Tutorial: Intuition and basic methods for causal inference over social media data

Emre Kıcıman, emrek@microsoft.com

Includes work done in collaboration with:

Scott Counts, Munmun De Choudhury, Melissa Gasser, Alexandra Olteanu, Matt Richardson, Onur Varol

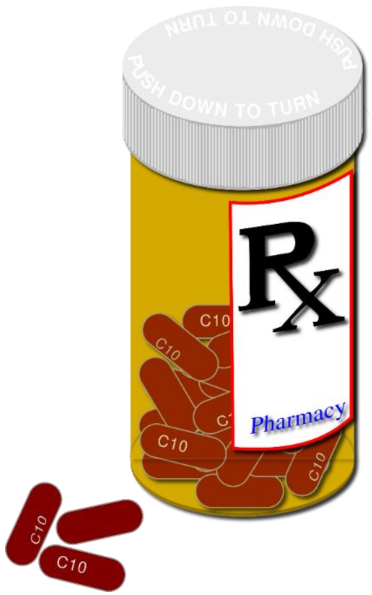
What to expect from this tutorial?

What this tutorial is:

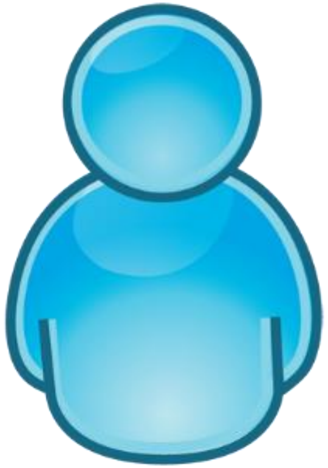
- Intuition to re: causal inference and counterfactual framework
- Examples designing observational studies over social media timelines (Covariates, Treatments, …)

What this tutorial is not:

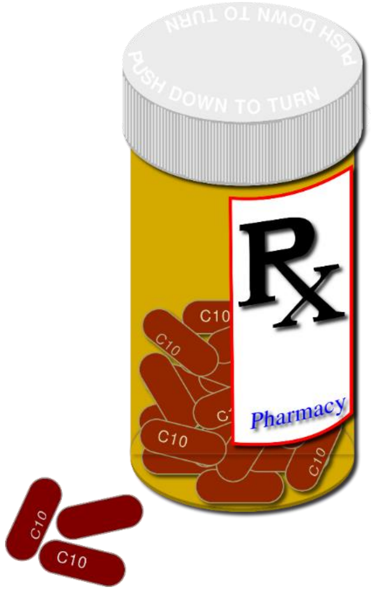
- Algorithmic details
- Coding how-to
- Social network effects


