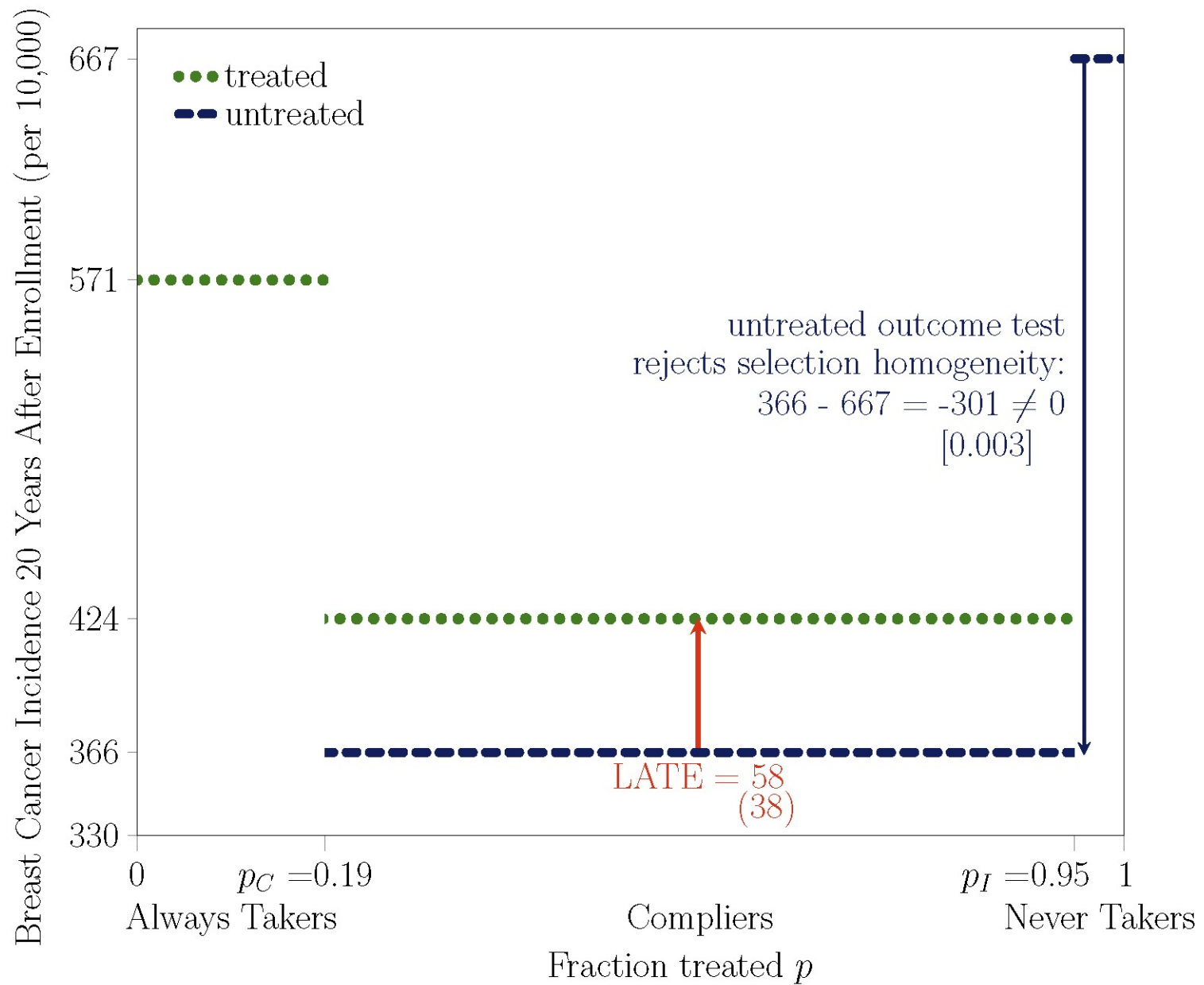
All-Cause Mortality
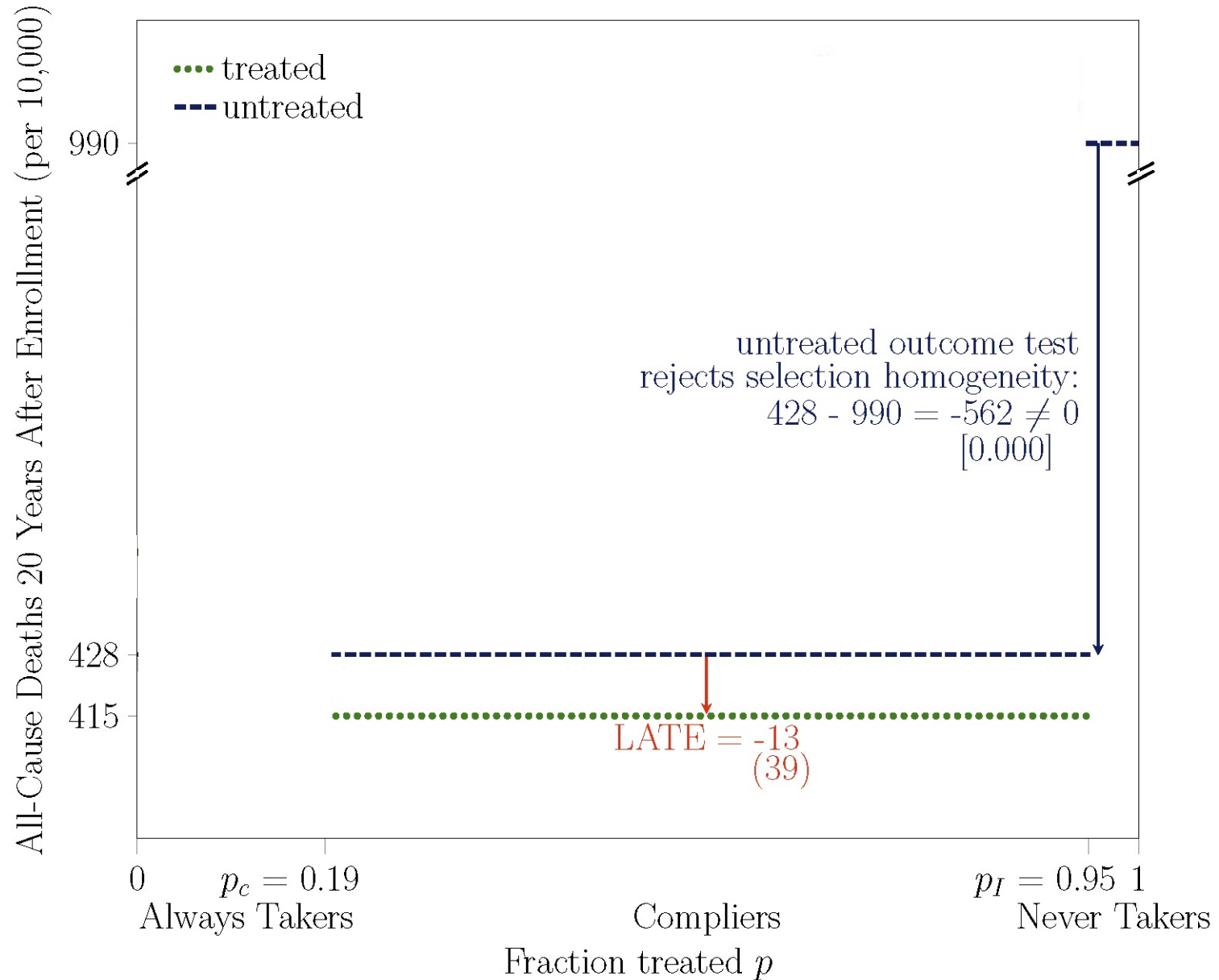
Women more likely to receive mammograms have higher socioeconomic status

