Behavioral Applications of Voter Files

Eitan Hersh

Tufts University
Department of Political Science

June 14, 2018

Overview

Overview

1. Voter files as is

Overview

1. Voter files as is
2. Linked geographically

Overview

1. Voter files as is
2. Linked geographically
3. Linked individually

Overview

1. Voter files as is
2. Linked geographically
3. Linked individually
4. Some lessons

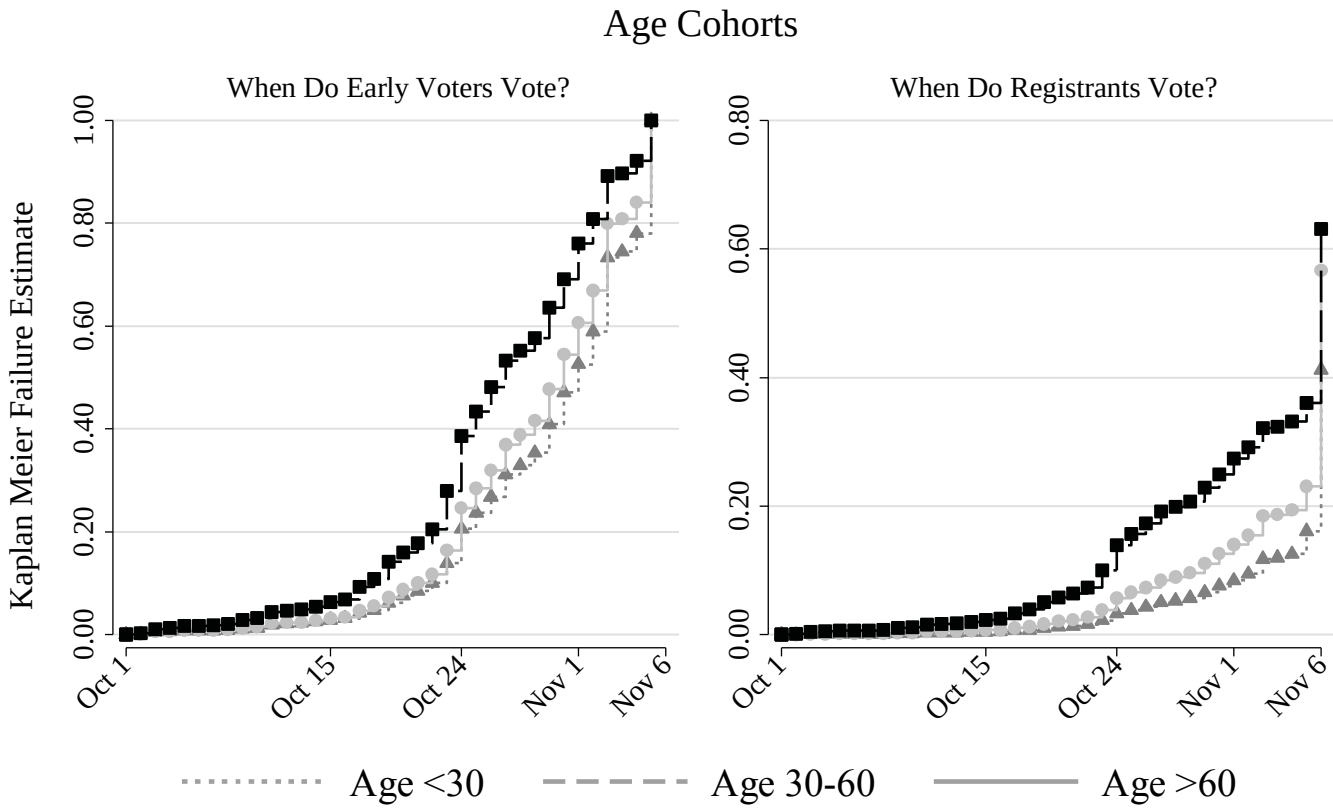
Voters files, as is

Voters files, as is

Example 1: Vivekinan Ashok, et al, "The Dynamic Election: Patterns of Early Voting across Time, State, Party, and Age," *Election Law Journal*, 2016.

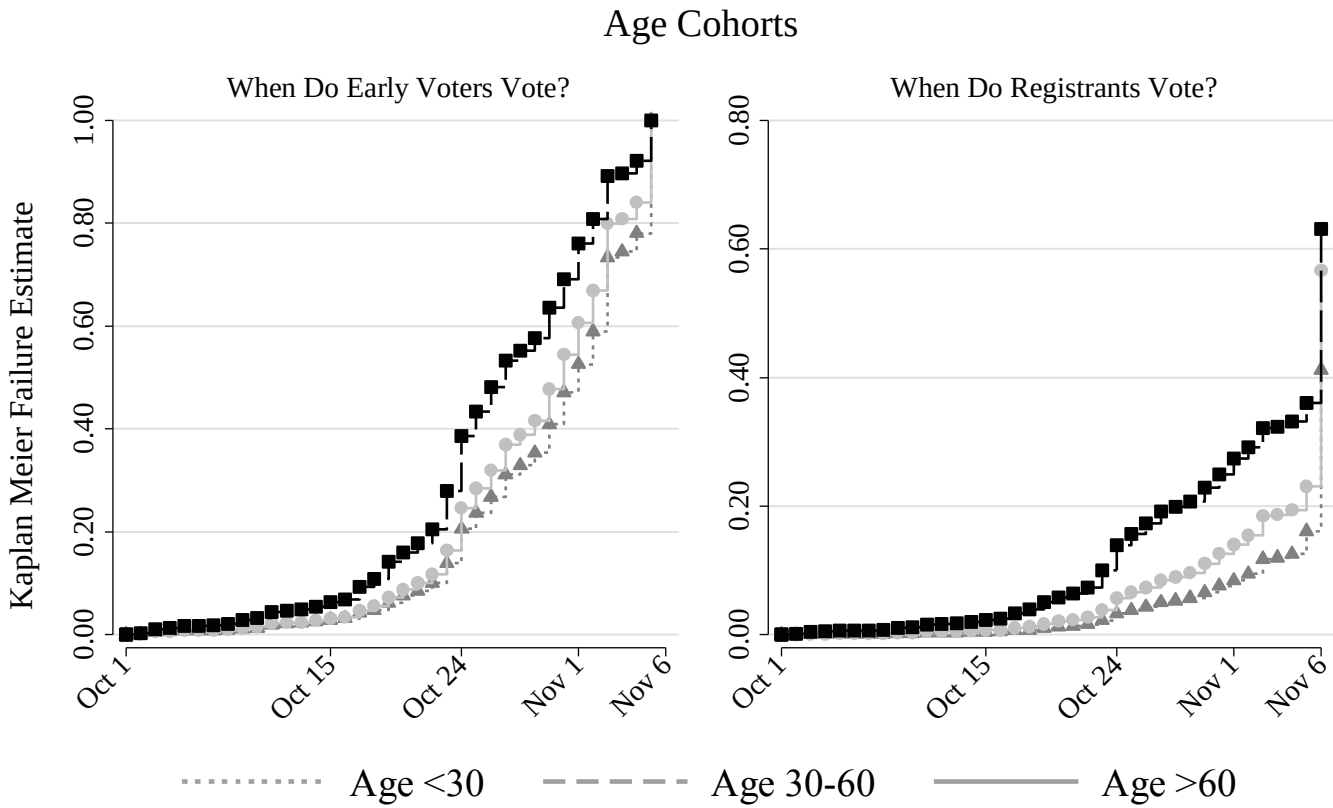
Voters files, as is

Example 1: Vivekinan Ashok, et al, "The Dynamic Election: Patterns of Early Voting across Time, State, Party, and Age," *Election Law Journal*, 2016.