PUSH DOWN TO TURN
PUSH DOWN TO TURN
Rx
Pharmacy
C10
C10
C10
C10
C10

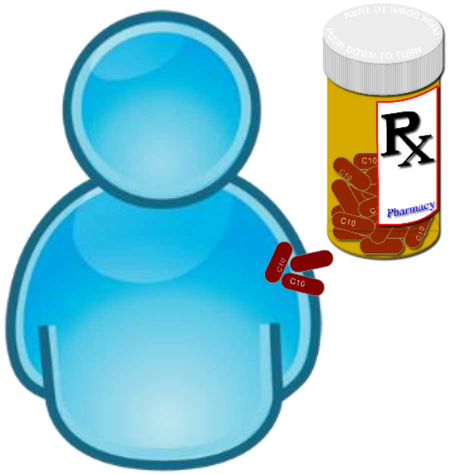
Treatment



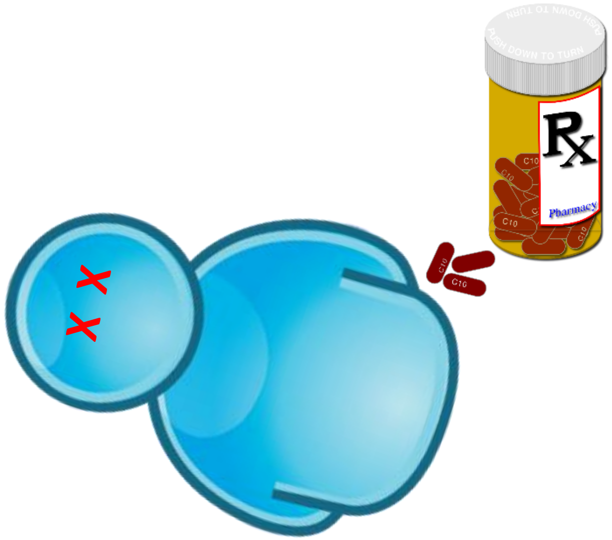
Alice



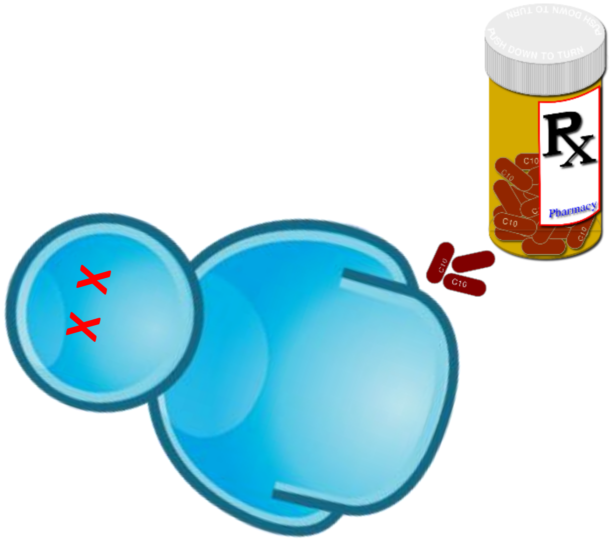
Treatment



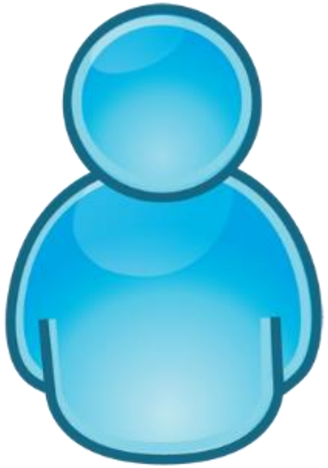
Alice



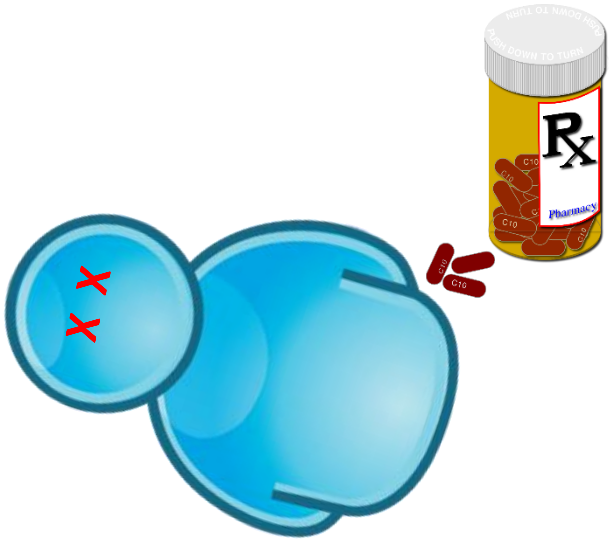
Alice



$$Y^{T=1}$$



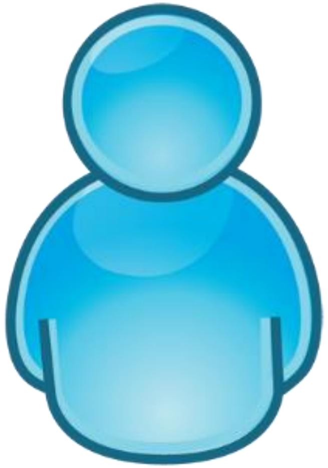
$$Y^{T=0}$$



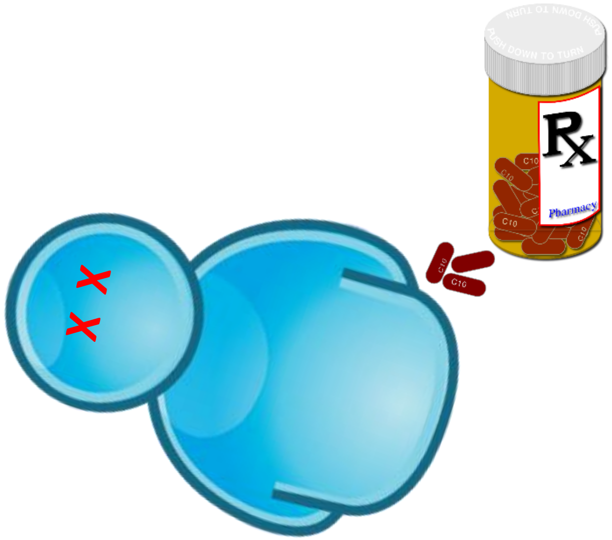
$$Y^{T=1}$$

Treatment effect = $Y^{T=1} - Y^{T=0}$

Treatment effect = $Y^{T=1} - Y^{T=0}$



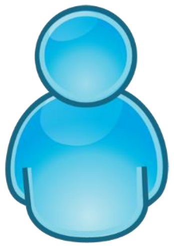
$Y^{T=0}$



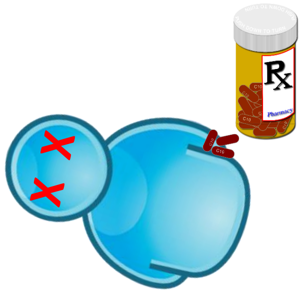
$Y^{T=1}$

Missing data problem

- Estimate missing data values using various methods



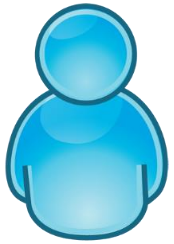
$$\tilde{Y}^{T=0}$$



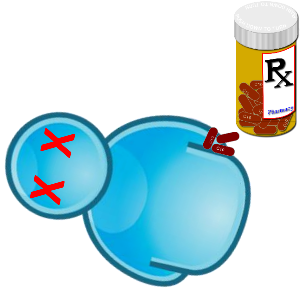
$$Y^{T=1}$$

Missing data problem

- Estimate missing data values using various methods



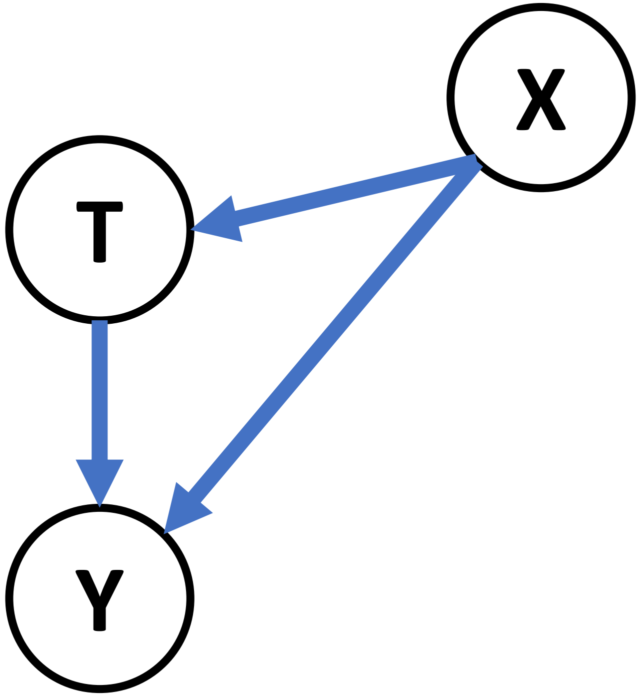
$$Y^{T=0}$$



$$\tilde{Y}^{T=1}$$

Confounding variables

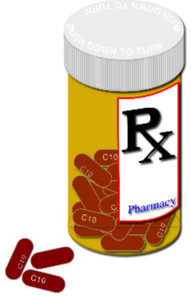
- Goal: learn relationship T → Y
- But other covariates X may also influence Y
- And X may also influence T



- Do T and Y co-vary because of X or because of T → Y?
- To estimate T → Y, we need to break relationship X → T

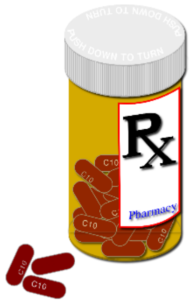
Disentangling covariate: Randomized Expts.

- Randomized assignment of treatment status provides independence: $T \perp X$



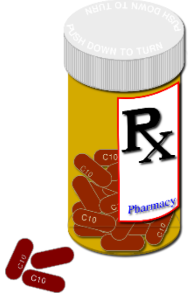
Disentangling covariates: Randomized Expts.

- Randomized assignment of treatment status provides independence: $T \perp X$



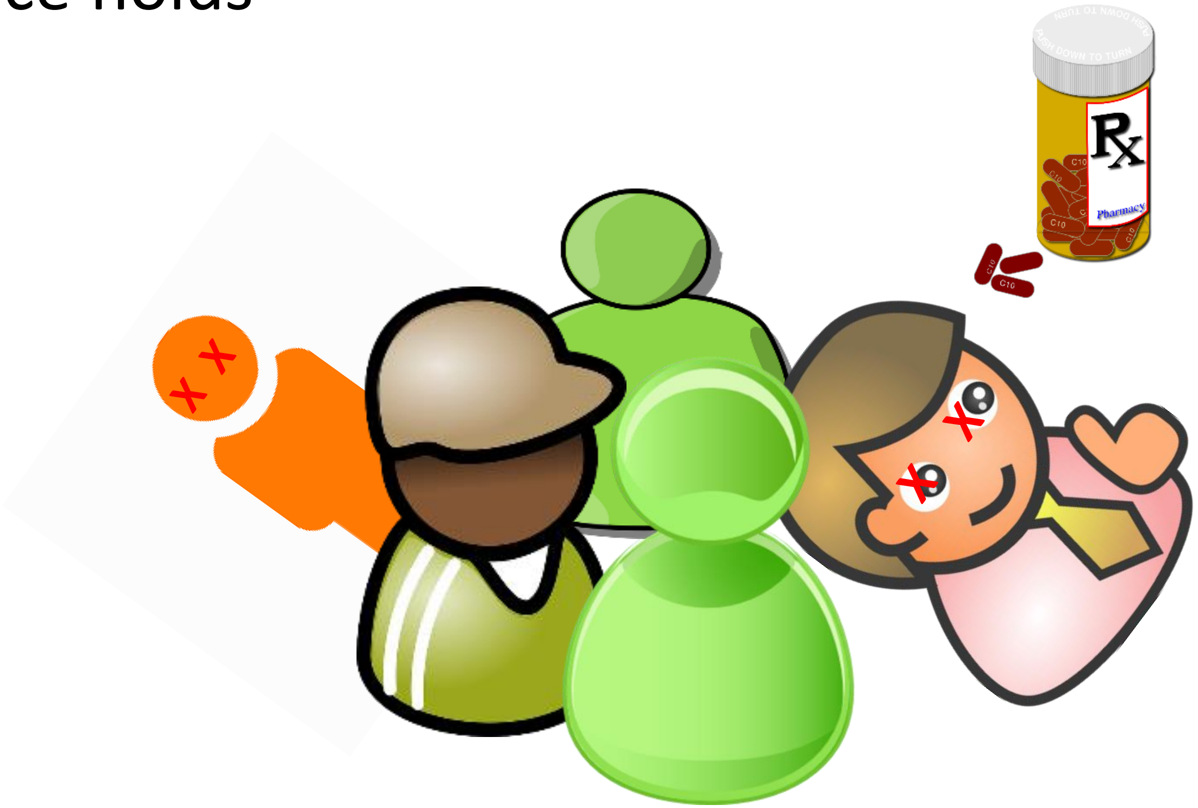
Disentangling covariates: Randomized Expts.

- Randomized assignment of treatment status provides independence: $T \perp X$



Observational Studies (not random)

- … Identify comparable treated and untreated subgroups such that independence holds

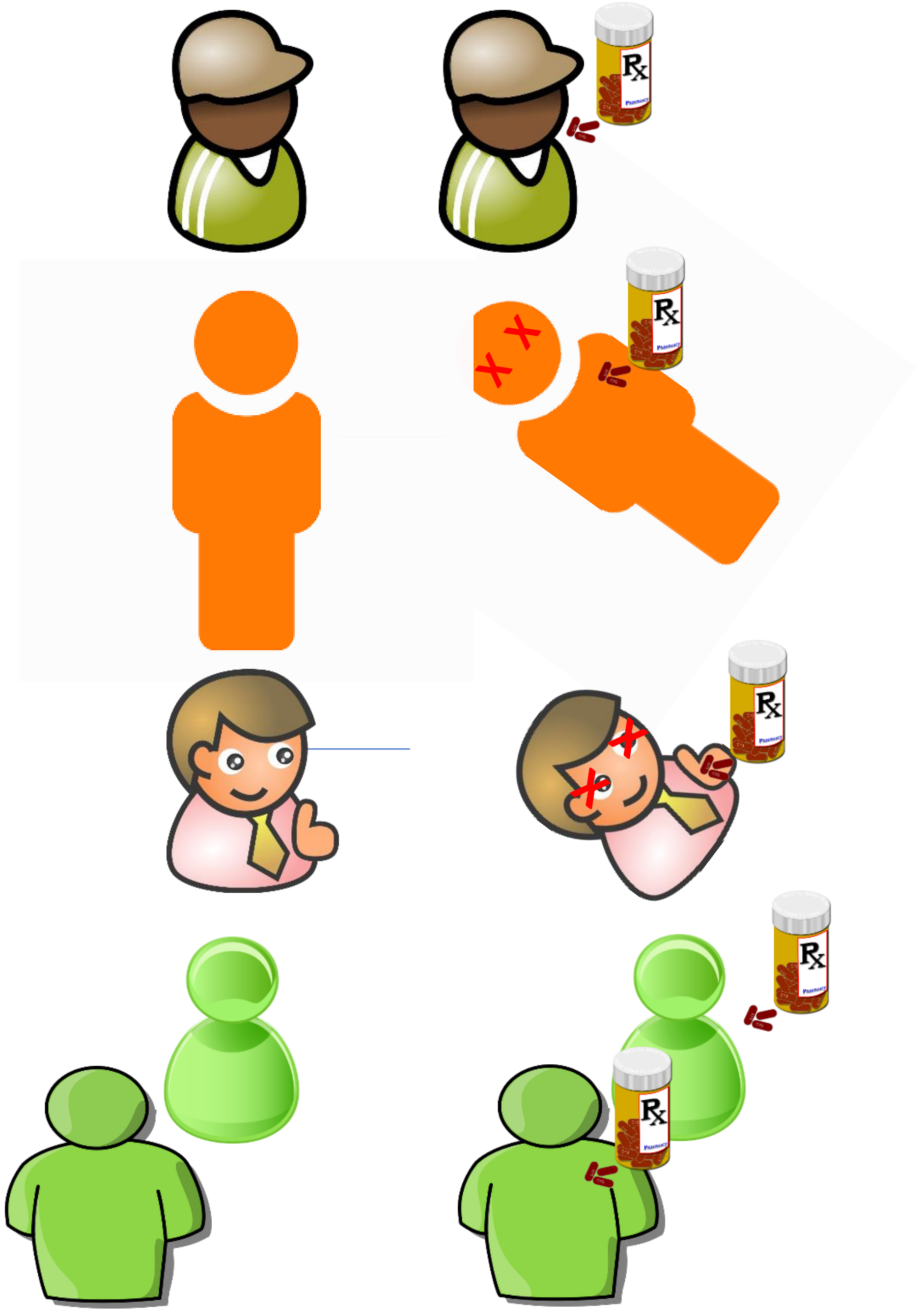


Exact Matching

- For every person who received treatment, find another with identical covariates X who did not.