	Means		
	Always Takers	Compliers	Never Takers
Baseline Socioeconomic Status			
In work force	0.65	0.64	0.65
Age at first birth	24.28	23.98	23.57
No live birth	0.16	0.15	0.13
Married	0.80	0.81	0.75
Husband in work force if alive	0.81	0.81	0.76
Baseline Health Behavior			
Non-Smoker	0.78	0.75	0.63
Body Mass Index	23.87	24.42	24.48
Used oral contraception	0.74	0.71	0.67
Used estrogen	0.13	0.13	0.15
Any mammograms prior to enrollment	0.23	0.13	0.13
Practiced breast self-examination	0.47	0.44	0.38

Natural Experiments Corroborate Selection Heterogeneity

- Kim and Lee (2017)

- Einav, Finkelstein, Oostrom, Ostriker, Williams (2020)
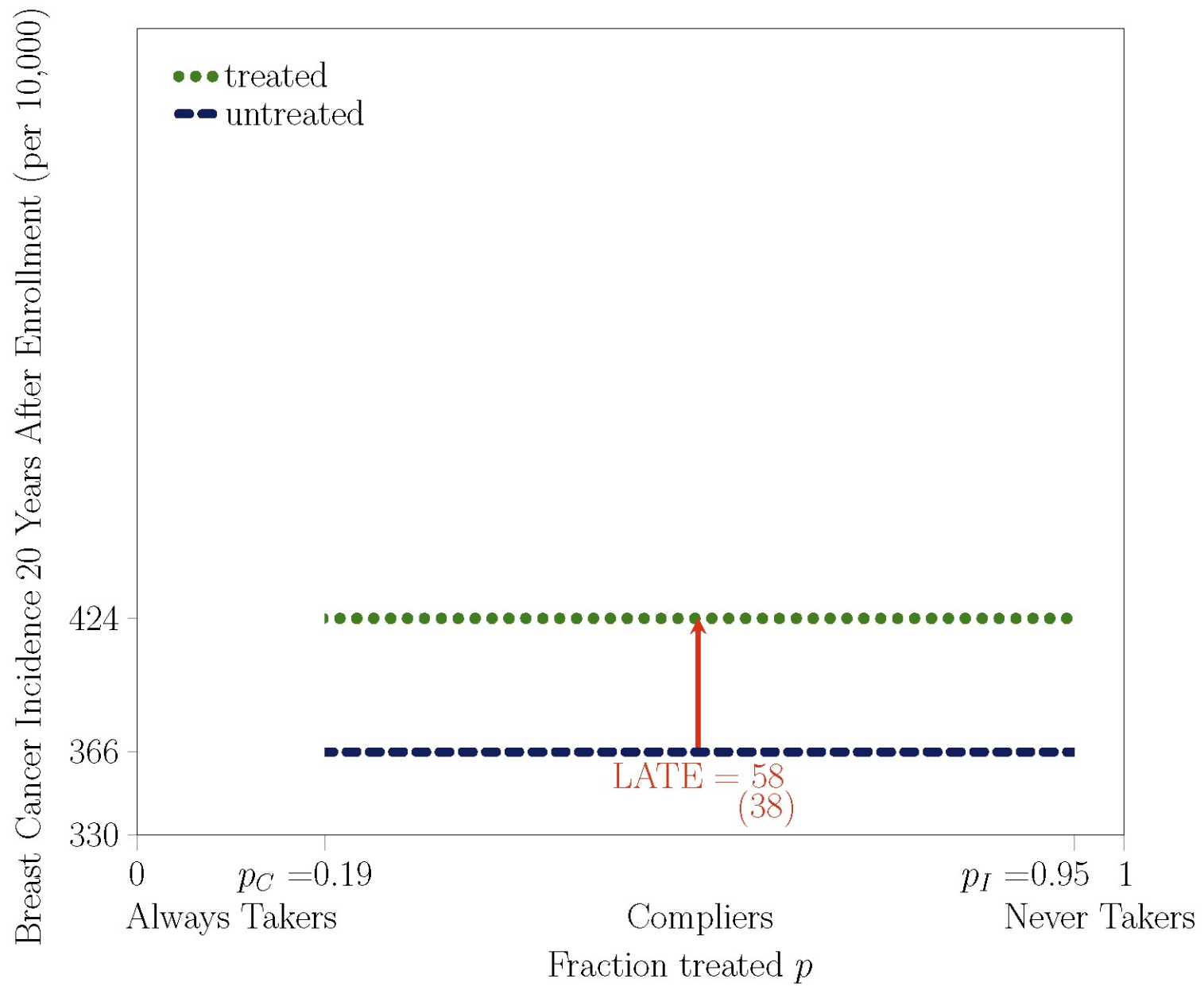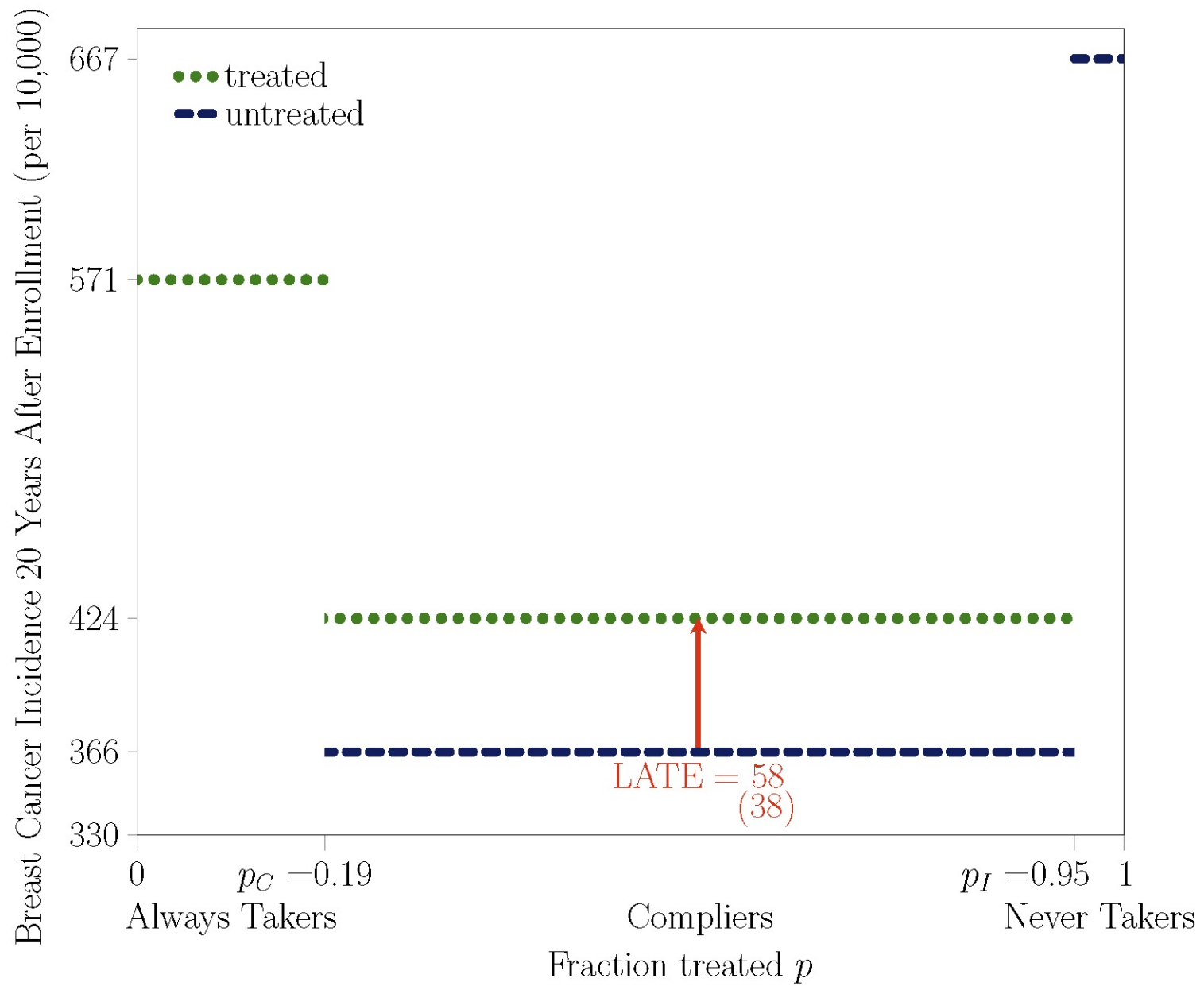
Behavior within a Clinical Trial and Implications for Mammography Guidelines