Voters files, as is

Example 1: Vivekinan Ashok, et al, "The Dynamic Election: Patterns of Early Voting across Time, State, Party, and Age," *Election Law Journal*, 2016.



Basic Advantages: Sample size, official turnout, demographics, location data, daily updates

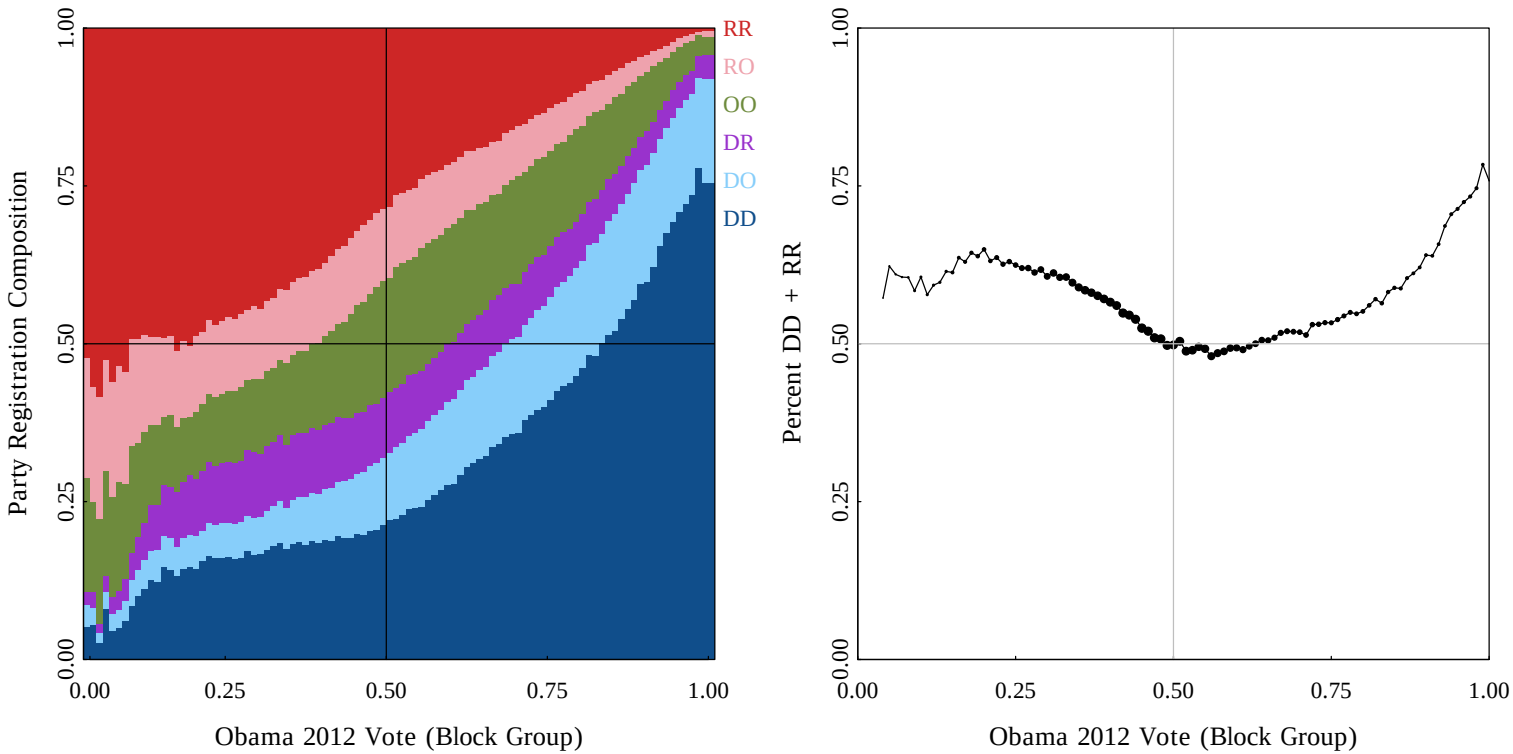
Voter files, linked to geographic data

Voter files, linked to geographic data

Example 2: Eitan Hersh and Yair Ghitza, "Mixed Partisan Households and Electoral Participation in the United States," Under Review.

Voter files, linked to geographic data

Example 2: Eitan Hersh and Yair Ghitza, "Mixed Partisan Households and Electoral Participation in the United States," Under Review.