- By construction, now $\mathrm{T} \perp X$

- Exact matching is hard in high dimensions.



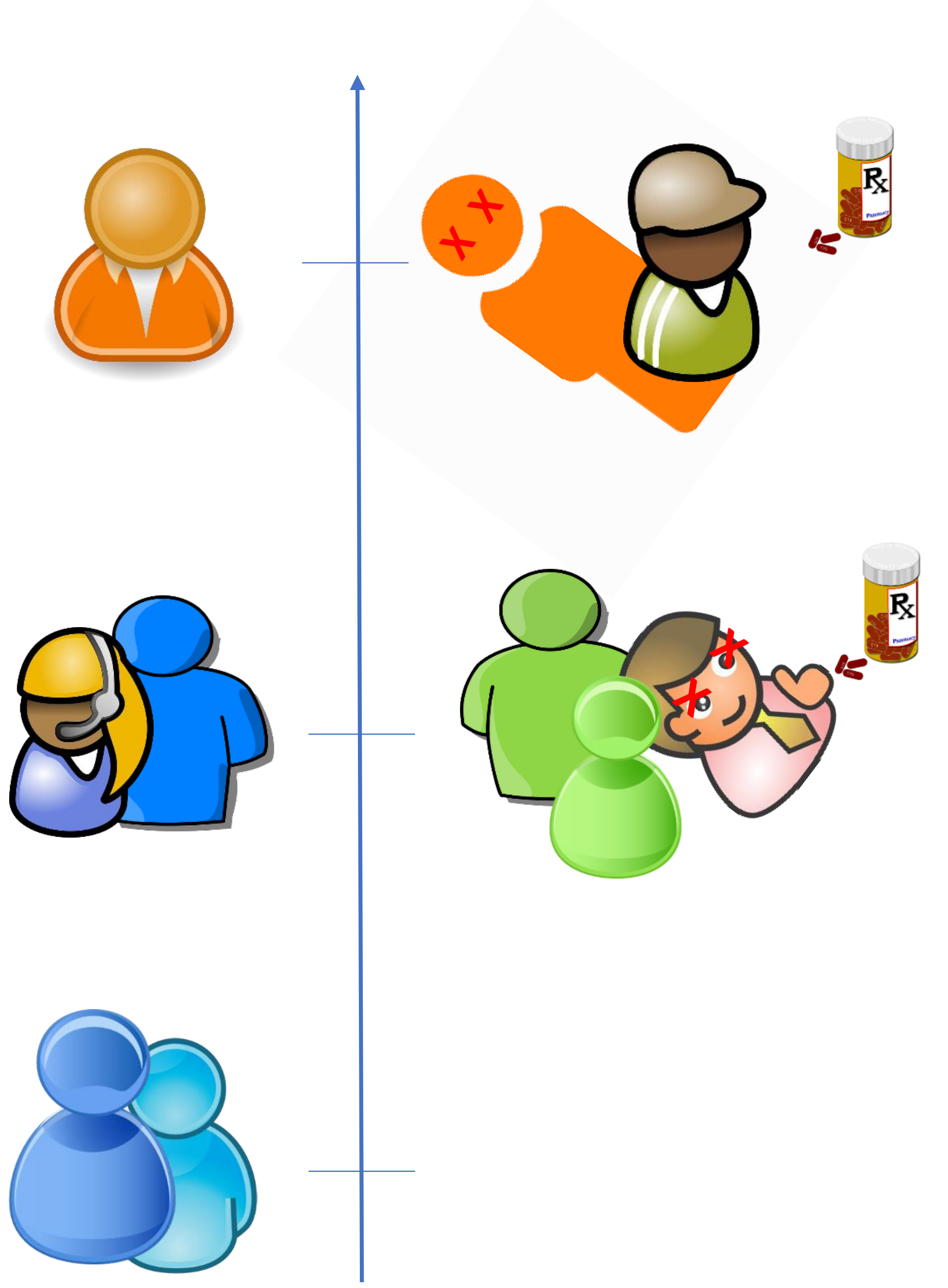
A *balancing score* subdivides observational data so that:

$$T \perp X \mid score$$

Estimated propensity is one possible balancing score

Match on propensity score.

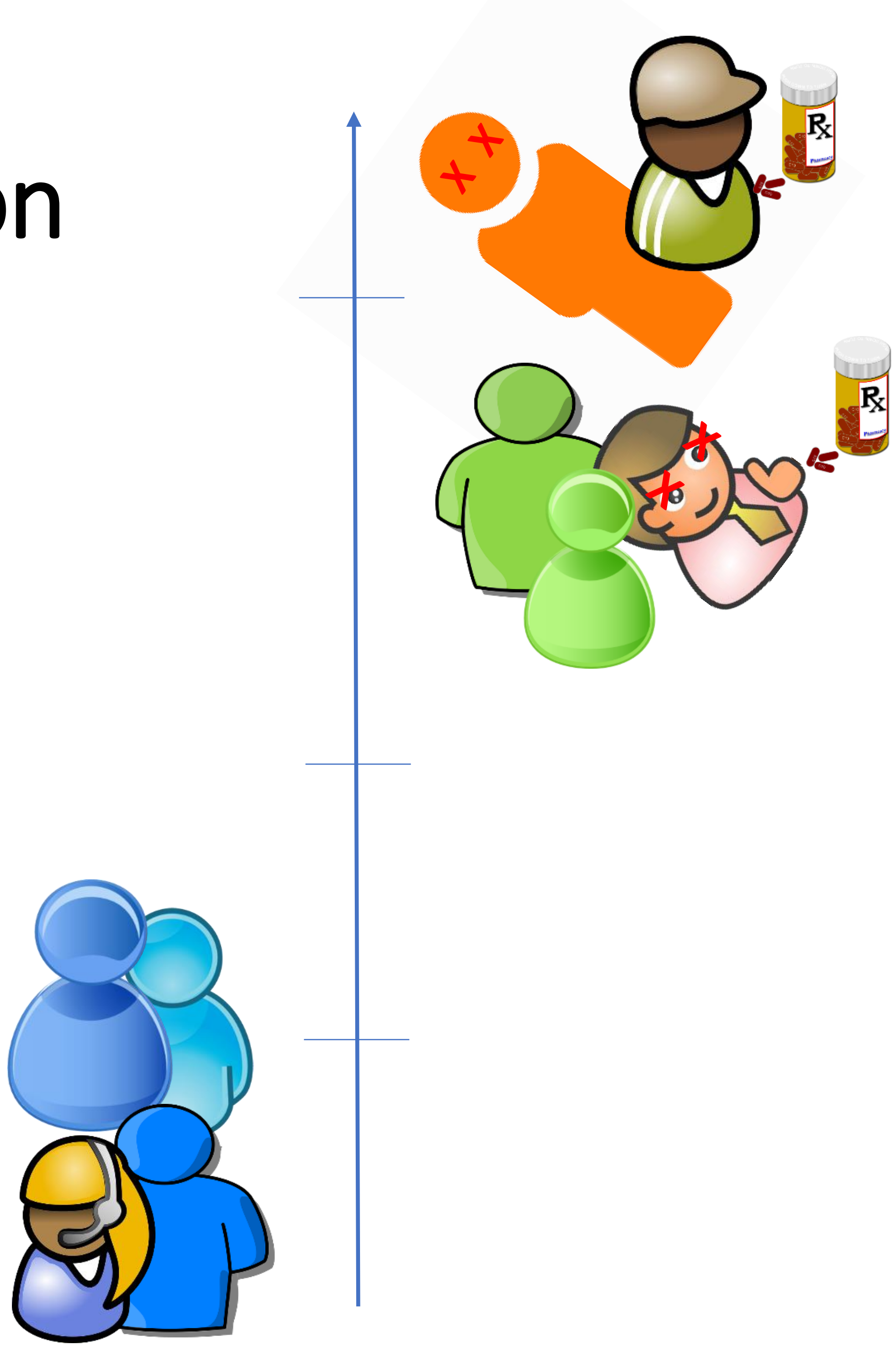
Generalizes to stratification and weighting approaches



Common support assumption

- The treated and untreated populations have to be similar enough

- Otherwise, cannot estimate counterfactual outcomes



Ignorability assumption

- Any unmeasured covariates are irrelevant.

