BASICALLY BILL'S VARIOUS ERRORS AND HOW TO DEAL WITH THEM

Advanced Data Analysis, Apr 23, 2019, Bill Yuanchen Liu

Breast Cancer Survival based on diagnosis reporting types, a survival analysis

Contents

- Background
- Research question
- Methods
  - Data cleaning
  - Descriptives
  - Analysis plan
- Results
  - Table 1
  - Univariate
  - Multivariate
  - Survival
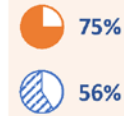- Discussion

Background

- On my DC Policy trip I advocated for Planned Parenthood.

- Found that in some counties, PP are the only source of women's wellness checks.

- There is a major difference between quality of care offered at FQHCs and at PP.

Protect Our Health?
S. 141 Strips Funding and Causes Harm

Planned Parenthood Federation of America is the leading provider of reproductive and sexual health services.[1]
In 2017, **2.4 million** people relied on Planned Parenthood for services at more than 600 affiliate centers.[2]

75% Low-income patients (Below 150% FPL)[1]

56% Centers in rural or underserved area[3]

4.7 Million Sexually Transmitted Infection testing[1]

2.8 Million Contraception services[2]

Other services provided:
- Cancer screening
- Well woman exams
- Pregnancy test
- Prenatal services
And much more...[2,4]

S. 141 Protect Funding for Women's Health Care Act will prohibit federal funding of Planned Parenthood. Although S. 141 *claims* to ensure that women will continue to receive necessary health care services, **defunding Planned Parenthood will deprive millions of individuals of essential health care.**

Drastic health impacts:
Texas removed Planned Parenthood from its state family planning program, resulting in a **35% decline** in contraceptive access and a dramatic **27% increase** in unintended births.[5]

Planned Parenthood in Missouri
17 of Missouri's **34** Senate districts would have **zero** providers that deliver accessible and comprehensive care without Planned Parenthood.[7]

$$ Heavy cost of defunding:
Prohibiting federal funding of Planned Parenthood would result in **a net cost of $130 million** to taxpayers over the next 10 years, due to the expected increase in unplanned pregnancies and birth.[6]

Without Planned Parenthood, FQHC sites would need to increase their contraceptive client caseloads:

| Less than double | At least double | At least triple |

⚠ Not enough capacity:
"As many as **1 in 4** people would face reduced access to care, even with available funding going to other clinics."[6]

Source: Guttmacher Institute[8]
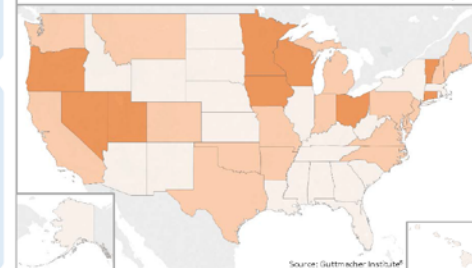
S. 141 will deny MILLIONS of individuals of essential sexual and reproductive care, which is both **HARMFUL** and **COSTLY** to the American people.
Do not support Senate Bill 141!

Question

- The objective of this study is to determine whether the different types of cancer diagnosis reporting source are associated with a difference in breast cancer survival in the U.S.

Methods:

Data cleaning

- SEER*Stat 18
  - Case listing session
  - Diagnosis and survival data, individual level
- Variables
  - Survival, specific death, reporting types
  - Controls: diagnosis age, sex, race, insurance, stage
- Preparing for MICE
  - Missing data
- Preparing for coxph
  - Na.omit
  - Complete case data analysis

Methods:

Descriptives

- Package called "TableOne" is very handy

Methods:

Analysis plan

- Univariate coxph
  - Survival time, death event, reporting
- Multivariate coxph
  - Age adjusted
- Multivariate coxph
  - Control variables
- Survival curves

```
                                         Stratified by reporting_omit
                                         Hospital        Laboratory      Nursing home    Surgery center  Physicians office Oncology center p       test
n                                        230949          2746            153             6680            2528              2871
sex = Male (%)                             1771 ( 0.8)     22 ( 0.8)       0 ( 0.0)       73 ( 1.1)        24 ( 0.9)         18 ( 0.6)      0.035
race (%)                                                                                                                                  <0.001
    Black                                 25860 (11.2)    208 ( 7.6)      15 ( 9.8)      749 (11.2)      126 ( 5.0)        249 ( 8.7)
    Other (American Indian/AK Native, Asian/Pacific Islander)  19198 ( 8.3)    544 (19.8)       6 ( 3.9)     1281 (19.2)      142 ( 5.6)        353 (12.3)
    Unknown                                 988 ( 0.4)    278 (10.1)       1 ( 0.7)       44 ( 0.7)      229 ( 9.1)         35 ( 1.2)
    White                               184903 (80.1)   1716 (62.5)     131 (85.6)     4606 (69.0)     2031 (80.3)       2234 (77.8)
diagnosis_age (mean (SD))                 61.71 (13.73) 65.04 (14.71)  79.38 (13.85)  62.10 (13.67)   67.59 (16.09)     62.23 (12.75)    <0.001
insurance_cleaned (%)                                                                                                                    <0.001
    Insured                             165063 (71.5)    850 (31.0)      15 ( 9.8)     4780 (71.6)      826 (32.7)       1975 (68.8)
    Insured/No specifics                 30745 (13.3)    231 ( 8.4)      24 (15.7)     1055 (15.8)      409 (16.2)        503 (17.5)
    Any Medicaid                         27291 (11.8)    115 ( 4.2)      17 (11.1)      523 ( 7.8)       83 ( 3.3)        225 ( 7.8)
    Uninsured                             4407 ( 1.9)     33 ( 1.2)       0 ( 0.0)      140 ( 2.1)       35 ( 1.4)         35 ( 1.2)
    Unknown                               3443 ( 1.5)   1517 (55.2)      97 (63.4)      182 ( 2.7)     1175 (46.5)        133 ( 4.6)
stage_cleaned (%)                                                                                                                        <0.001
    Stage 1                             112927 (48.9)    944 (34.4)       4 ( 2.6)     3670 (54.9)      793 (31.4)       1516 (52.8)
    Stage 2                              72259 (31.3)    536 (19.5)       2 ( 1.3)     1985 (29.7)      400 (15.8)        757 (26.4)
    Stage 3                              25462 (11.0)    161 ( 5.9)       2 ( 1.3)      548 ( 8.2)      143 ( 5.7)        322 (11.2)
    Stage 4                              13282 ( 5.8)    119 ( 4.3)      18 (11.8)      235 ( 3.5)      119 ( 4.7)        201 ( 7.0)
    Stage Unknown                         7019 ( 3.0)    986 (35.9)     127 (83.0)      242 ( 3.6)     1073 (42.4)         75 ( 2.6)
survival (mean (SD))                      51.99 (19.84) 43.28 (22.33)  14.18 (15.70)  51.97 (18.92)   40.97 (25.81)     53.61 (19.96)    <0.001
died (mean (SD))                           0.08 (0.28)   0.11 (0.31)    0.82 (0.39)    0.06 (0.23)     0.21 (0.41)       0.08 (0.27)    <0.001
```

Results:
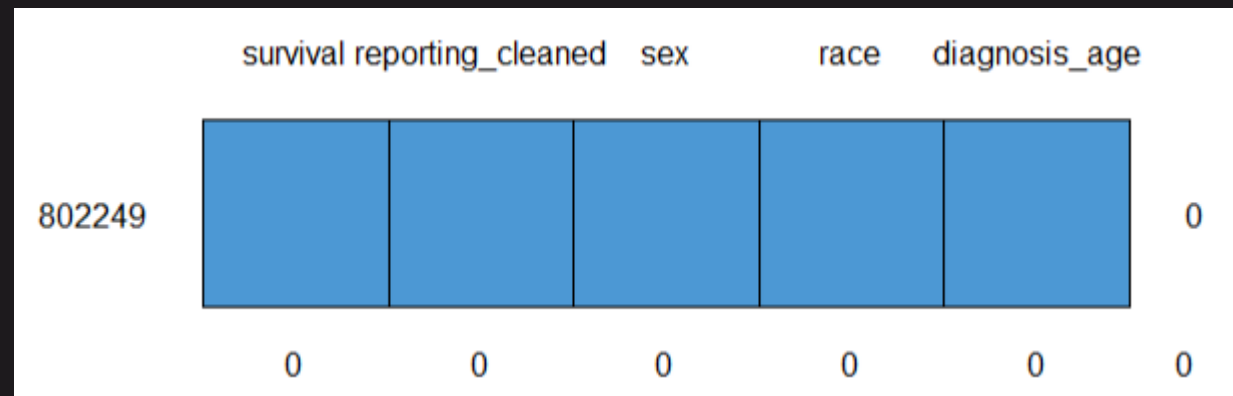## *Table 1, or should I call it… Tableone*