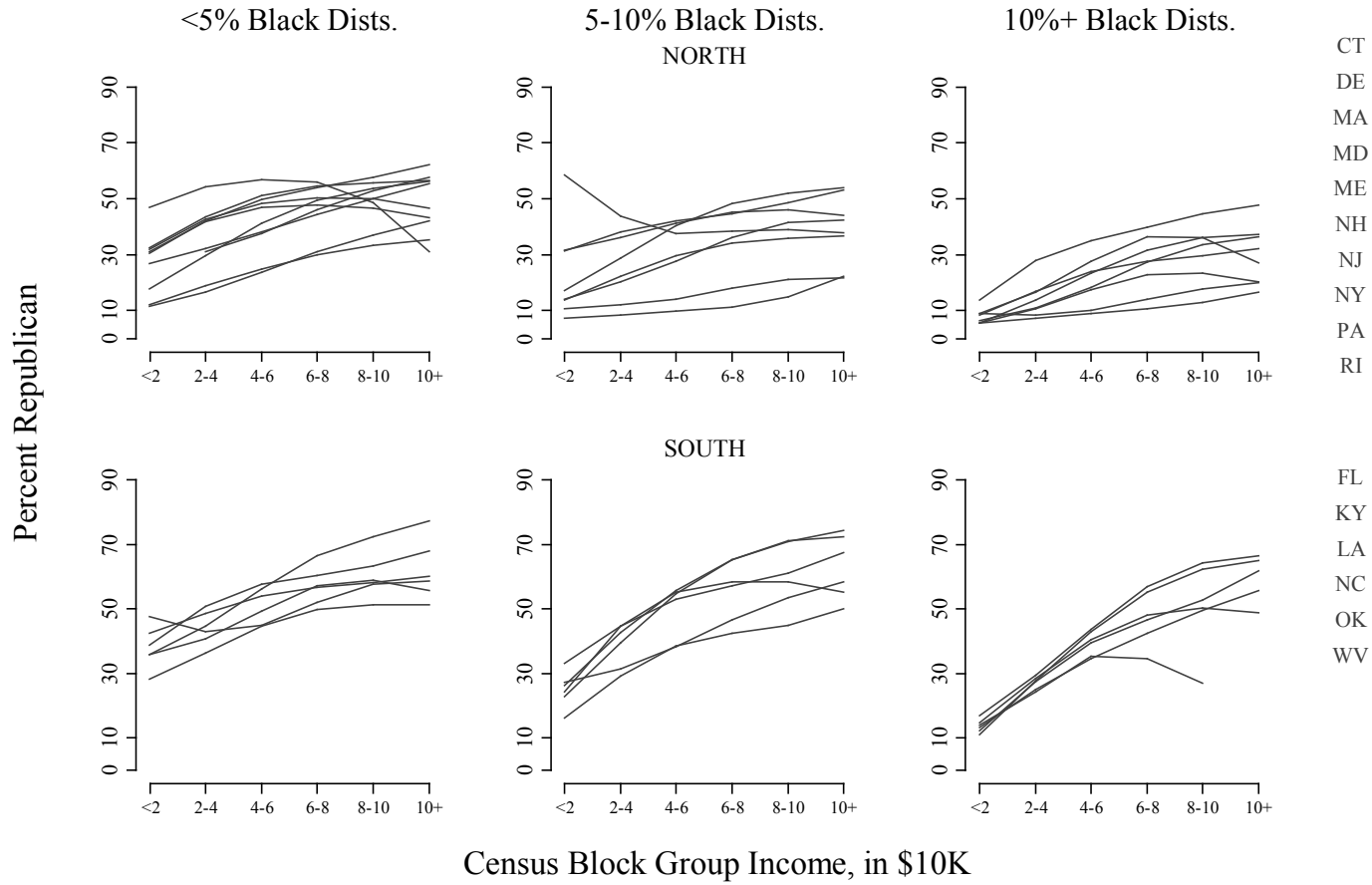


Voter files, linked to geographic data

Example 3: Eitan Hersh and Clayton Nall, "The Primacy of Race in the Geography of Income-Based Voting," *American Journal of Political Science*, 2016.

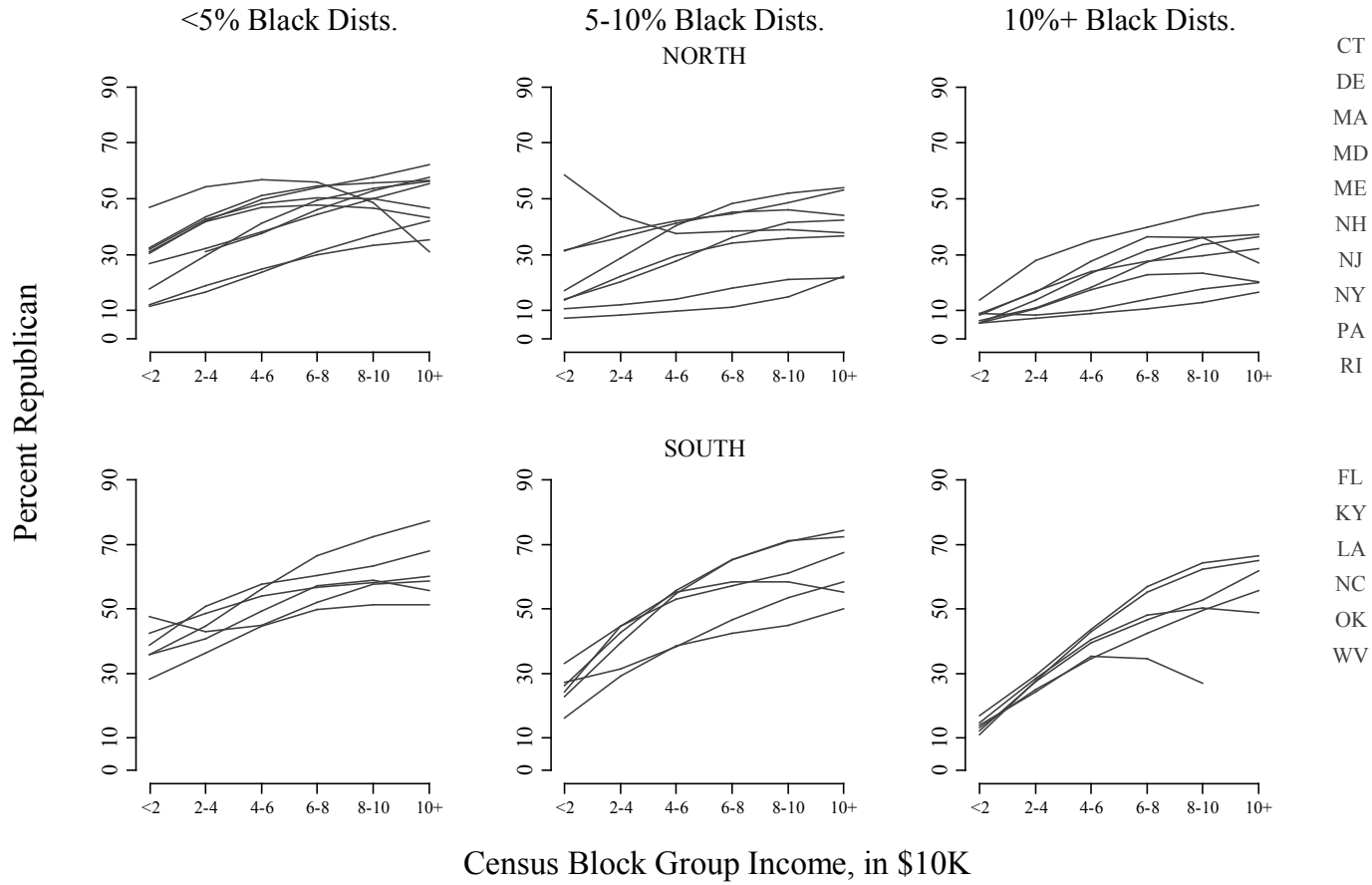
Voter files, linked to geographic data

Example 3: Eitan Hersh and Clayton Nall, "The Primacy of Race in the Geography of Income-Based Voting," *American Journal of Political Science*, 2016.



Voter files, linked to geographic data

Example 3: Eitan Hersh and Clayton Nall, "The Primacy of Race in the Geography of Income-Based Voting," *American Journal of Political Science*, 2016.



Advantages: nested geographies; households and neighborhoods, custom defined; vendors do the hard work

Voter files, individuals linked to external data

Voter files, individuals linked to external data

1. Completely outsource to vendor

Voter files, individuals linked to external data

1. Completely outsource to vendor
2. Completely in-house linkage

Voter files, individuals linked to external data

1. Completely outsource to vendor
2. Completely in-house linkage
3. Hybrid outsourced/custom linkage

Voter files, individuals linked to external data

Voter files, individuals linked to external data

1. Completely outsource to vendor

Voter files, individuals linked to external data

1. Completely outsource to vendor
   - Example 4: Validation of CCES with Catalist voter file (Ansolabehere and Hersh, "Validation," *Political Analysis*, 2012.)