- Under random experiments, $T \perp X$ for both observed and unobserved covariates

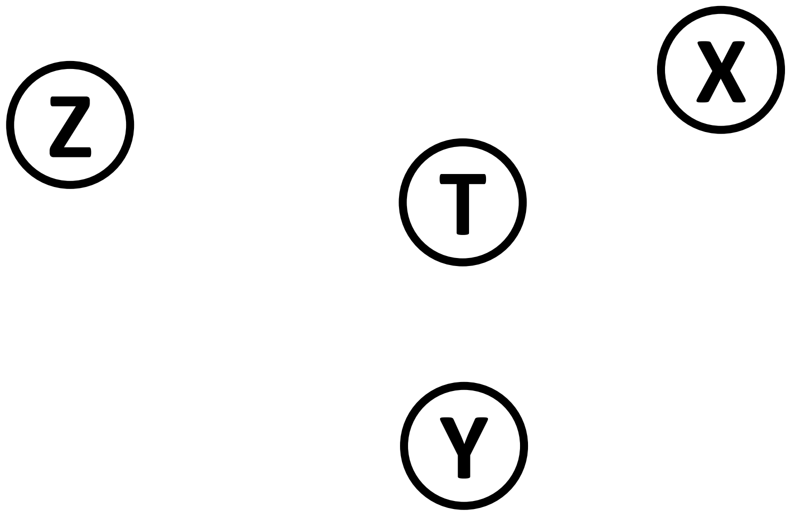
- But matching and related techniques can only construct $T \perp X$ for observed covariates

Other causal inference methods

Regression Discontinuities

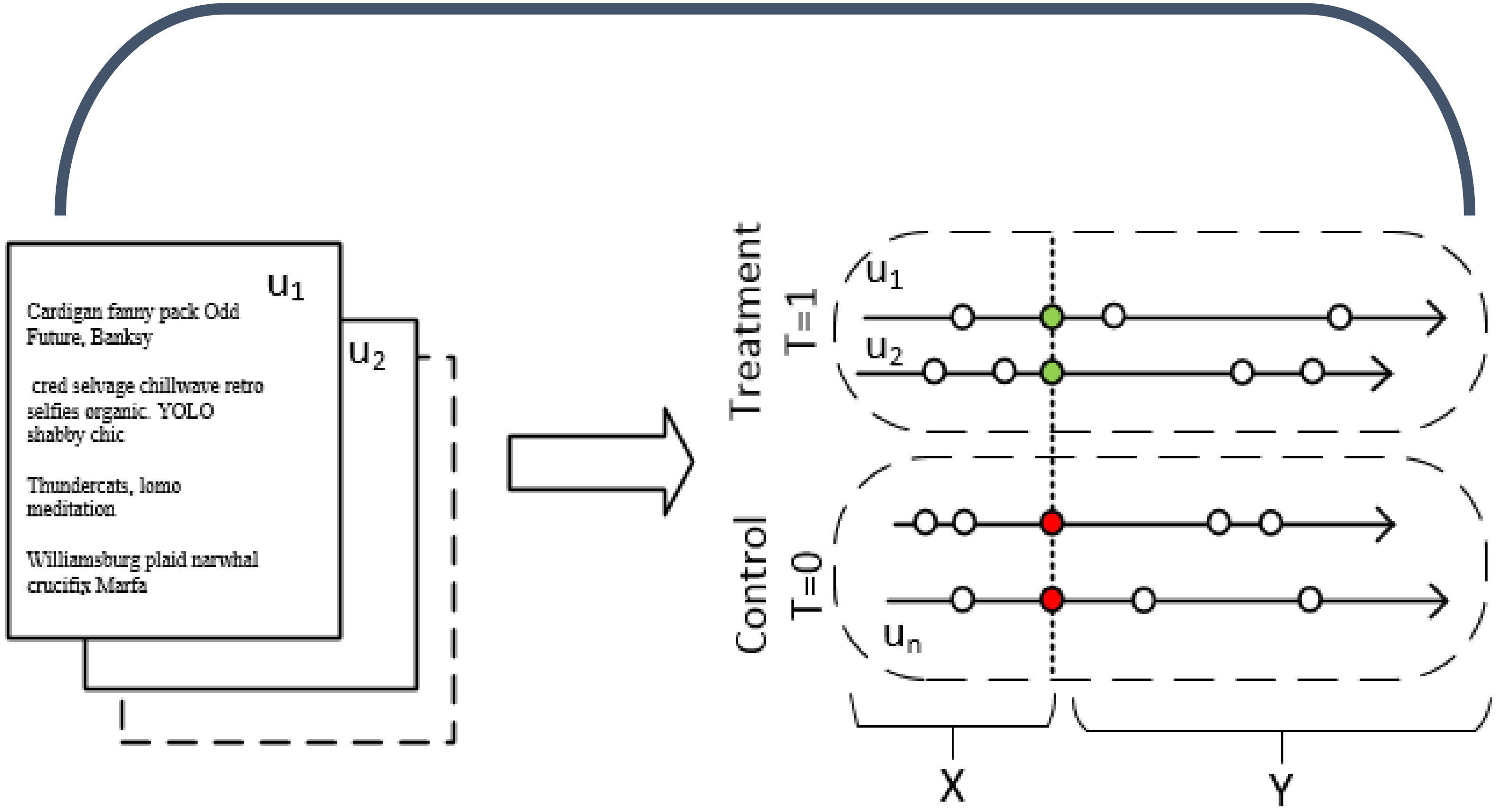
Instrumental Variables

Disentangle X and T through another variable Z.

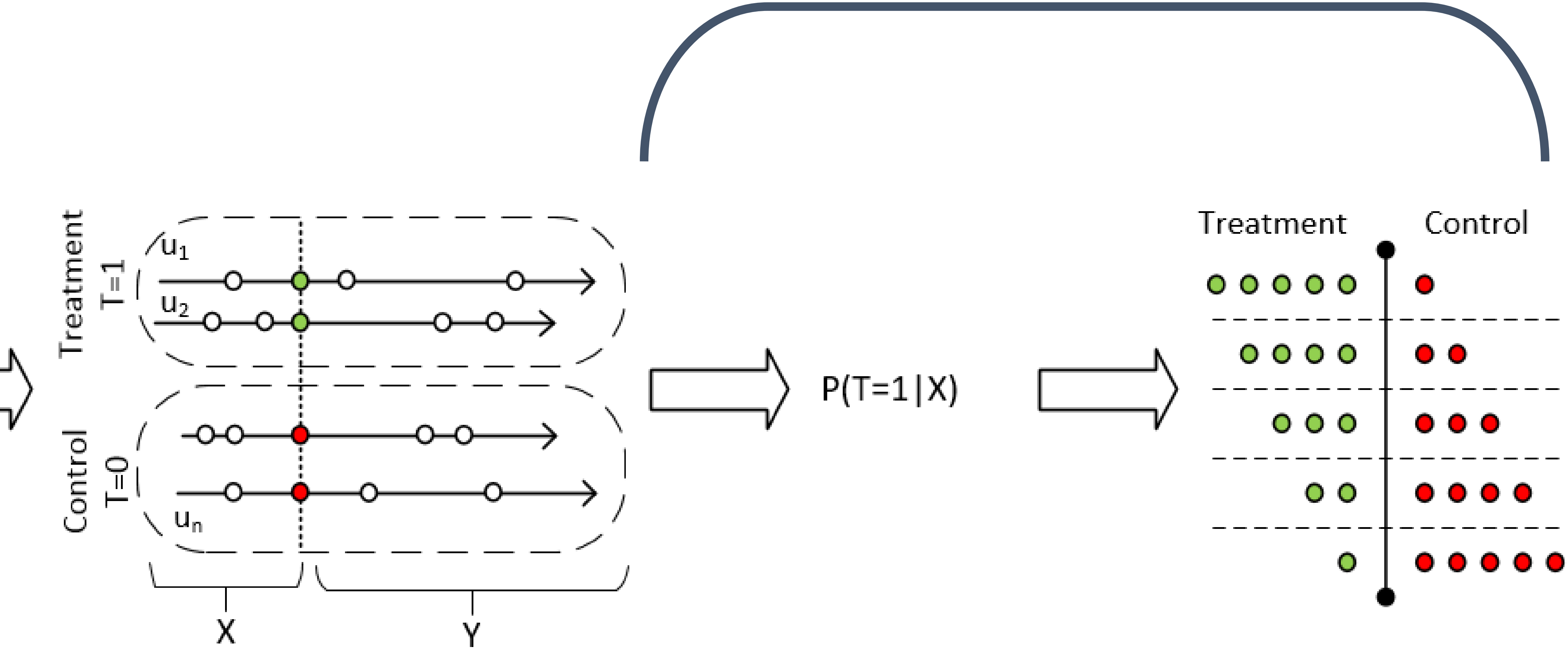


Framing an Analysis over Social Media

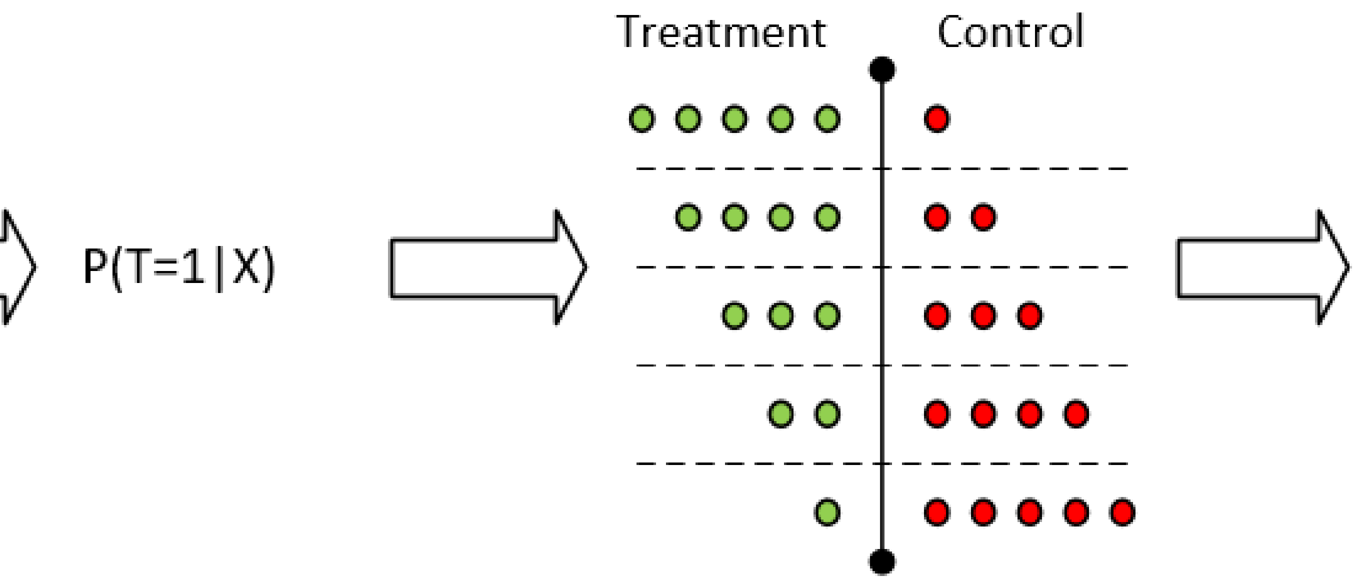
1. Extract timeline of event occurrences from texts; Identify treatment and control groups



2. Learn propensity score estimator and stratify users



3. Calculate population average outcomes



For every observed outcome $y \in Y$, compute $P(y=1|T=1)$ per strata

Setting up an analysis

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

Each has implications and trade-offs for internal and external validity of research outcomes

Case Study: Effects of Alcohol Use During College

With Scott Counts (MSR) and Melissa Gasser (UW)

College is an important transition period

- Success in college predicts career success, career happiness, economic achievement.
  - High rates of college graduates drive regional income levels and other positive macro-economic indicators
- Over 40% of college students leave without earning a degree.
  - Many factors: academic and social integration, financial pressures, …
- Excessive alcohol consumption negatively associated w/ college success, as well as other long-term negative consequences

5-year longitudinal social media analysis

- Existing study methods primarily use surveys.
  - Limited to single institutions and/or small number of participants
  - Rely on participant recall
  - Response biases

- Social media studies can complement
  - Large number of participants
  - In situ reporting of experiences
  - (different) reporting biases
  - Granular observations

What insight are we looking for?