- Model of Behavior within a Clinical Trial
  - First Stage: Mammography
  - Second Stage: Breast Cancer Incidence
- **Results**
  1. Selection Heterogeneity
     - *Women more likely to receive mammograms are healthier*
  2. **Treatment Effect Heterogeneity**
     - *Women more likely to receive mammograms are more likely to be overdiagnosed by them*
- Robustness
- Implications for Mammography Guidelines
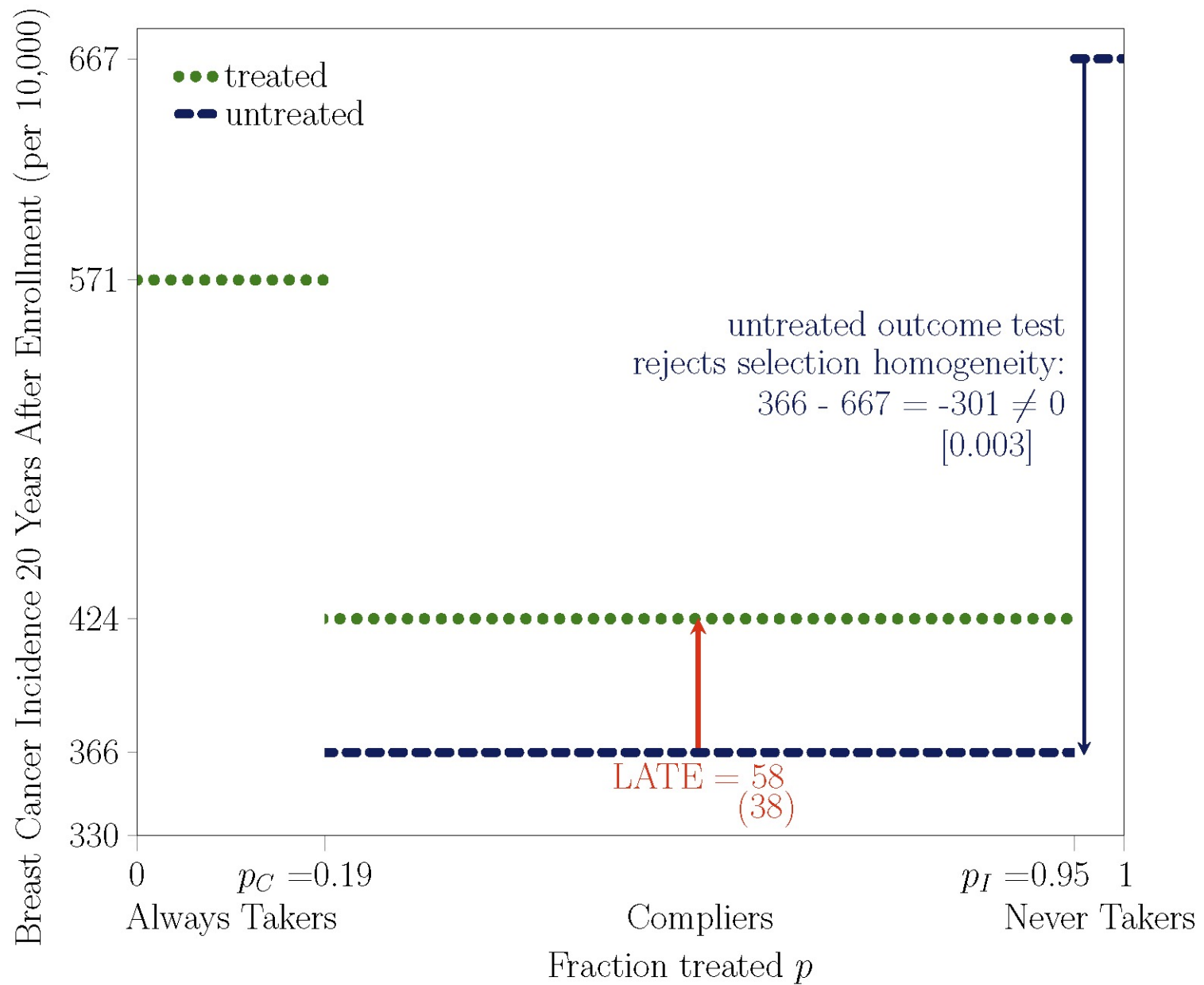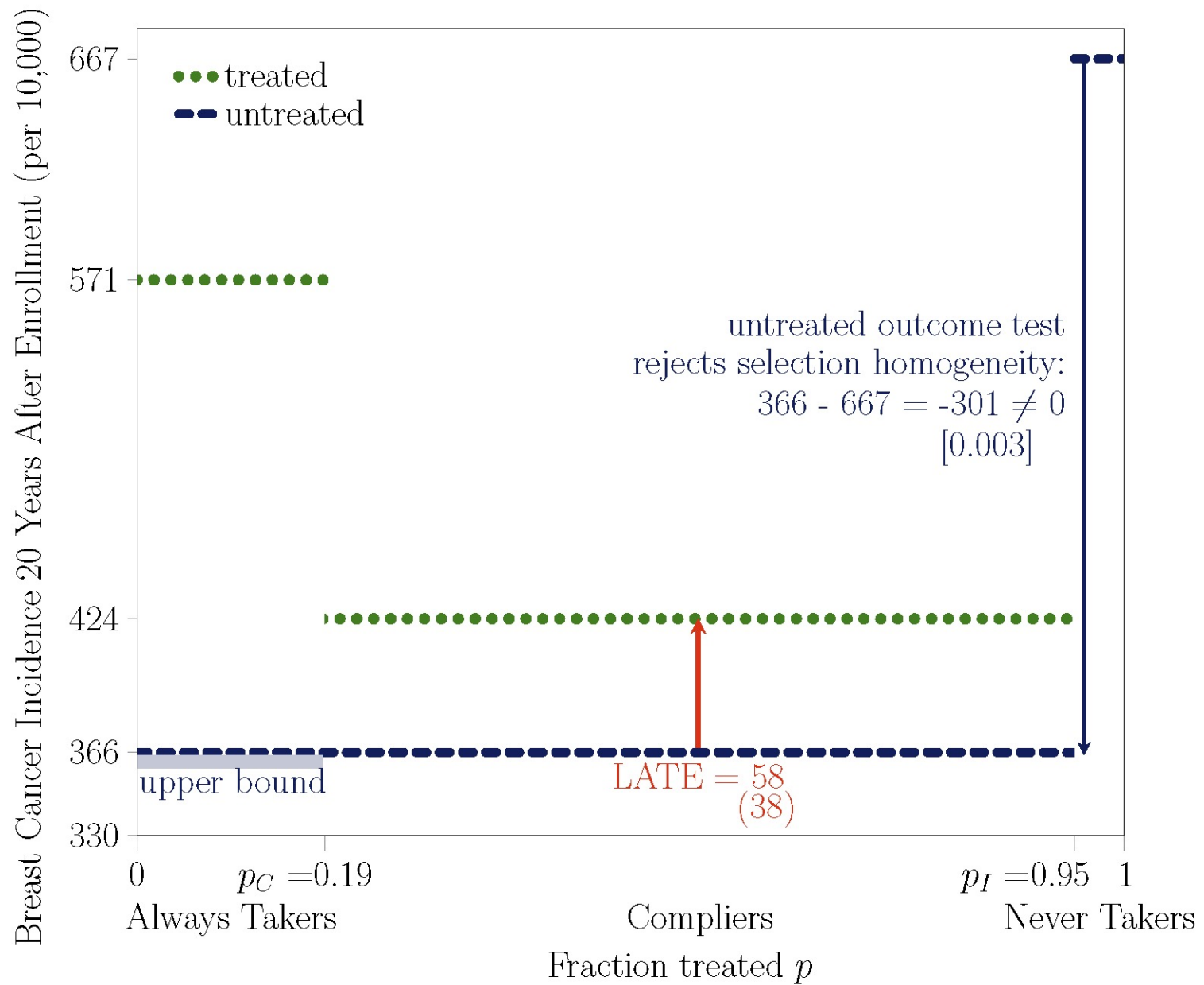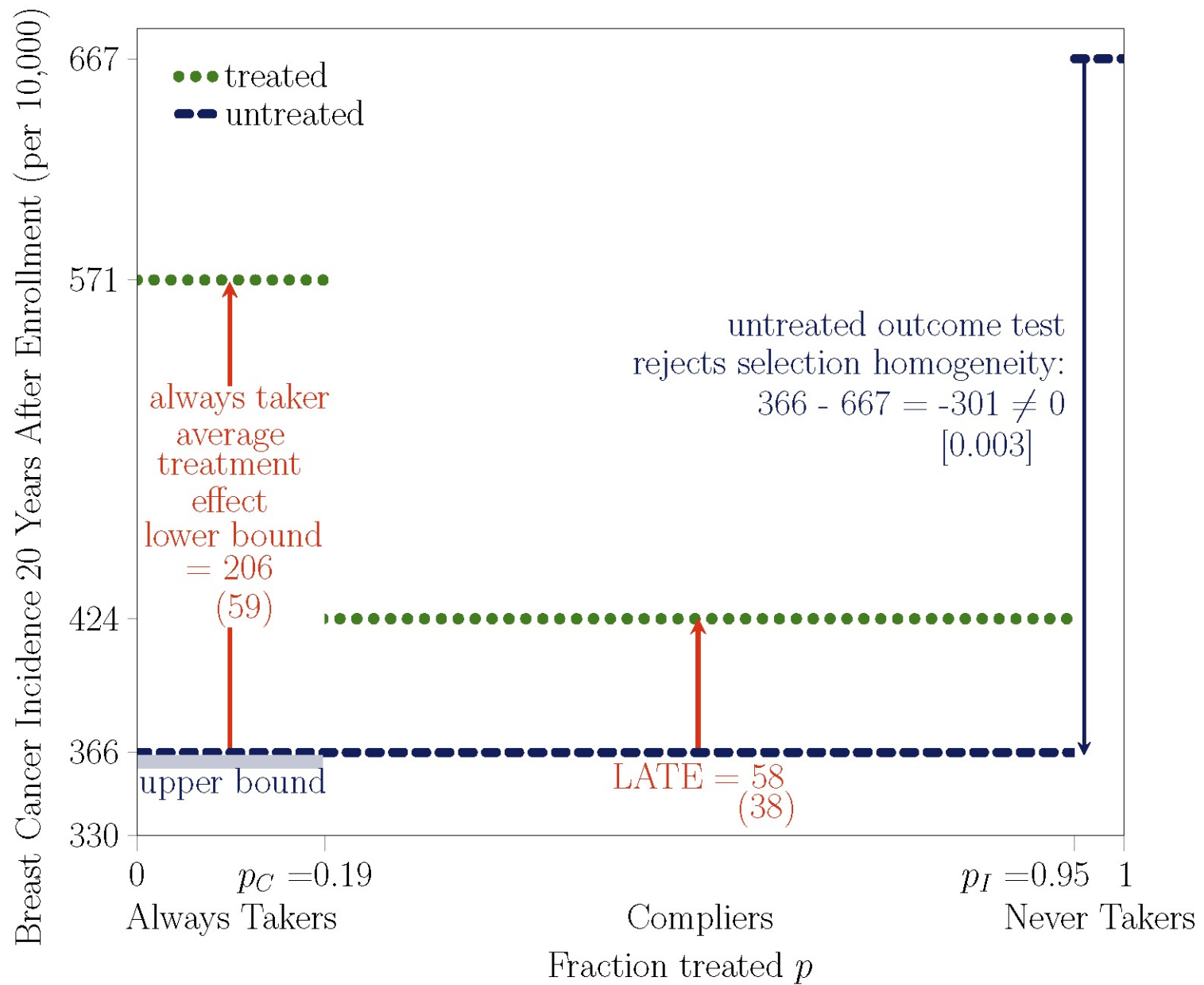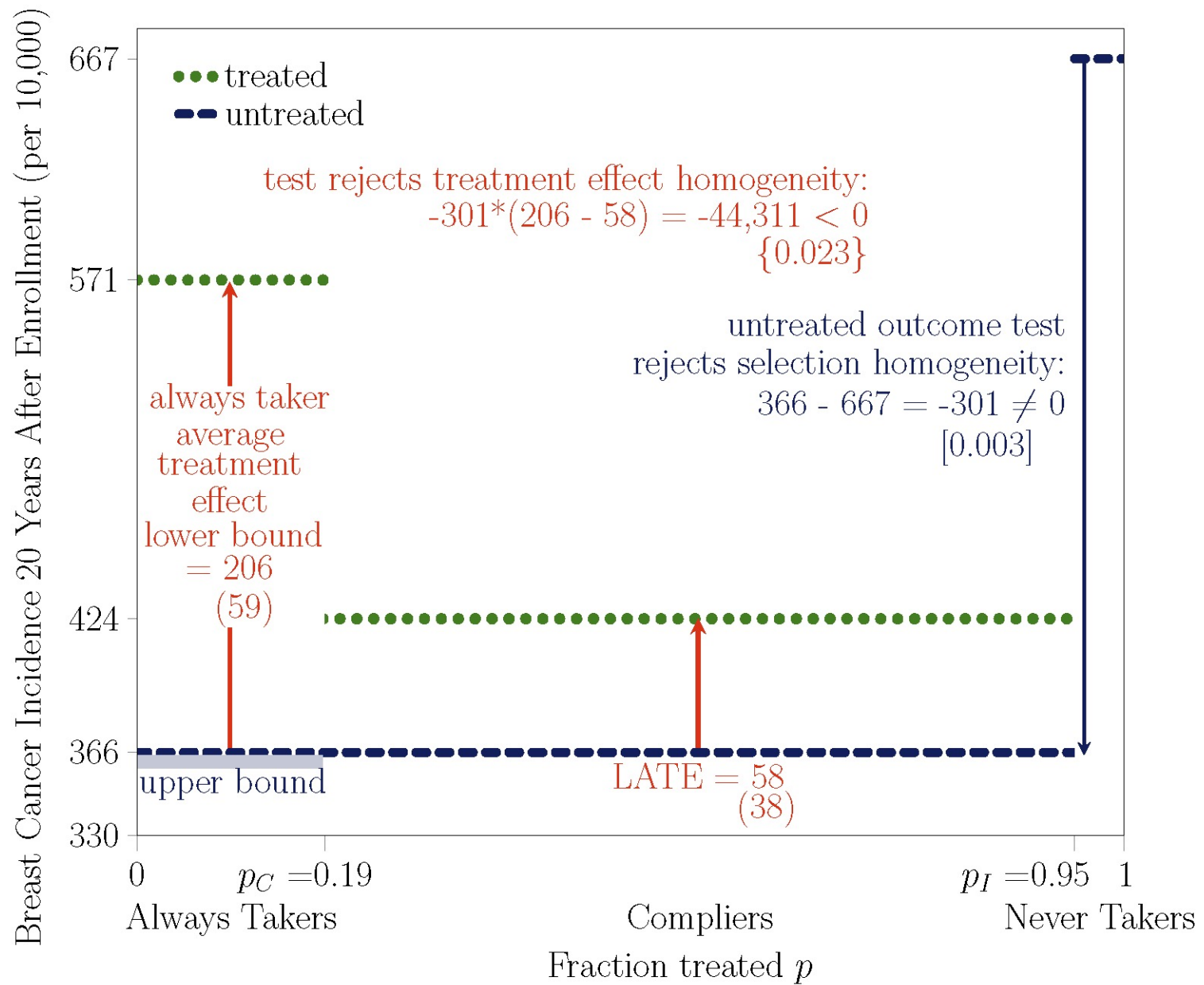
Women of higher socioeconomic status are exposed to increased *"observational intensity"* such that "*they are likely to be screened more often and by means of such tests...that can detect smaller abnormalities, undergo more follow-up testing, and undergo more biopsies, and they may be served by health systems that have a lower threshold for labeling results as abnormal.*"