Voter files, individuals linked to external data

1. Completely outsource to vendor
   - Example 4: Validation of CCES with Catalist voter file (Ansolabehere and Hersh, "Validation," *Political Analysis*, 2012.)
   - Example 5: Survey of 2016 max-out donors (Hersh and Schaffner, in draft)

Voter files, individuals linked to external data

1. Completely outsource to vendor
   - Example 4: Validation of CCES with Catalist voter file (Ansolabehere and Hersh, "Validation," *Political Analysis*, 2012.)
   - Example 5: Survey of 2016 max-out donors (Hersh and Schaffner, in draft)
2. Completely in-house linkage

In-House Individual Link

Example 6: Eitan Hersh, "Long Term Effect of September 11 on the Political Behavior of Victims' Families and Neighbors" *Proceedings of the National Academy of Sciences*, 2013.

In-House Individual Link

Example 6: Eitan Hersh, "Long Term Effect of September 11 on the Political Behavior of Victims' Families and Neighbors" *Proceedings of the National Academy of Sciences*, 2013.

Research question: Did 9/11 change the politics of victims' families and neighbors?

In-House Individual Link

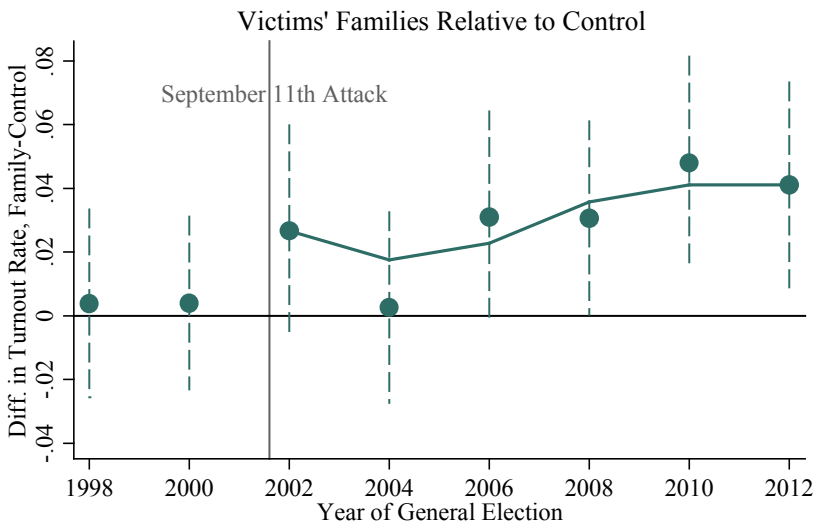
Example 6: Eitan Hersh, "Long Term Effect of September 11 on the Political Behavior of Victims' Families and Neighbors" *Proceedings of the National Academy of Sciences*, 2013.

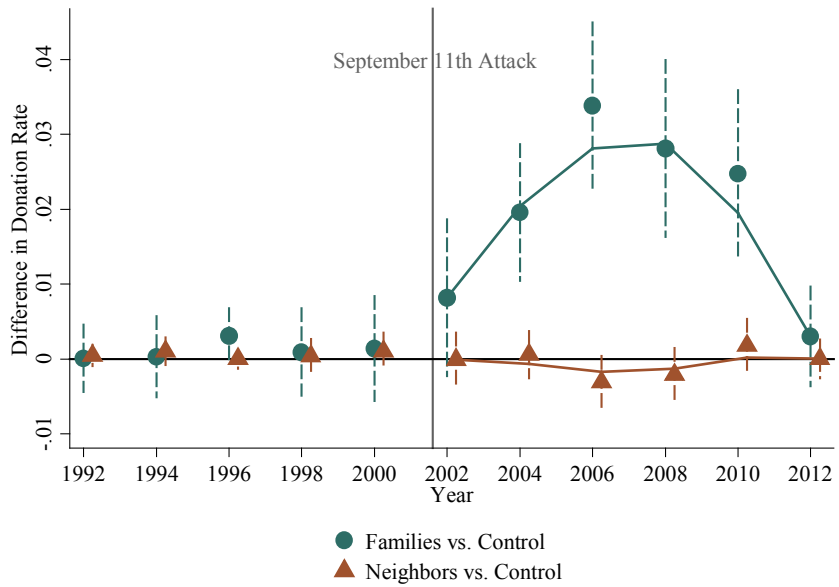
Research question: Did 9/11 change the politics of victims' families and neighbors?

Data

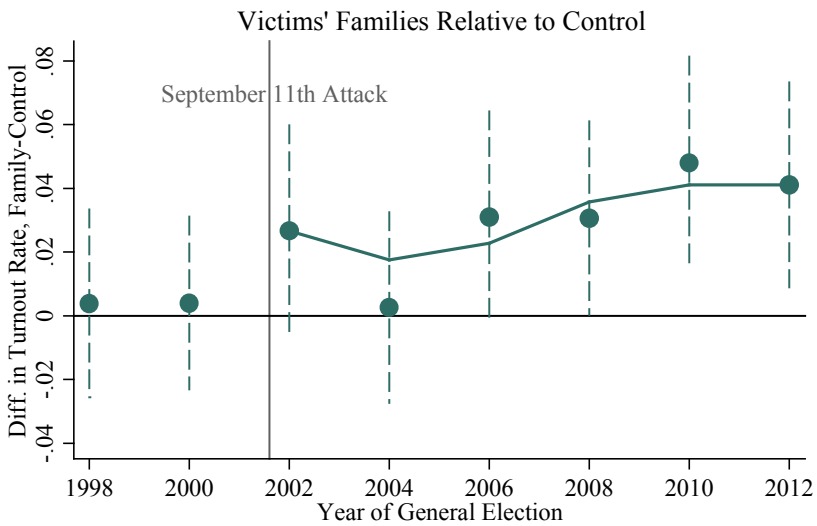
- Obituary data from NYT
- Pre-9/11 NY voter file (L2)
- Year 2013 NY voter file (Catalist)

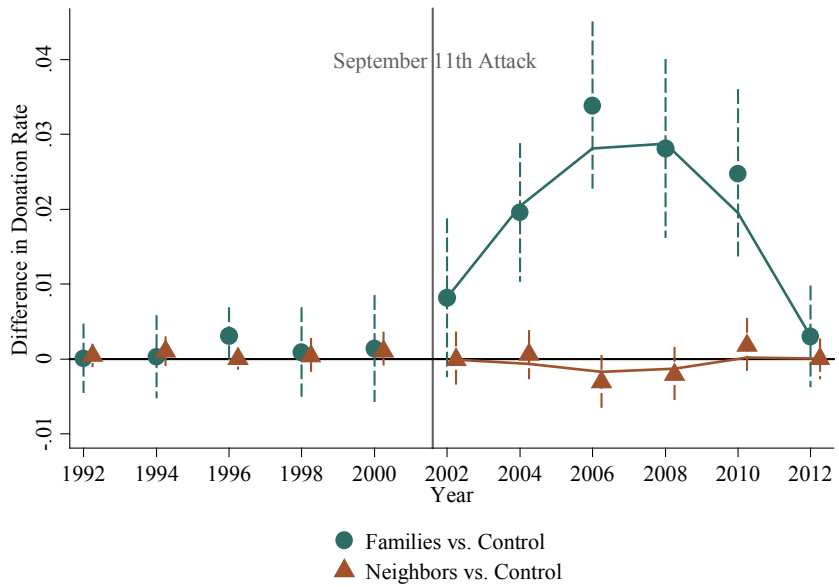
In-House Individual Link





In-House Individual Link





Advantages: long time frame, precise geography, pseudo control group, no misreporting, highly customized

In-House Individual Link

Example 7: Stephen Ansolabehere and Eitan Hersh, "ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name," *Statistics and Public Policy*, Forthcoming.