Goal: might intervening to stop early alcohol usage aid college success?

Does early alcohol usage have measurable effects on topics linked to college success?

Does early alcohol usage have measurable effects on future alcohol usage?

5-year longitudinal social media analysis

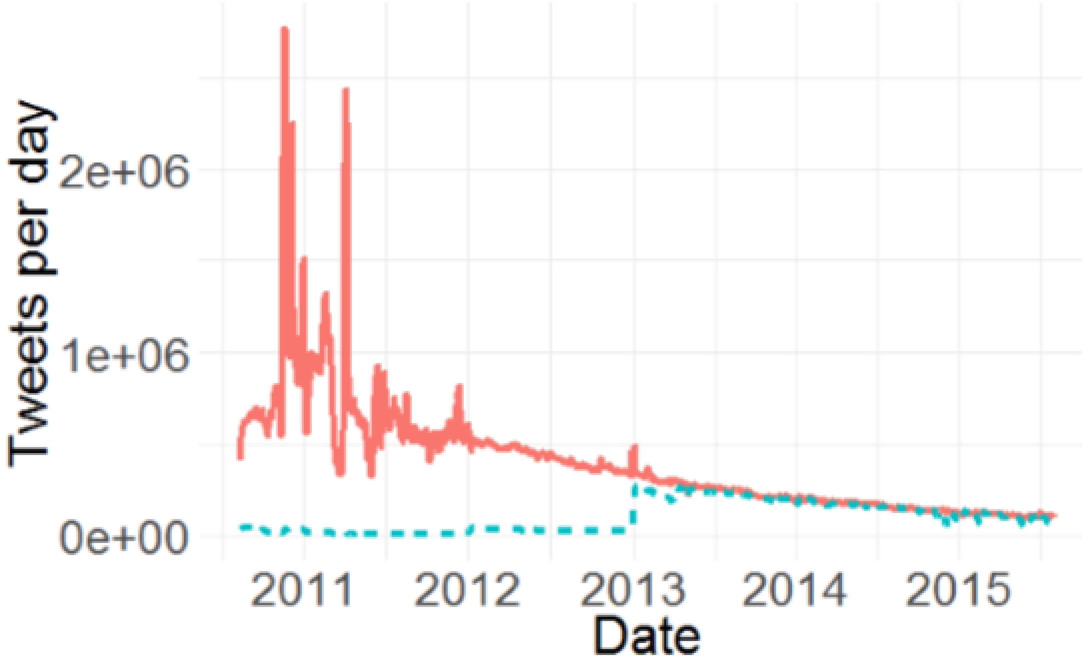
1. **Build a dataset of college students twitter timelines:**
   - Identify twitter accounts of entering college students in Fall 2010
   - Gather their tweets from Fall 2010 through Summer 2015
2. **Identify relevant events and topics:**
   - Drinking and alcohol mentions in Fall 2010
   - Topics known to be related to college success: financial pressures, negative academic outcomes, studying, family, friends, criminal/legal issues.
   - We could not find a simple, reliable indicator of college graduation
3. **Infer effects of early drinking on college-success linked topics**
   - Stratified propensity score analysis.

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

Identifying a college cohort

1. Find all tweets matching a high-recall, low-precision phrases
2. Build and apply high-precision classifier

Keyword Phrase	**Users (Tweets)**	
	Pre-classifier	**Post-classifier**
day of college	22k (26k)	14k (17k)
college tomorrow	23k (37k)	11k (15k)
start college	13k (17k)	6k (7k)
going to college	38k (46k)	5k (6k)
my first college	9k (9k)	5k (5k)
my first semester	10k (10k)	3k (3k)
first semester of	9k (9k)	3k (3k)
college starts	6k (6k)	3k (3k)
week of college	4k (4k)	2k (2k)
Total (over all phrases)	320k (639k)	49k (68k)



1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

Drinking/alcohol mentions

- Previous studies find link between alcohol mentions on social media and real-world alcohol usage
- Identify all tweets that contain a validated list of high-precision alcohol phrases
- Alcohol mention in first semester (fall 2010) will be our treatment
  - Check that fall 2010 window allows for enough covariate data

Table 4: Phrases indicating drinking alcohol

Phrase	**Matched Tweets**	**Distinct Users**
drunk	371011	25318
drinking	182033	32017
beer	138609	18678
wine	128732	18353
drinks	103969	19077
alcohol	91359	17576
vodka	43635	11160
drank	41133	13081
gin	35018	11934
tequila	19169	6715
(other words)	545975	112834
Total	1700643	277933

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

Control group

- All untreated individuals in cohort who do not mention alcohol.
- Assign placebo time to each untreated individual, to match distribution of treatment times of treated individuals.

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

Covariate representation

- Rule of thumb: include all words before treatment event
  - Word counts for top 50k words
- Daily tweet frequency and tweet length statistics
  - Tweet frequency is linked to likelihood of reporting any given experience
- Featurize word counts as word likelihood, not absolute count
  - Increasing tweet frequency post-treatment means absolute counts less meaningful
- Words in days leading up to drinking “give away” treatment status
  - Including these words as covariates means we lose support
  - ➔ We don’t use words in the week before alcohol mention as covariates
  - Trade-off: treatment event is more complicated, includes events leading up to drinking (e.g., talking about parties)

Outcomes representation

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

- Focus on topics linked to college success:
    - Financial pressures, peer and family relationships, study habits, negative academic outcomes, interactions with police/crime.
- Outcome measures are likelihood of topic mentions
    - Using a list of 15-30 related words/topic
- Measured over a week-long sliding window, starting from drinking mention for ~5 years.
- Note: we could not find a reliable indicator of actual college success. Too much variation in success/failure declarations and too few declarations.

Outcomes representation

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

Table 2: Topics linked to college success

Concept	**Seed words**	**Empath expansions**	**Example Tweets**
Peer group interaction	friend, boyfriend, girlfriend	buddy, roommate, bandmate, fiance, *+23 more*	Yup buddy I found my bandmate!!
Family responsibilities	mother, father, brother, sister	stepdad, children, grandmother, *+25 more*	Thankful for my little bro and mom I have a sister #fact
Negative school performance	flunk, fail, miss, skip	flunk, fail, lose, cancel, retake, late, skip, *+5 more*	retake my microecon exam today maybe ima jus flunk
Study habits	study, library, homework	math, tutor, textbooks, worksheets, *+56 more*	anyone that wants to study for history we're in the library but anyways ima off to study
Financial pressures	debt, student loan, loans	wages, afford, utilities, tuition, evicted, fees, *+45 more*	finally my wages wooo @anon its all about money. Im in debt. dont want more loans
Legal/criminal challenges	arrested, police, cops, jail, parole	restraining, agency, authorities, *+65 more*	meeting my parole officer cops pulling out breathalyzer f***k we drunk

Outcomes representation 2

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

- Measure effect on tweet rate
  - Average tweet rate in sliding window

Variant: Difference-in-differences

- Measure the individual-level difference between tweet rate before and after treatment.
  - Helps when initial values of studied feature are not well-balanced

Summary: Details on causal inference setup

- Covariates:
  - Word counts for top 50k words
  - Daily tweet frequency and tweet length statistics
  - Featurize word counts as proportions
  - Don’t use words in the week before alcohol mention as covariates
- Treatment:
  - Marker: Alcohol mention in first semester
  - Semantically: the treatment is everything that happens the week before a drinking mention.
- Outcomes
  - Topics linked to college success
  - Week-long sliding window, starting from drinking mention for ~5 years.