- Welch and Fisher (2017)

Breast Cancer Characteristics Corroborate Treatment Effect Heterogeneity

	Means		Difference in Means
	(1)	(2)	(1) - (2)
	Always Takers	Treated Compliers	
Tumor Size Among Breast Cancers (in mm)	13	18	-5
	(2)	(3)	(4)
Share of Invasive Breast Cancer Among Breast Cancers (%)	73	75	-2
	(9)	(7)	(13)

Procedures Corroborate Treatment Effect Heterogeneity

|	Means		Difference in Means
	(1)	(2)	(1) - (2)
	Always Takers	Treated Compliers	
Tumor Size Among Breast Cancers (in mm)	13	18	-5
	(2)	(3)	(4)
Share of Invasive Breast Cancer Among Breast Cancers (%)	73	75	-2
	(9)	(7)	(13)
Share of Mastectomy Among Breast Cancers	45	23	22
with Mastectomy or Lumpectomy (%)	(9)	(7)	(14)

Behavior within a Clinical Trial
and Implications for Mammography Guidelines

- Model of Behavior within a Clinical Trial
  - First Stage: Mammography
  - Second Stage: Breast Cancer Incidence
- Results
  1. Selection Heterogeneity
     - *Women more likely to receive mammograms are healthier*
  2. Treatment Effect Heterogeneity
     - *Women more likely to receive mammograms are more likely to be overdiagnosed by them*
- **Robustness**
- Implications for Mammography Guidelines

Results Are Robust Along Many Dimensions

- Alternative outcome
  - All-cause mortality
- Alternative sample restrictions
  - Excluded participants aged 40-49
  - Aged 40-49 at enrollment
  - Aged 50-59 at enrollment
  - All participants
- Alternative definitions of mammography
  - Narrower
- Alternative follow-up lengths

Suggestive Evidence for All-Cause Mortality
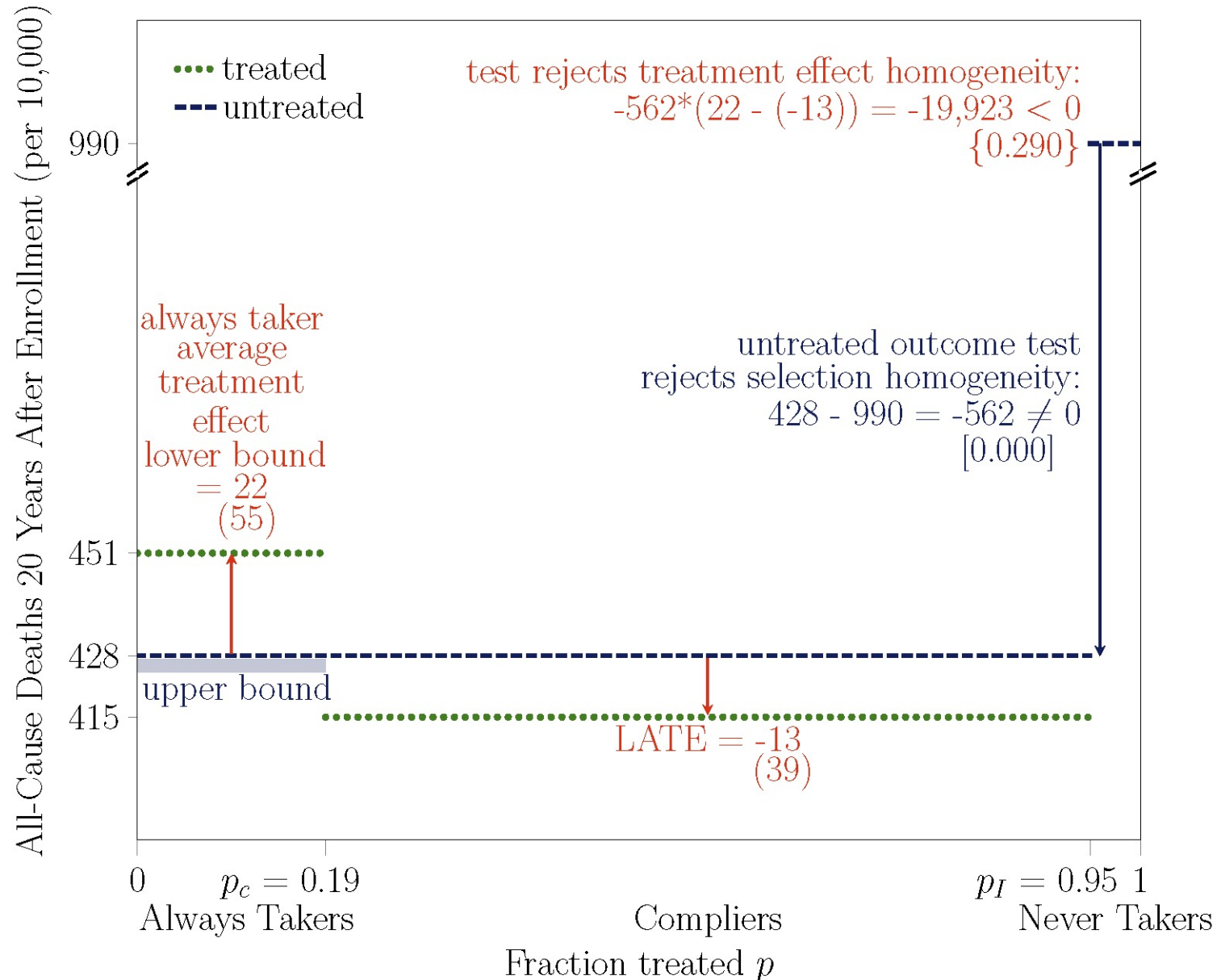
Behavior within a Clinical Trial and Implications for Mammography Guidelines

- Model of Behavior within a Clinical Trial
  - First Stage: Mammography
  - Second Stage: Breast Cancer Incidence
- Results
  1. Selection Heterogeneity
     - *Women more likely to receive mammograms are healthier*
  2. Treatment Effect Heterogeneity
     - *Women more likely to receive mammograms are more likely to be overdiagnosed by them*
- Robustness
- **Implications for Mammography Guidelines**

Implications for Mammography Guidelines

U.S. Preventive Services Task Force (USPSTF) 2016 Guidelines for Women in 40's:

"The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences"