In-House Individual Link

Example 7: Stephen Ansolabehere and Eitan Hersh, "ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name," *Statistics and Public Policy*, Forthcoming.

Research question: Are protected racial groups less likely to possess valid photo identification in Texas?

In-House Individual Link

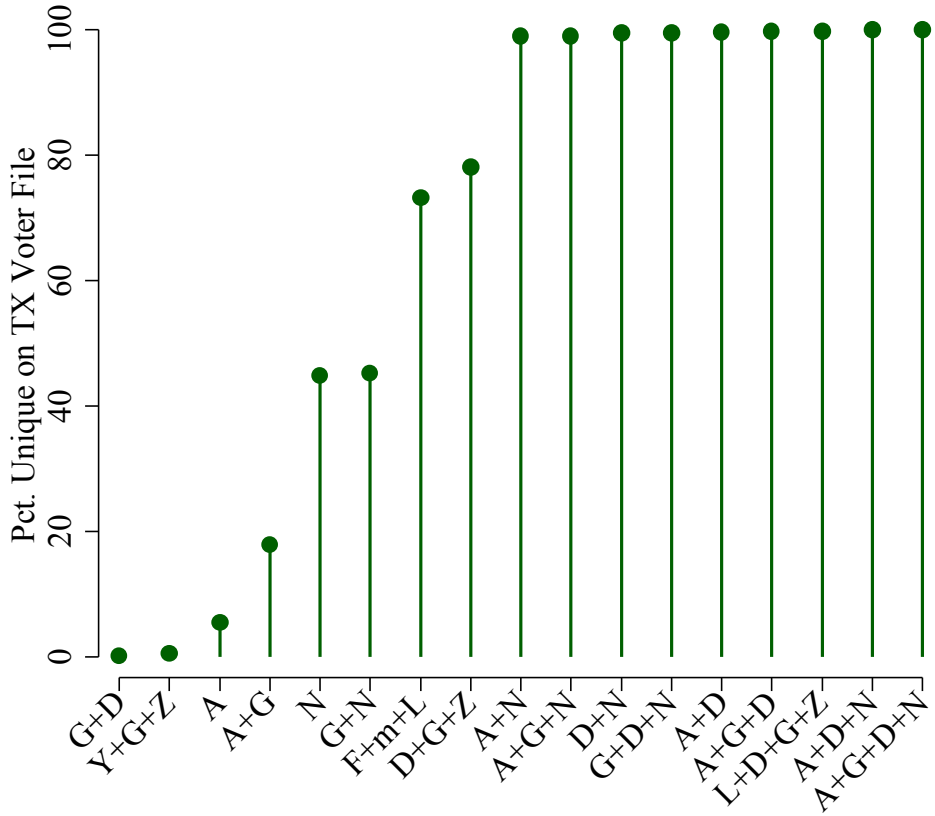
Example 7: Stephen Ansolabehere and Eitan Hersh, "ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name," *Statistics and Public Policy*, Forthcoming.

Research question: Are protected racial groups less likely to possess valid photo identification in Texas?

Data

- Texas voter file
- 10 ID databases (e.g. drivers, passport holders)

In-House Individual Link



A=Zip5+Street Num; G=Gender; D=MDY of Birth; N=First+Last Name
Y=Year of Birth; Z=Zip5; F=First Name; m=Middle Initial; L=Last Name

In-House Individual Link

Table: Number and Percent of instances of NO-MATCH and MATCH by Racial Group, Using Catalist Racial Classification

Race	NO-MATCH	MATCH	Total
Anglo	296,156 (3.6%)	7,949,860 (96.4%)	8,246,016
Black	127,908 (7.5%)	1,569,861 (92.5%)	1,707,769
Hispanic	174,715 (5.7%)	2,867,782 (94.2%)	3,042,497
Other	9,691 (2.0%)	481,621 (98.0%)	491,312
All	608,470 (4.5%)	12,879,124 (95.5%)	13,487,594

In-House Individual Link

Table: Number and Percent of instances of NO-MATCH and MATCH by Racial Group, Using Catalist Racial Classification

Race	NO-MATCH	MATCH	Total
Anglo	296,156 (3.6%)	7,949,860 (96.4%)	8,246,016
Black	127,908 (7.5%)	1,569,861 (92.5%)	1,707,769
Hispanic	174,715 (5.7%)	2,867,782 (94.2%)	3,042,497
Other	9,691 (2.0%)	481,621 (98.0%)	491,312
All	608,470 (4.5%)	12,879,124 (95.5%)	13,487,594

Advantages: direct answer to question, customized algorithm for legal purpose

Voter files, individuals linked to external data

1. Completely outsource to vendor
2. Completely in-house linkage

Voter files, individuals linked to external data

1. Completely outsource to vendor
2. Completely in-house linkage
3. Hybrid outsourced/custom linkage

Hybrid Individual Link

Example 8: Eitan Hersh and Matthew Goldenberg, "Democratic and Republican Physicians Provide Different Care on Politicized Health Issues" *Proceedings of the National Academy of Sciences*, 2016.

Hybrid Individual Link

Example 8: Eitan Hersh and Matthew Goldenberg, "Democratic and Republican Physicians Provide Different Care on Politicized Health Issues" *Proceedings of the National Academy of Sciences*, 2016.

Research question: Do Democratic and Republican physicians treat patients differently?

Hybrid Individual Link

Example 8: Eitan Hersh and Matthew Goldenberg, "Democratic and Republican Physicians Provide Different Care on Politicized Health Issues" *Proceedings of the National Academy of Sciences*, 2016.

Research question: Do Democratic and Republican physicians treat patients differently?