Stratified propensity score analysis

1. Define a cohort
2. Find who was treated
3. Extract covariates
4. Extract outcomes
5. (Run inference algorithm)
6. Analyze effects
7. … and iterate

- Learn a probabilistic classifier: maps from covariates to treatment likelihood
  - Supervised learning: we have all the treatment status labels
  - Logistic regression, SVM, … all reasonable choices.
- Stratify into 100 strata, drop strata with insufficient support
  - Report distribution of treatment and control individuals across strata.
  - Report population covered by strata with sufficient support

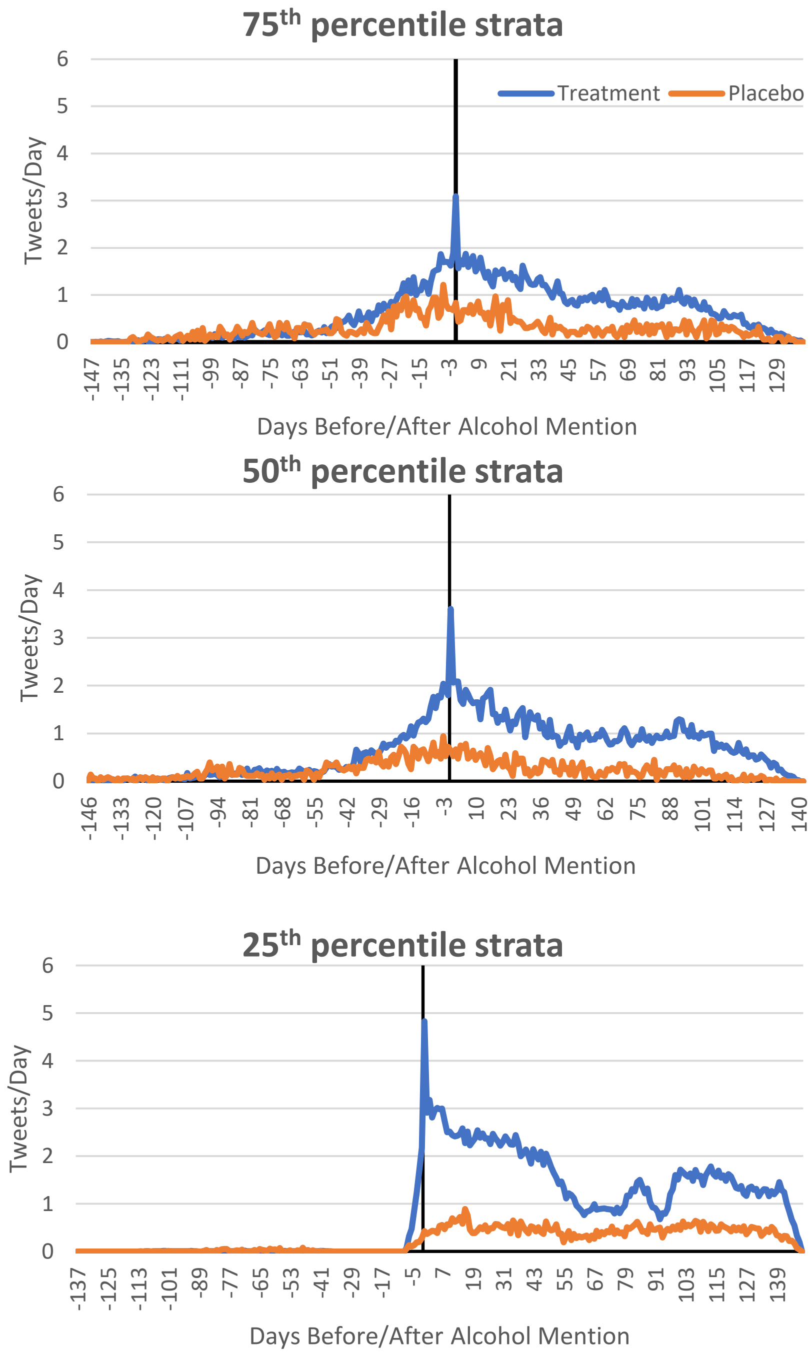
High-level results

- Increase in tweet rates after drinking mentions
- Mentioning drinking indicates higher rates of alcohol mentions for ~next 2 years as compared to control.
- Drinking mentions has 6mo-2yr effect on most college-success linked topics; with a longer effect on study habits and friend mentions.

Increase in tweet rate

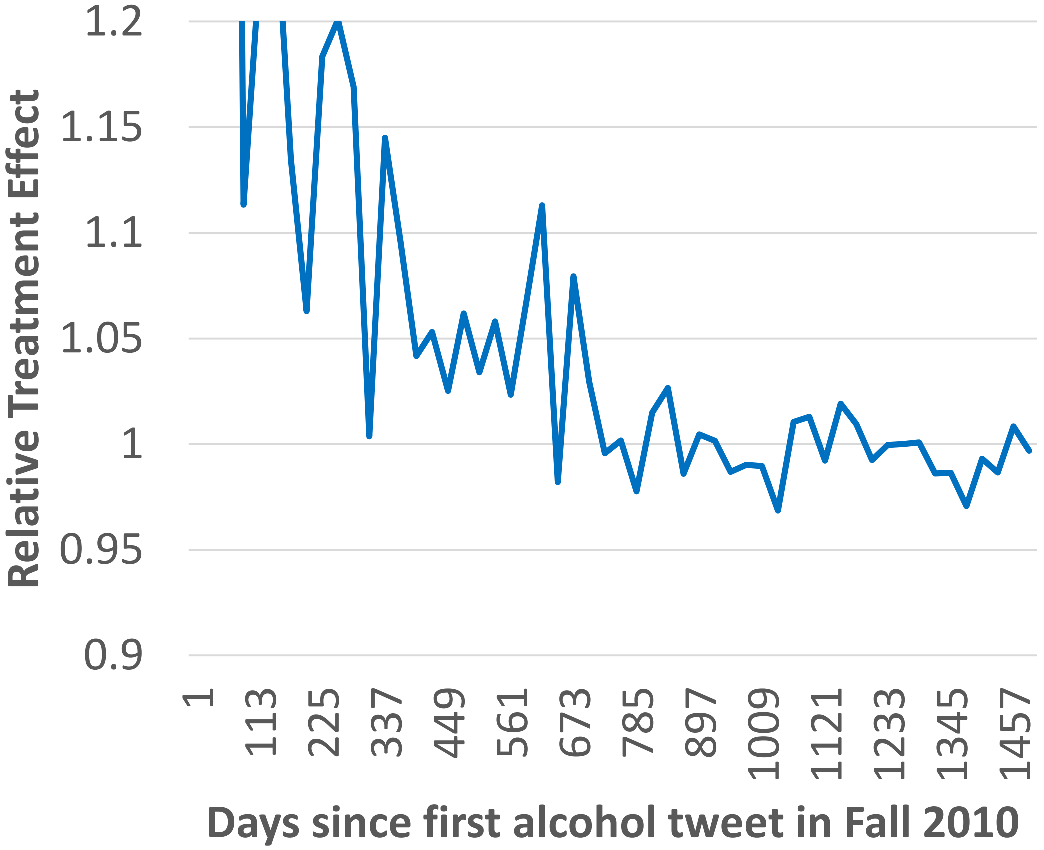
- Before drinking, tweet rates are approximately balanced
- After drinking, tweet rates increase by 0.98 tweets/day
- Repeated with a difference-in-differences analysis, with effect persisting (~1 tweet/day)

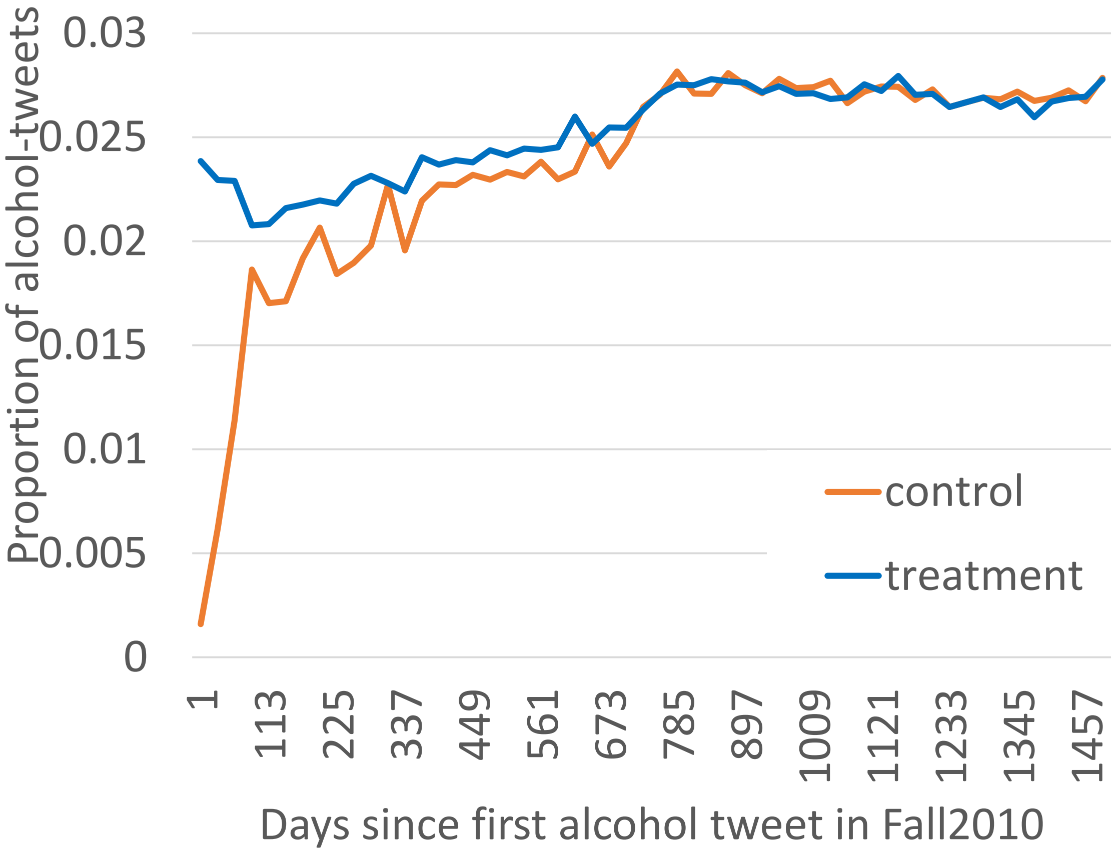
- Minor implications for analysis: we have to represent words and tokens as proportions, not counts



Effect on future drinking

Effect on future drinking



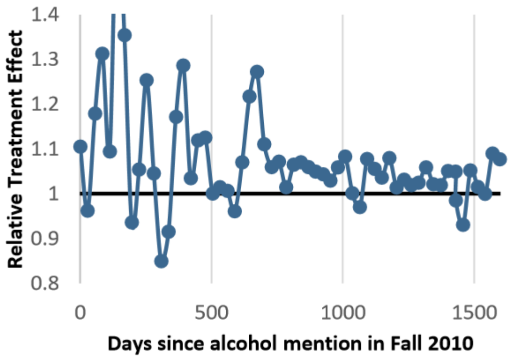


Relative effect shows early drinking leads to more drinking, with diminishing effect over time. This is because the non-drinkers "catch up", increasing the drinking mentions over time.