Appendix

Results Are Robust Along Many Dimensions

- Alternative outcome
  - All-cause mortality
- Alternative sample restrictions
  - Excluded participants aged 40-49
  - Aged 40-49 at enrollment
  - Aged 50-59 at enrollment
  - All participants
- Alternative definitions of mammography
  - Narrower
- Alternative follow-up lengths

Main Specification For Comparison

	N	(1) Untreated Outcome Test	(2) Always Taker Average Treatment Effect Lower Bound	(3) Local Average Treatment Effect LATE	(4) Test Rejects Treatment Effect Homogeneity $(1)*((2)-(3))<0$
Main Specification					
Breast cancer incidence	19,505	-301 [0.003]	206 (59)	58 (38)	-44,311 {0.023}

Results Are Robust Along Many Dimensions

- **Alternative outcome**
  - **All-cause mortality**
- Alternative sample restrictions
  - Excluded participants aged 40-49
  - Aged 40-49 at enrollment
  - Aged 50-59 at enrollment
  - All participants
- Alternative definitions of mammography
  - Narrower
- Alternative follow-up lengths

Robust to Alternative Outcome

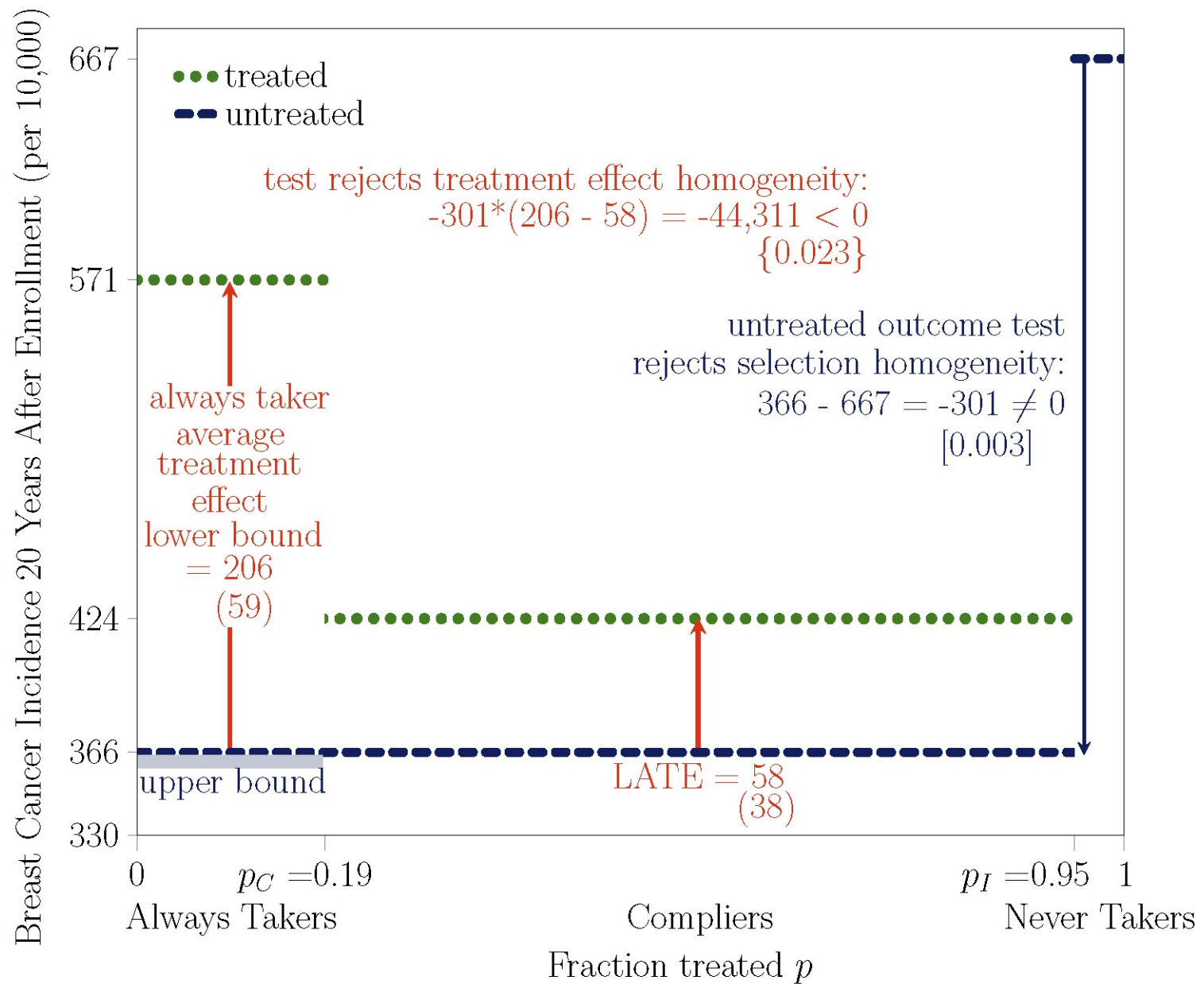|	N	(1) Untreated Outcome Test	(2) Always Taker Average Treatment Effect Lower Bound	(3) Local Average Treatment Effect LATE	(4) Test Rejects Treatment Effect Homogeneity (1)*((2)-(3))<0
Main Specification					
Breast cancer incidence	19,505	-301 [0.003]	206 (59)	58 (38)	-44,311 {0.023}
Alternative Outcomes					
All-cause mortality	19,505	-562 [0.000]	22 (55)	-13 (38)	-19,923 {0.290}

Results Are Robust Along Many Dimensions

- Alternative outcome
  - All-cause mortality
- **Alternative sample restrictions**
  - **Excluded participants aged 40-49**
  - **Aged 40-49 at enrollment**
  - **Aged 50-59 at enrollment**
  - **All participants**
- Alternative definitions of mammography
  - Narrower
- Alternative follow-up lengths

Robust to Alternative Sample Restrictions

	N	(1) Untreated Outcome Test	(2) Always Taker Average Treatment Effect Lower Bound	(3) Local Average Treatment Effect LATE	(4) Test Rejects Treatment Effect Homogeneity (1)*((2)-(3))<0
Main Specification					
Breast cancer incidence	19,505	-301 [0.003]	206 (59)	58 (38)	-44,311 {0.023}
Alternative Sample Restrictions					
All excluded participants aged 40-49 at enrollment	30,925	-1,237 [0.000]	309 (45)	79 (44)	-284,634 {0.000}
All participants aged 40-49 at enrollment	50,430	-826 [0.000]	298 (36)	69 (30)	-189,397 {0.000}
All participants aged 50-59 at enrollment	39,405	-1,555 [0.000]	419 (53)	39 (34)	-591,037 {0.000}
All participants	89,835	-1,156 [0.000]	332 (30)	55 (22)	-319,660 {0.000}

Results Are Robust Along Many Dimensions

- Alternative outcome
  - All-cause mortality
- Alternative sample restrictions
  - Excluded participants aged 40-49
  - Aged 40-49 at enrollment
  - Aged 50-59 at enrollment
  - All participants
- **Alternative definitions of mammography**
  - **Narrower**
- Alternative follow-up lengths

Robust to Alternative Definitions of Mammography

	N	(1) Untreated Outcome Test	(2) Always Taker Average Treatment Effect Lower Bound	(3) Local Average Treatment Effect LATE	(4) Test Rejects Treatment Effect Homogeneity (1)*((2)-(3))<0
Main Specification					
Breast cancer incidence	19,505	-301 [0.003]	206 (59)	58 (38)	-44,311 {0.023}
Alternative Definitions of Mammography					
At least two active study period years after enrollment	19,505	-341 [0.000]	239 (90)	54 (35)	-63,347 {0.019}
At least three active study period years after enrollment	19,505	-330 [0.000]	167 (142)	55 (36)	-36,927 {0.206}
All active study period years after enrollment	19,505	-178 [0.005]	158 (181)	64 (42)	-16,656 {0.312}