Data

- National Provider Identification (NPI) File, CMS
- Voter file in 29 party-registration states
- Survey of primary care physicians

Targeting the Population

Targeting the Population

Step 1 The National Provider Identification File

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches,

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches, plus 5,820 confident matches (57% match rate)

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches, plus 5,820 confident matches (57% match rate)
3. Of 24,250 confident matches, 20,296 mailable addresses

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches, plus 5,820 confident matches (57% match rate)
3. Of 24,250 confident matches, 20,296 mailable addresses

Step 3 Target sample

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches, plus 5,820 confident matches (57% match rate)
3. Of 24,250 confident matches, 20,296 mailable addresses

Step 3 Target sample

Of 20,296, 13,678 when restricted to registered Ds and Rs

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches, plus 5,820 confident matches (57% match rate)
3. Of 24,250 confident matches, 20,296 mailable addresses

Step 3 Target sample

Of 20,296, 13,678 when restricted to registered Ds and Rs

Oversample doctors in mixed-partisan practices

1. Classify practices by size and partisan composition
2. Draw 100% physicians from mid-size bipartisan stratum (754) and 6% sample of all other strata (775)

Targeting the Population

Step 1 The National Provider Identification File

1. 560,896 U.S.-based physicians
2. 149,936 in primary care specialties (internal, family, adult, general practice)
3. 85,722 in 29 party-registration states
4. 42,861 in 50% simple random sample

Step 2 Match to Voter File (Catalist)

1. 161,553 plausible matches sent by Catalist
2. 18,430 records with unique matches, plus 5,820 confident matches (57% match rate)
3. Of 24,250 confident matches, 20,296 mailable addresses

Step 3 Target sample

Of 20,296, 13,678 when restricted to registered Ds and Rs

Oversample doctors in mixed-partisan practices

1. Classify practices by size and partisan composition
2. Draw 100% physicians from mid-size bipartisan stratum (754) and 6% sample of all other strata (775)
3. Survey 1,529 physicians

The Survey

...We are trying to better understand how doctors take a patient's social history and what factors may impact differences in the ways physicians approach a patient's social history...

The Survey

Nine Vignettes

A healthy-appearing, 38-year old, male patient comes to your office for a physical. This is his first appointment with you. He does not have any known prior chronic medical issues. During the patient interview, the patient...

The Survey

Nine Vignettes

A healthy-appearing, 38-year old, male patient comes to your office for a physical. This is his first appointment with you. He does not have any known prior chronic medical issues. During the patient interview, the patient...

- ... acknowledges consuming about 20 alcoholic beverages in a typical week but denies any related physical concerns.

The Survey

Nine Vignettes

A healthy-appearing, 38-year old, male patient comes to your office for a physical. This is his first appointment with you. He does not have any known prior chronic medical issues. During the patient interview, the patient...