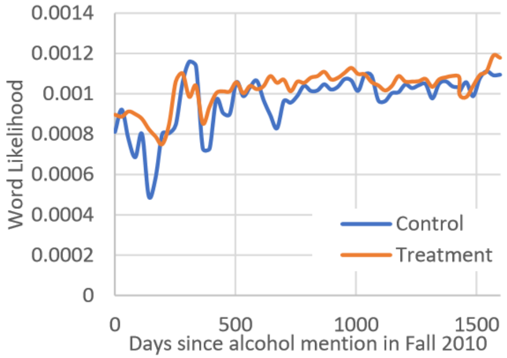
Effect on topics related to college success

Both control and treatment users follow similar temporal patterns.

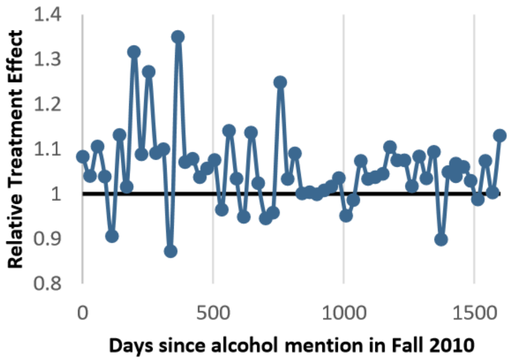
But generally see positive effect on each outcome initially, with diminishing effect over time.



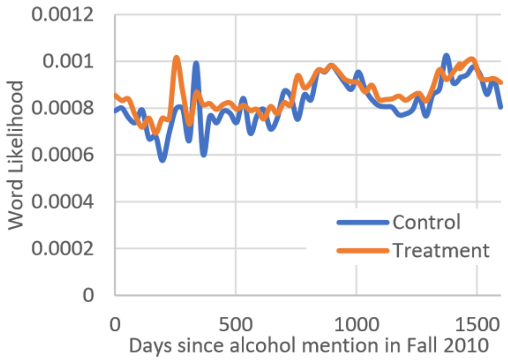
(a) Financial RTE



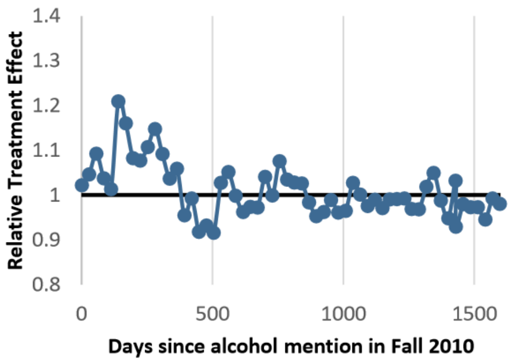
(b) Financial Outcomes



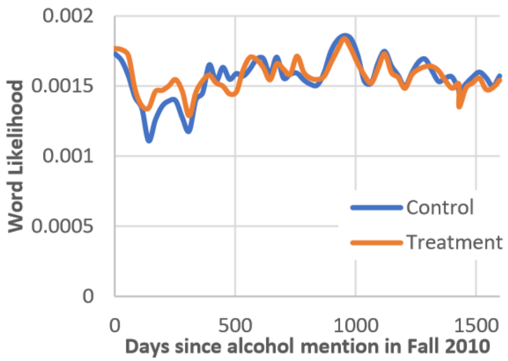
(c) Criminal RTE



(d) Criminal Outcomes



(e) Family RTE



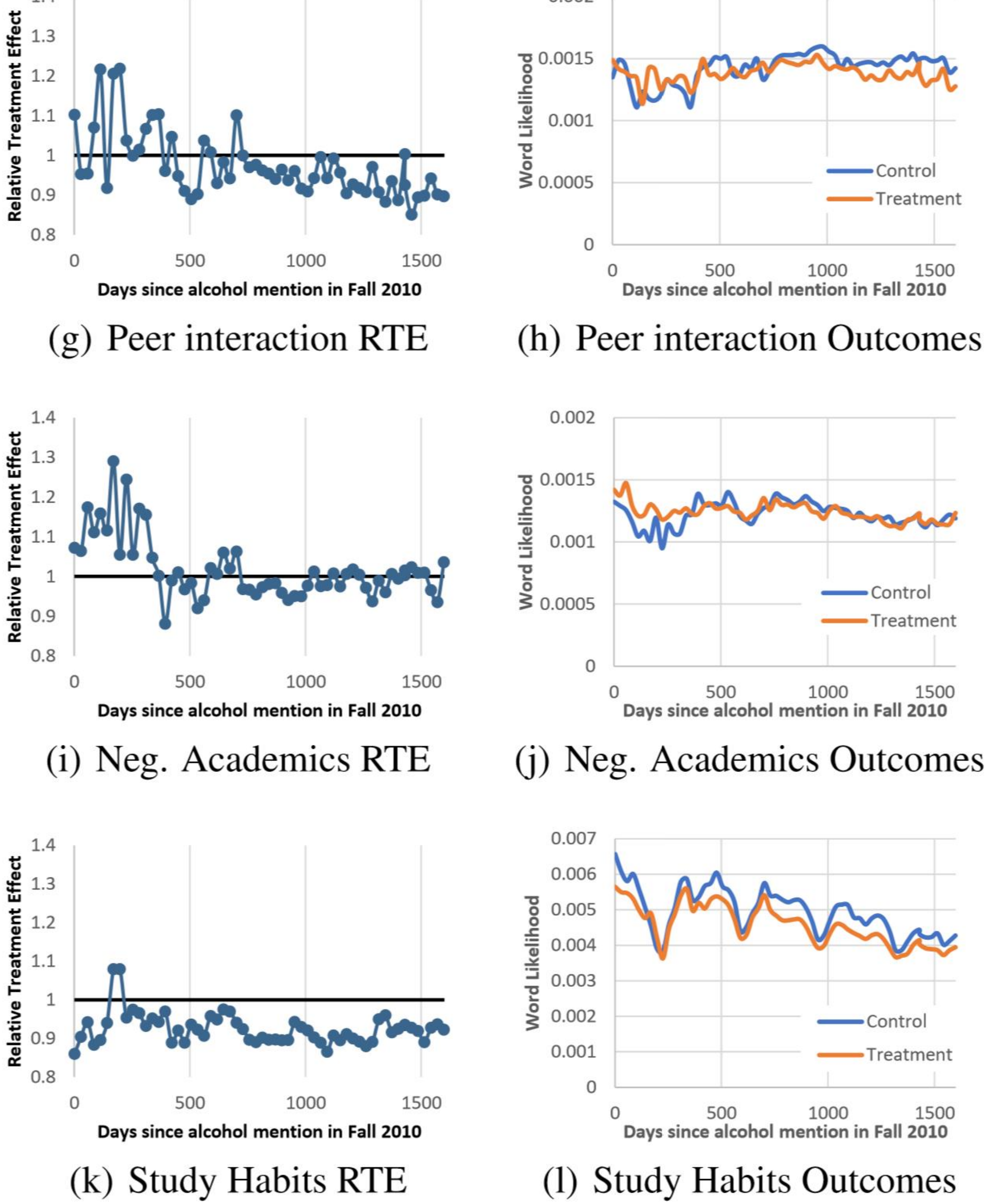
(f) Family Outcomes

Effect on topics related to college success

Initially, drinkers are more social, but after about 1-1.5 years, drinkers mention peers less than control group.

Strong effect on negative academic outcome mentions in year, then no difference.

Persistently lower proportion of study habit mentions.



(g) Peer interaction RTE
(h) Peer interaction Outcomes
(i) Neg. Academics RTE
(j) Neg. Academics Outcomes
(k) Study Habits RTE
(l) Study Habits Outcomes

Case study findings & insights

Q: Does early alcohol usage have measurable effects on topics linked to college success?

A: Yes, especially in short-term (6mo-2yrs). However, effects might be due to other factors closely associated with drinking (e.g., socialization at parties)

Q: Does early alcohol usage have measurable effects on future alcohol usage?

A: Not long term. In fact, control group catches up over time. Note, this does not account for frequency of drinking.

Goal: might intervening to stop early alcohol usage aid college success?