Results:

MICE
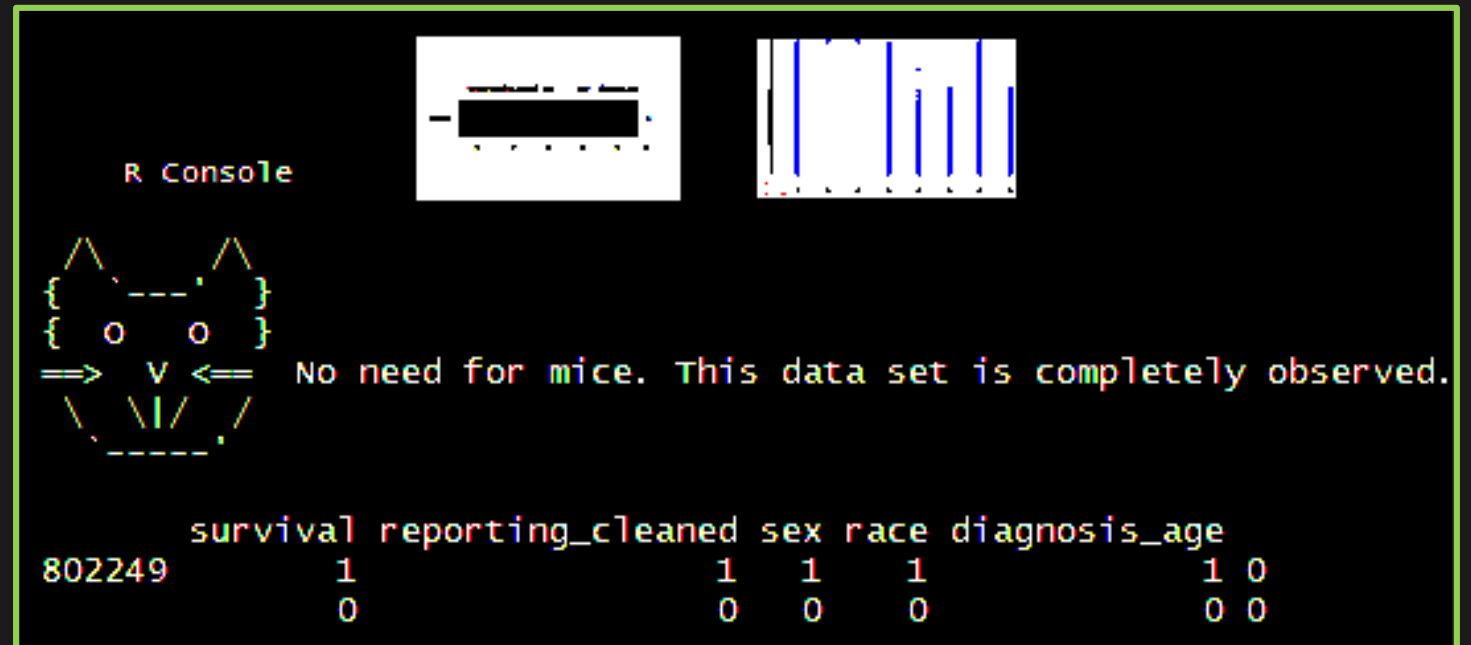
...no mice

Results:

Univariate analysis

	Hazard	P value
Laboratory	1.47	<.001
Nursing home	30.46	<.001
Surgery center	0.67	<.001
Physicians office	3.14	<.001
Oncology center	0.95	>.05

Results:

Multivariate analysis

age adjusted

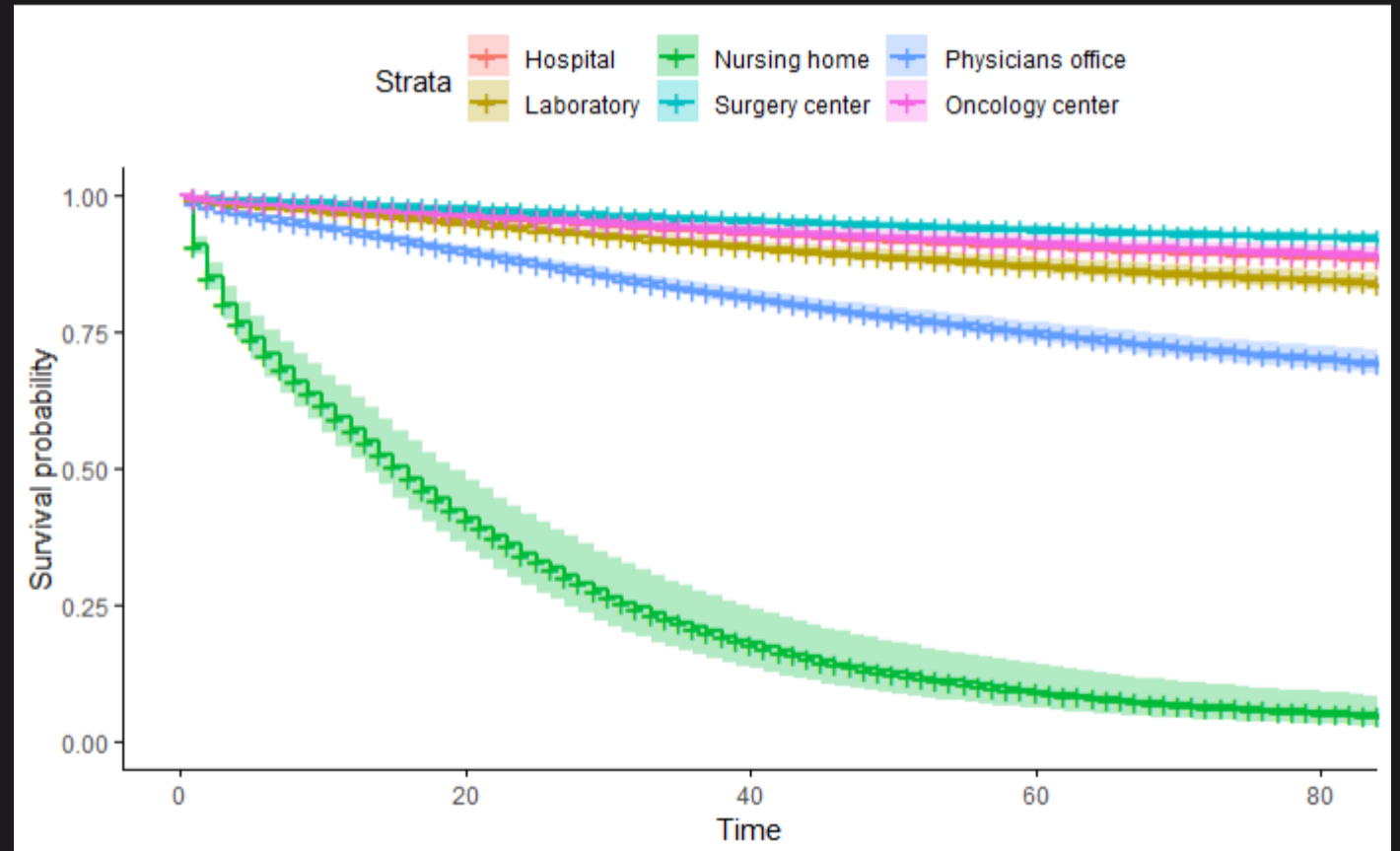|	Hazard	P value
Laboratory	1.43 ▼	<.001
Nursing home	24.74 ▼	<.001
Surgery center	0.67	<.001
Physicians office	3.00 ▼	<.001
Oncology center	0.94	>.05

Results:

Multivariate analysis

more robust

	Hazard	P value
Laboratory	0.98	>.05
Nursing home	7.89 ▼	<.001
Surgery center	0.83 ▲	<.001
Physicians office	2.08 ▼	<.001
Oncology center	0.91	>.05

Results:

Dem curves

Discussion:

Method
Limitations
Inference

- Multivariate and the problem of missing data

- Rounding down the dataset

- χ_2 test and mediators, hypothesis testing

- What is it about nursing homes?

- Prioritizing checkup programs

Questions

- *?*

Acknowledgements

- Google
- Package TableOne, Kazuki Yoshida
- Dr. Johnson's hour long help session
- Kyle's emails
- My group mates in DC course