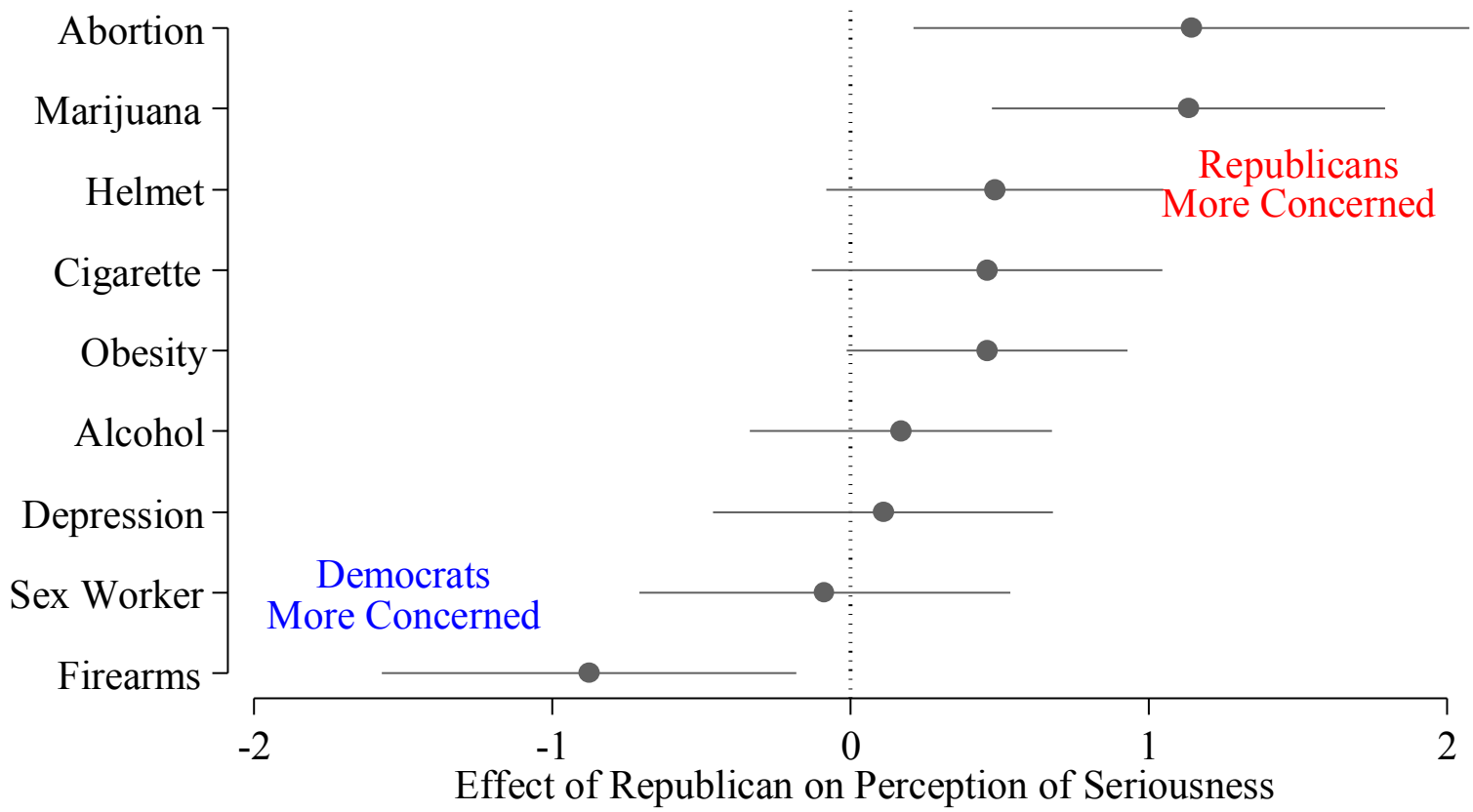
The Survey

Nine Vignettes

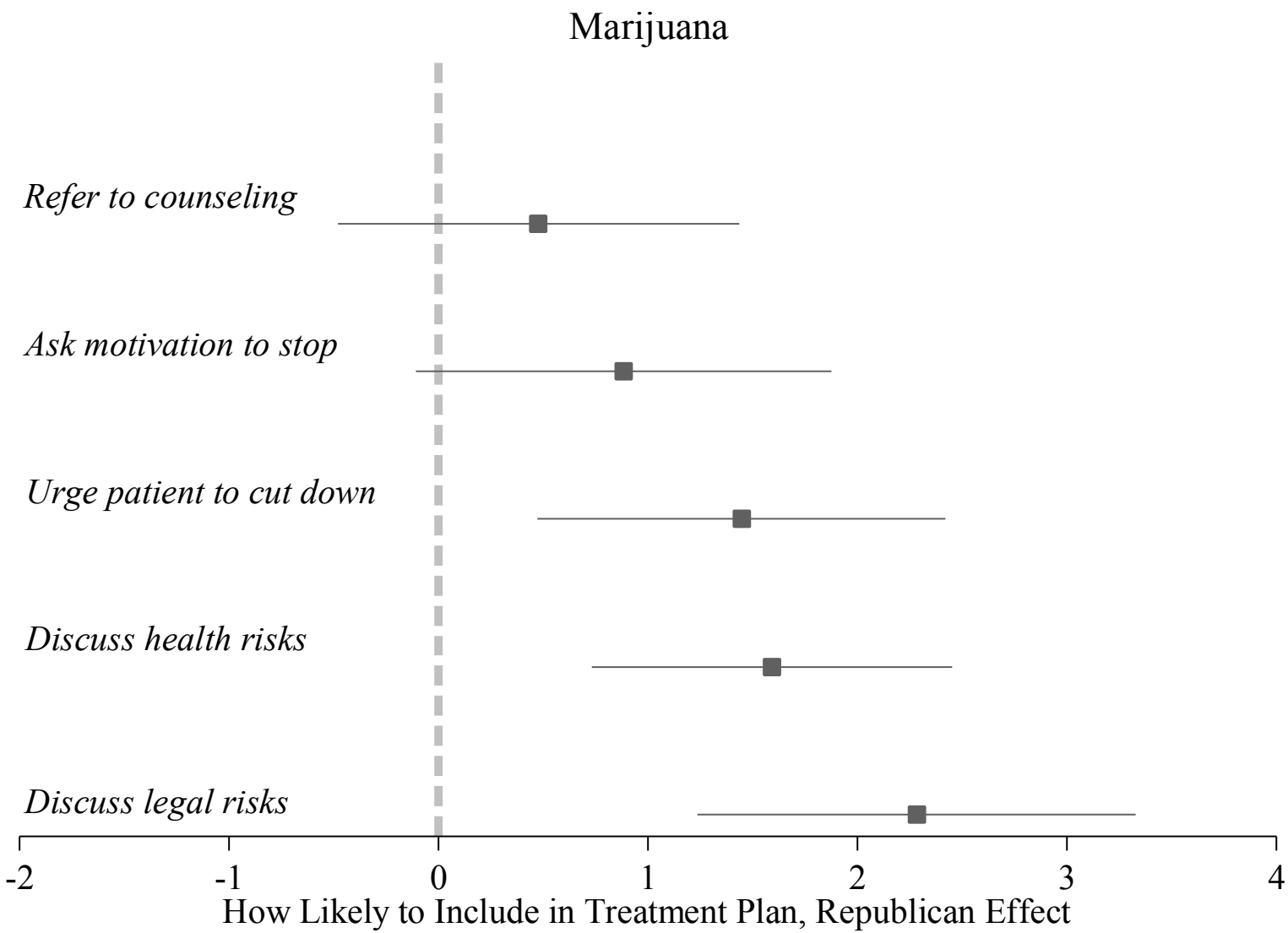
A healthy-appearing, 38-year old, male patient comes to your office for a physical. This is his first appointment with you. He does not have any known prior chronic medical issues. During the patient interview, the patient...

- ... acknowledges using recreational marijuana approximately three times per week but denies any related physical concerns.

Results


Abortion
Marijuana
Helmet
Cigarette
Obesity
Alcohol
Depression
Sex Worker
Firearms
Republicans
More Concerned
Democrats
More Concerned
-2
-1
0
1
2
Effect of Republican on Perception of Seriousness

Results


Marijuana
Refer to counseling
Ask motivation to stop
Urge patient to cut down
Discuss health risks
Discuss legal risks
-2
-1
0
1
2
3
4
How Likely to Include in Treatment Plan, Republican Effect

Hybrid Individual Link

Example 9: Eitan Hersh and Gabrielle Malina, "Partisan Pastor" In Progress

Research questions: Do pastors reflect the political views of their congregants? Do Democratic and Republican pastors lead differently?

Hybrid Individual Link

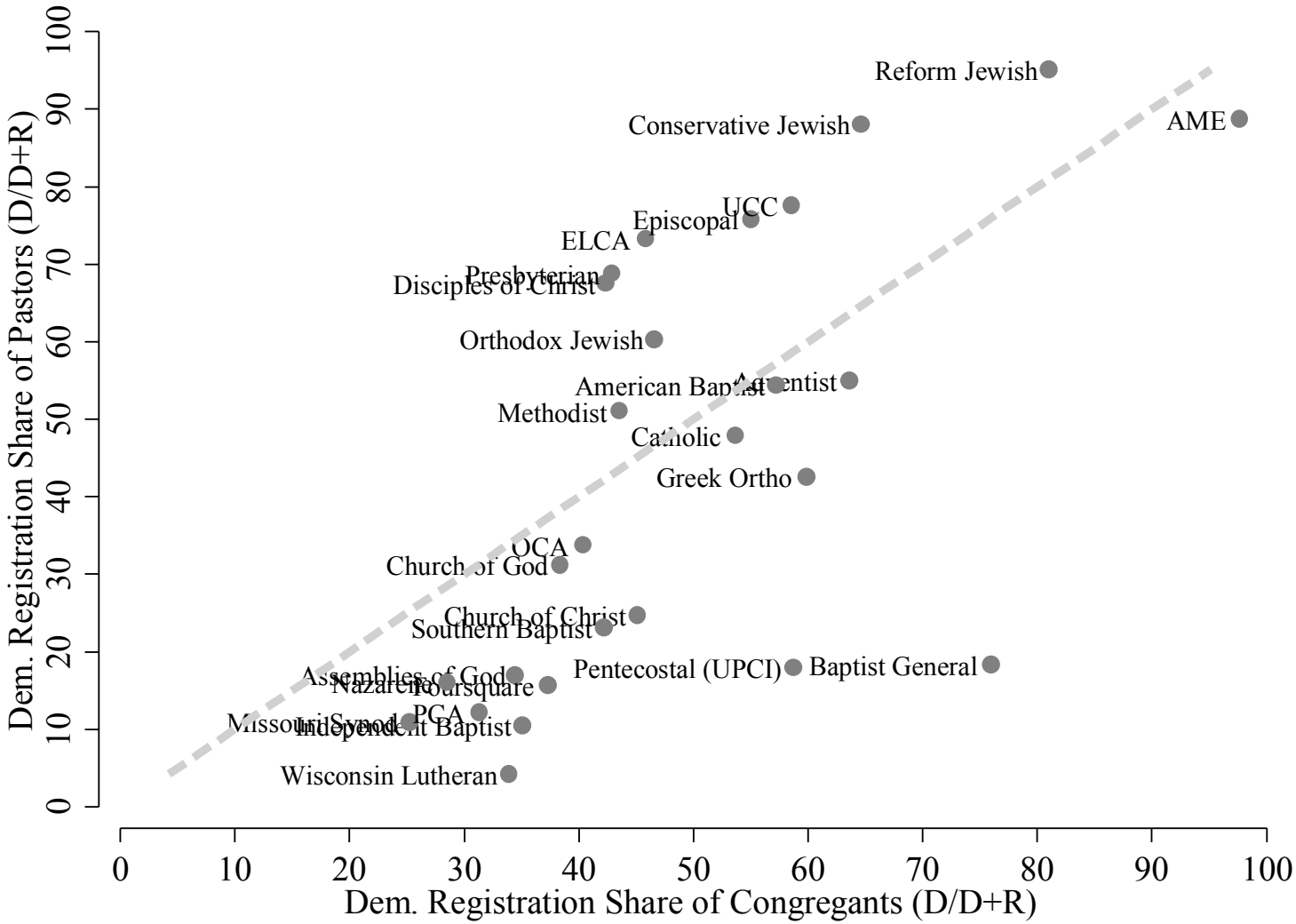
Example 9: Eitan Hersh and Gabrielle Malina, "Partisan Pastor" In Progress