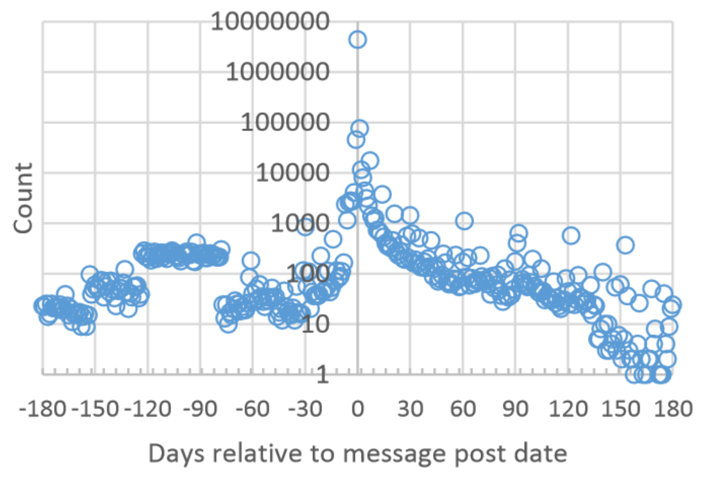
Additional Notes

Be aware

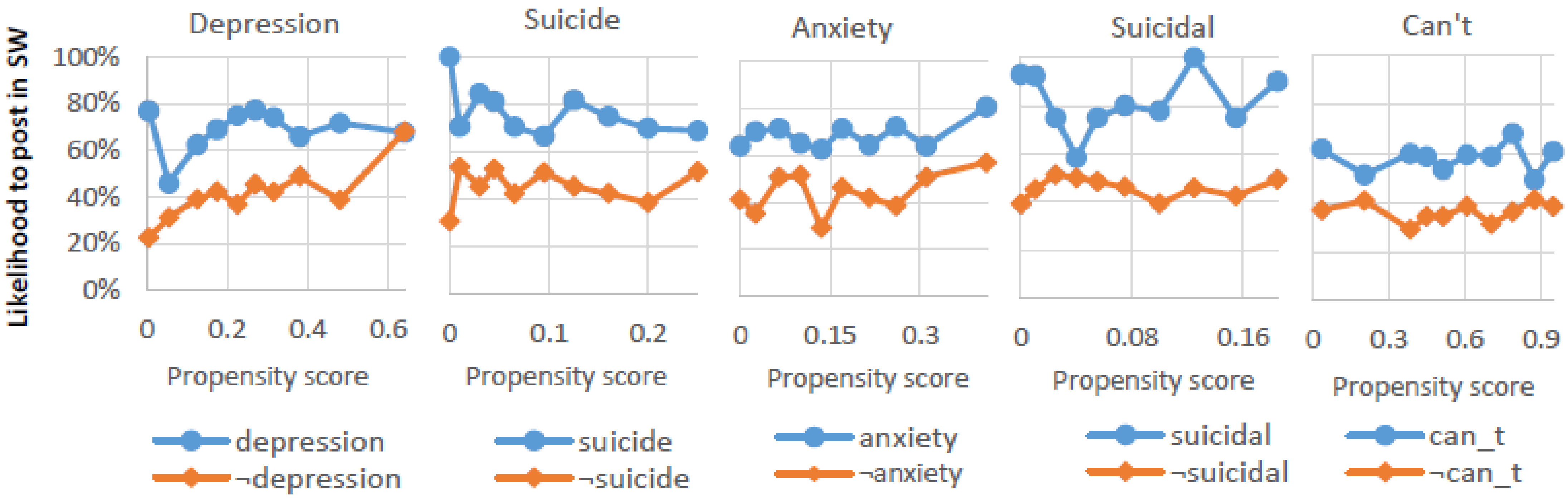
- In social media: “conversational space” is not “real world”
- Heterogeneous effects: treatment doesn’t affect everyone the same
- Balance tests
- Importance of qualitative analysis
  - Do the features mean what you think they mean?

Awareness: Conversations & real-world experiences

1. Bots (Assuming we've already removed bots and organizations)
2. Consider focusing on experiential messages
   - ~25% of tweets are experiential.
   - Non-experiential: Discussions of news, chit-chat, etc.
   - Choice depends on research question
3. People do not always mention experiences in order
   - Implies causal relationships may be reversed
   - E.g., people more likely to mention disease after mentioning drug treatment



Heterogeneous effects



Even opposite effects across strata

	Increase likelihood of MH → SW			*Decrease likelihood of MH → SW*		
treat. token	propensity (strata #)	effect	distinguishing historical words	propensity (strata #)	effect	distinguishing historical words
baby	0.03-0.04(5)	+44%	me_on, problem, mind, week, was_in	0.00-0.01(2)	-39%	wait, okay, thanks, re, her
have_sex	0.01-0.02(4)	+43%	instead_of, months, isn, well, a_girl	0.03-0.05(6)	-53%	months, dating, isn, such, there
medication	0.06-0.09(6)	+46%	seem_to, its, back, myself_and, seem	0.02-0.03(3)	-51%	but_then, not_to, d, i_never, everything
own	0.30-0.36(7)	+37%	all_i, can_be, we, of_this, this	0.45-0.59(9)	-32%	stop, my_own, yourself, you_can, needs
relationship_with	0.00-0.01(2)	+44%	spent, finally, my_mind, etc, think_it	0.13-0.17(9)	-36%	had_a, good, i_as, m_not, do_i
stressed	0.08-0.09(9)	+44%	i_do, as, i_hate, by, i_m	0.02-0.04(3)	-33%	there_and, but_they, deal_with, ve_had, took
upset	0.13-0.16(8)	+58%	but_i, some, a, haven_t, we	0.08-0.11(6)	-28%	living, is_this, with_the, i_started, sorry

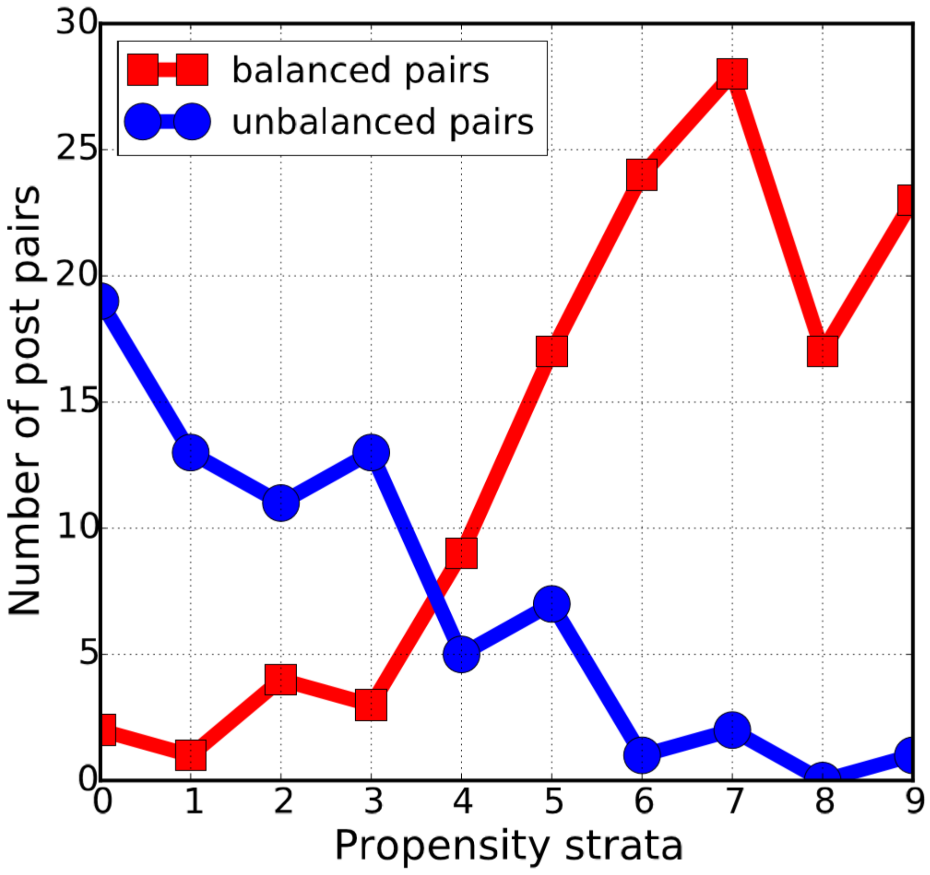
Are confounding variables balanced across compared treatment and control groups?

Are confounding variables balanced?

- Many ways to test balance, harder in high
- In higher dimensions, approximate balance can be sufficient to get bounded error.

Manual judgements of balance in stratified posts from a mental health forum.

- Generally, people are poor at judging balance.
- Here, treatment status tied to human reading, thus human judgement of similarity relevant.
- Found imbalances at lower strata!



Importance of qualitative analysis

- Feature extraction is a serious threat to experiment validity
  - Words can have multiple meanings, ambiguities
  - Learned classifiers can have systematic errors
- Unlike some other data domains, we can read experiential messages for validation and mechanism clues

Table 4: Paired treatment and outcome messages for selected users, carefully paraphrased for anonymity.

Treatment	Example tweet	Outcome	Example tweet
Dealing with jealousy issues	*ironically, u ask why I have jealousy issues*	wake up	*@user I need u to wake up because im bored*
Suffering from depression	*if u think depression is eccentric or cute u can have mine bc i dont wanna deal with it*	thoughts	*hate small talks, dont talk abt weather, tell me what keeps u up at night, ur thoughts abt dying*
Suffering from depression	*if u think depression is eccentric or cute u can have mine bc i dont wanna deal with it*	self harm	*@user dont self harm, remember yr worth so much better, u dont deserve this pain, stay safe*
Suffering from anxiety	*if hadnt spent years dealing with anxiety, I wouldnt have my sense of humor*	yelling	*dont have any anger issues at all, im really happy when yelling at people*
Paying credit	*seriously, my soul was deep hurt when I*	apartment	*Im checking some apartments in NYC lol*

Importance of qualitative analysis

- Feature extraction is a serious threat to experiment validity
  - Words can have multiple meanings, ambiguities
  - Learned classifiers can have systematic errors
- Unlike some other data domains, we can read experiential messages for validation and mechanism clues

Example:

- taking “Xanax” increases “getting drunk”, “smoking weed”

Not a panacea:

- One drug increases swear words. Looks like noise, but turns out irritability is a known side-effect

Conclusions

1. Estimating causal relationships among experience reports on social media is a rich and promising approach to understanding broad set of phenomenon

2. Causal inference reduces some kinds of bias

   But, still need to worry about measurement validity and generalizability.

   Achieve through separate validation, and/or expt design & scoping of goals

Questions?

emrek@microsoft.com; @emrek

Referenced papers:

- Towards Decision Support and Goal Achievement: Identifying Action-Outcome Relationships from Social Media. Kıcıman, Richardson. KDD15
- Shifts to Suicidal Ideation from Mental Health Content in Social Media. De Choudhury, Kıcıman, Dredze, Coppersmith, Kumar. CHI16
- Distilling the Outcomes of Personal Experiences: A Propensity-scored Analysis of Social Media. Olteanu, Varol, Kıcıman. CSCW17
- The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk, De Choudhury, Kıcıman ICWSM17
- Using Social Media to Understand the Effects of Alcohol Use During College. Kıcıman, Counts, Gasser (email for draft)
- Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Olteanu, Castillo, Diaz and Kıcıman. Working paper