Results Are Robust Along Many Dimensions

- Alternative outcome
  - All-cause mortality
- Alternative sample restrictions
  - Excluded participants aged 40-49
  - Aged 40-49 at enrollment
  - Aged 50-59 at enrollment
  - All participants
- Alternative definitions of mammography
  - Narrower
- **Alternative follow-up lengths**

Robust to Breast Cancer Incidence at Alternative Follow-Up Lengths: 11-20

Years Since Enrollment	N	(1) Untreated Outcome Test Rejects Selection Homogeneity	(2) Always Taker Average Treatment Effect Lower Bound	(3) Local Average Treatment Effect LATE	(4) Test Rejects Treatment Effect Homogeneity (1)*((2)-(3))<0
Main specification: 20	19,505	-301 [0.003]	206 (59)	58 (38)	-44,311 {0.023}
19	19,505	-269 [0.013]	196 (58)	52 (37)	-38,565 {0.023}
18	19,505	-311 [0.000]	210 (56)	54 (35)	-48,503 {0.010}
17	19,505	-322 [0.000]	214 (55)	49 (34)	-52,975 {0.005}
16	19,505	-342 [0.000]	232 (54)	56 (32)	-60,245 {0.003}
15	19,505	-381 [0.000]	211 (50)	84 (31)	-48,650 {0.015}
14	19,505	-404 [0.000]	201 (49)	80 (29)	-49,046 {0.020}
13	19,505	-431 [0.000]	223 (48)	75 (28)	-63,808 {0.007}
12	19,505	-443 [0.000]	191 (44)	64 (27)	-56,156 {0.010}
11	19,505	-423 [0.000]	195 (43)	55 (25)	-59,084 {0.004}

Robust to Breast Cancer Incidence at Alternative Follow-Up Lengths: 1-10

Years Since Enrollment	N	(1) Untreated Outcome Test Rejects Selection Homogeneity	(2) Always Taker Average Treatment Effect Lower Bound	(3) Local Average Treatment Effect LATE	(4) Test Rejects Treatment Effect Homogeneity (1)*((2)-(3))<0
10	19,505	-419 [0.000]	200 (42)	47 (23)	-64,017 {0.000}
9	19,505	-413 [0.000]	192 (40)	34 (22)	-64,955 {0.000}
8	19,505	-409 [0.000]	175 (37)	35 (21)	-57,386 {0.000}
7	19,505	-393 [0.000]	177 (35)	46 (18)	-51,740 {0.000}
6	19,505	-412 [0.000]	185 (33)	50 (17)	-55,761 {0.000}
5	19,505	-382 [0.000]	180 (32)	45 (15)	-51,581 {0.000}
4	19,505	-393 [0.000]	152 (29)	46 (13)	-41,568 {0.003}
3	19,505	-354 [0.000]	104 (23)	37 (11)	-23,679 {0.012}
2	19,505	-337 [0.000]	63 (18)	25 (9)	-12,632 {0.030}
1	19,505	-342 [0.000]	35 (11)	20 (6)	-5,194 {0.097}

2016 USPSTF Guidelines Based on RCT's

Annals of Internal Medicine — REVIEW

Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation

Heidi D. Nelson, MD, MPH; Rochelle Fu, PhD; Amy Cantor, MD, MPH; Miranda Pappas, MA; Monica Daeges, BA; and Linda Humphrey, MD, MPH

Author, Year (Reference)	Trial Name	Mean Follow-up, y	Relative Risk (95% CI)
Women aged 39–49 y			
Nyström et al, 2002 (30)*	MMST II	11.2	0.64 (0.39–1.06)
Tabár et al, 1995 (26)	Kopparberg	12.5	0.73 (0.37–1.41)
Tabár et al, 1995 (26)	Östergötland	12.5	1.02 (0.52–1.99)
Moss et al, 2015 (27)	Age	17.5	0.93 (0.80–1.09)
Bjurstam et al, 2003 (25)	Gothenburg	13.8	0.69 (0.45–1.05)
Habbema et al, 1986 (29)	HIP	14.0	0.75 (0.53–1.05)
Nyström et al, 2002 (30)*	Stockholm	14.3	1.52 (0.80–2.88)
Nyström et al, 2002 (30)*	MMST I	18.2	0.74 (0.42–1.29)
Miller et al, 2014 (15)	CNBSS-1	21.9	1.04 (0.87–1.24)
Overall ($I^2 = 25\%$; $P = 0.230$)			0.92 (0.75–1.02)

0.25   1.00   4.00

Favors Screening Group — Relative Risk (95% CI) — Favors Control Group

CNBSS Consistent with Meta-analysis of RCT's

Annals of Internal Medicine

REVIEW

Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation

Heidi D. Nelson, MD, MPH; Rochelle Fu, PhD; Amy Cantor, MD, MPH; Miranda Pappas, MA; Monica Daeges, BA; and Linda Humphrey, MD, MPH