Research questions: Do pastors reflect the political views of their congregants? Do Democratic and Republican pastors lead differently?

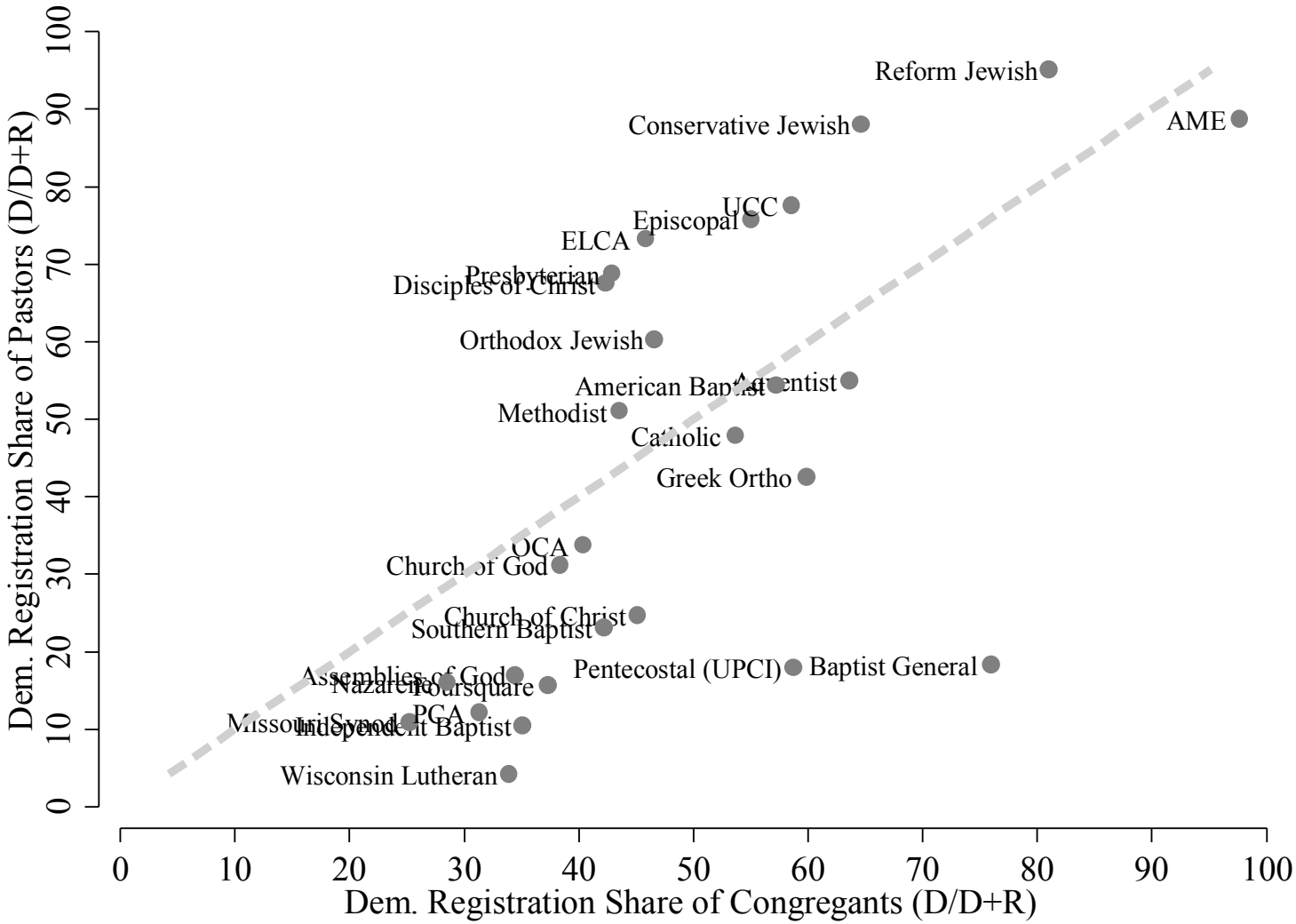
Data

- 41 find-a-church websites, scraped (Upwork)
- Additional pastors added (MTurk)
- Voter file link (Catalist)
- link to survey data

Hybrid Individual Link



Hybrid Individual Link



Advantages: comprehensive data on small but important population, within- and cross-denominational variation

Some Lessons

Some Lessons

Match rate to registered voters

Some Lessons

Match rate to registered voters

	Doctors	Pastors	9/11	Donors
First Pass	43%	44		

Some Lessons

Match rate to registered voters

	Doctors	Pastors	9/11	Donors
First Pass	43%	44		
Second Pass	57	63		

Some Lessons

Match rate to registered voters

	Doctors	Pastors	9/11	Donors
First Pass	43%	44		
Second Pass	57	63		
Only Pass			68	79*

Some Lessons

Mail survey response rates

Some Lessons

Mail survey response rates

Doctors: 20%

Donors: 17%

Some Lessons

Some Lessons

False positives

Some Lessons

False positives

1. 8/1,529 “doctors” (0.5%) said they weren’t the doctor

Some Lessons

False positives

1. 8/1,529 “doctors” (0.5%) said they weren’t the doctor
2. 21/6,500 “donors” (0.3%) said they weren’t the donor

Some Lessons

False positives

1. 8/1,529 “doctors” (0.5%) said they weren’t the doctor
2. 21/6,500 “donors” (0.3%) said they weren’t the donor
3. False positives in Texas case:

Some Lessons

False positives

1. 8/1,529 “doctors” (0.5%) said they weren’t the doctor
2. 21/6,500 “donors” (0.3%) said they weren’t the donor
3. False positives in Texas case:

ADGN Match	SSN9 Match		
	No SSN Match	SSN Match	Total
No ADGN Match	1,207,739 [TN]	135,686 [FN]	1,343,425
ADGN Match	119,601 [FP]	5,249,230 [TP]	5,368,831
Total	1,327,340	5,384,916	6,712,256

Behavioral Applications of Voter Files

Eitan Hersh

Tufts University
Department of Political Science

June 14, 2018