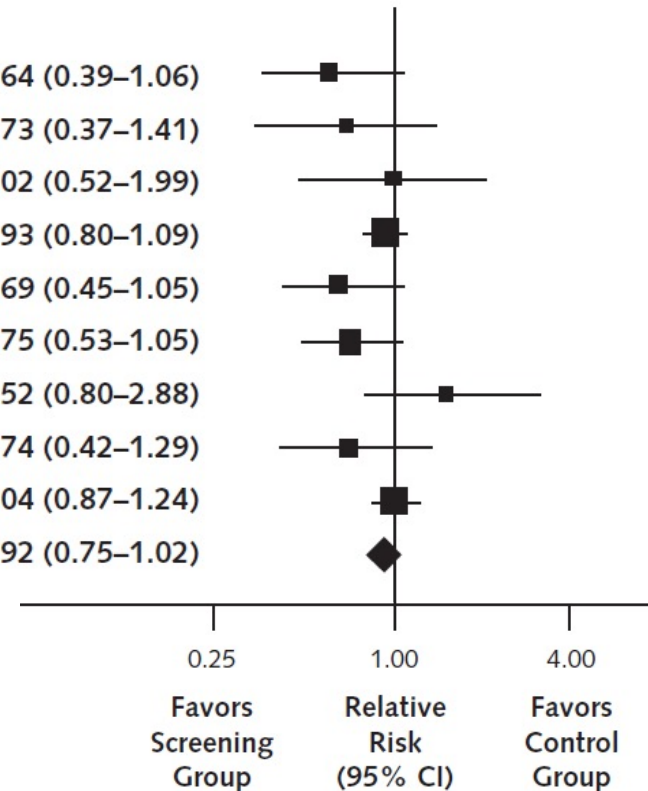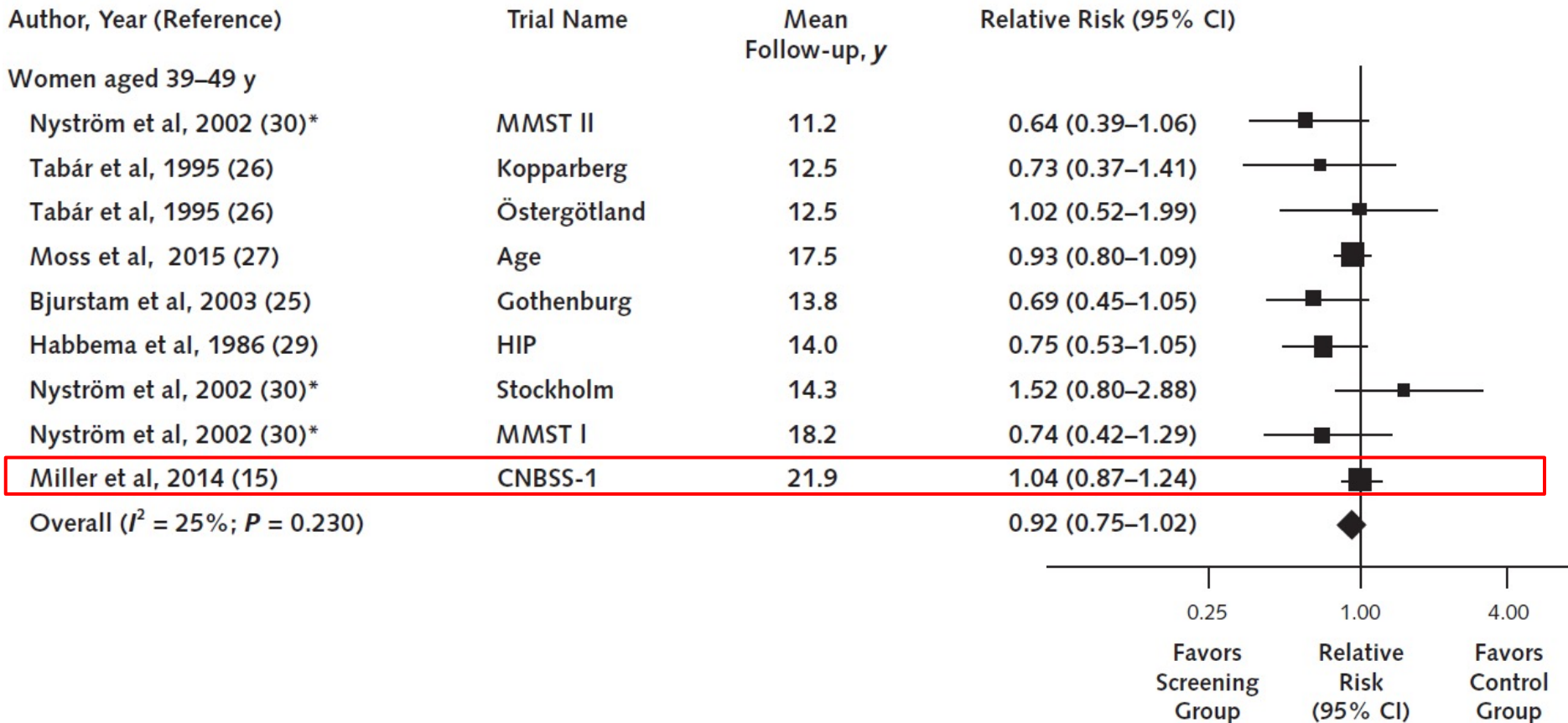| Author, Year (Reference)	Trial Name	Mean Follow-up, y	Relative Risk (95% CI)
Women aged 39–49 y			
Nyström et al, 2002 (30)*	MMST II	11.2	0.64 (0.39–1.06)
Tabár et al, 1995 (26)	Kopparberg	12.5	0.73 (0.37–1.41)
Tabár et al, 1995 (26)	Östergötland	12.5	1.02 (0.52–1.99)
Moss et al, 2015 (27)	Age	17.5	0.93 (0.80–1.09)
Bjurstam et al, 2003 (25)	Gothenburg	13.8	0.69 (0.45–1.05)
Habbema et al, 1986 (29)	HIP	14.0	0.75 (0.53–1.05)
Nyström et al, 2002 (30)*	Stockholm	14.3	1.52 (0.80–2.88)
Nyström et al, 2002 (30)*	MMST I	18.2	0.74 (0.42–1.29)
Miller et al, 2014 (15)	CNBSS-1	21.9	1.04 (0.87–1.24)
Overall ($I^2 = 25\%$; $P = 0.230$)			0.92 (0.75–1.02)

0.25   1.00   4.00

Favors Screening Group   Relative Risk (95% CI)   Favors Control Group
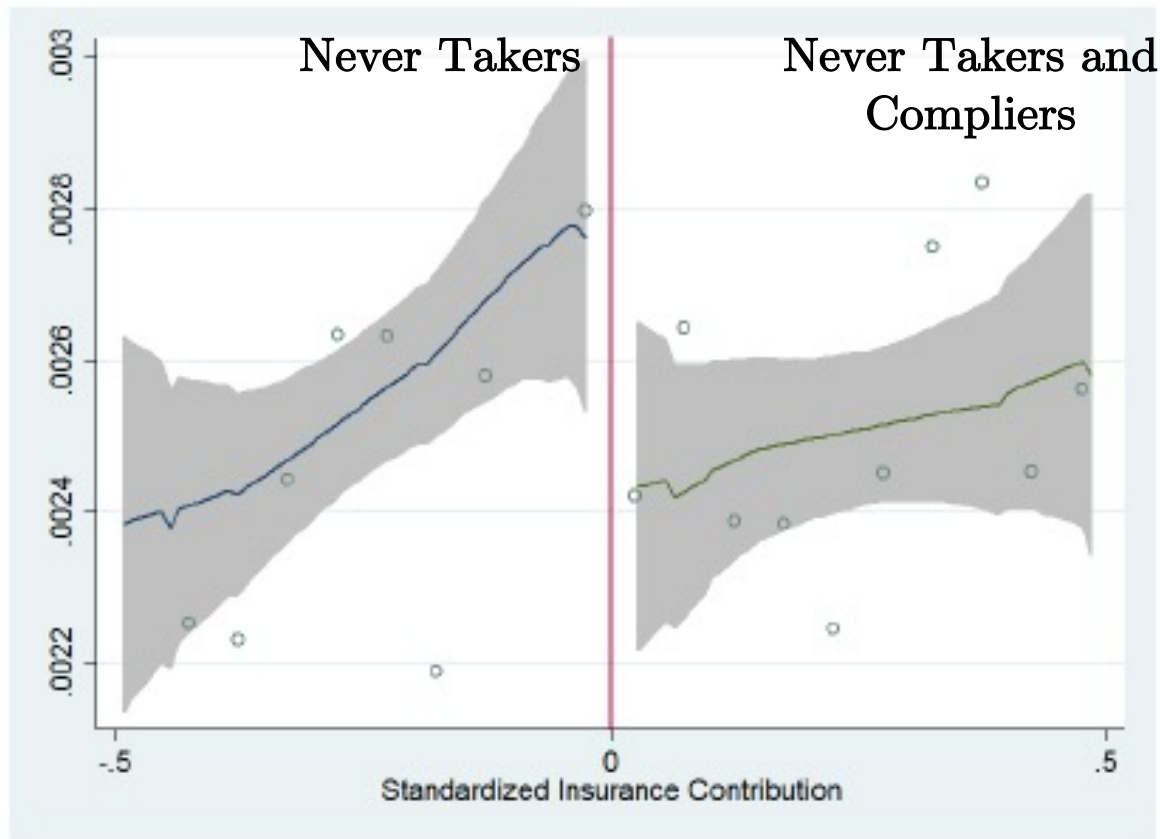
CNBSS Protocols Varied by Age

- Patients aged 40-49:
  - Intervention group: mammography + physical examination each year for 4-5 years, then return to usual care
  - Control group: usual care
- Patients aged 50-59:
  - Intervention group: mammography + physical examination each year for 4-5 years, then return to usual care
  - Control group: physical examination each year for 4-5 years, then return to usual care

USPSTF Recommendations Differ for Women in 40's and 50's

- The U.S. Preventive Services Task Force (USPSTF) Assigns "grades"
  - "A" and "B" grades fully-covered under ACA
- Different grades for 40's and 50+ (Siu, 2016)
  - "The decision to start screening mammography in women prior to age 50 years should be an individual one. (Grade C recommendation)"
  - "The USPSTF recommends biennial screening mammography for women aged 50 to 74 years. (Grade B recommendation)"

Never Takers Die More Than Compliers

- Women more likely to receive mammograms are healthier
- Breast cancer mortality <u>without</u> screening (Kim and Lee, 2017)

Female, Breast cancer