Longitudinal analysis of social and behavioral determinants of health in the EHR: exploring the impact of patient trajectories and documentation practices

Daniel Feller, Jason Zucker, Oliver Bear Don't Walk, Michael Yin, Peter Gordon, Noemie Elhadad

S80: Social Determinants of Health



Disclosure

• I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Introduction

 Social and behavioral determinants (SBDH) are non-medical factors that impede disease self-management and exacerbate existing comorbid conditions

 Targeted HIV and STI prevention strategies have created opportunities for a focus on the SBDH of sexual health

 Retrieval of SBDH a common source of frustration among nurses and social workers. (Weir 2015)





Background

- Healthcare providers predominantly document determinants of sexual health in free-text notes (Walsh 2014, Chen 2011)
- Previous research has focused on using NLP to extract a broad range of SBDH (Yestigen 2017, Bejan 2017, Dorr 2019)
 - These approaches 1) consider SBDH at a single time point and 2) treat EHR data as source of truth

Problem: Unclear whether SBDH related to sexual health change in the patient record, nor the expected frequency of such changes

To ensure that information extraction for determinants of sexual health are accurate

Research Questions

- 1. Does patient SBDH status change through time?
- 2. What is the incidence of changes in patient SBDH status across subsequent medical encounters?
- 3. Are there potential data quality issues with SBDH documentation?

Corpus Characteristics

- We curated a gold-standard corpus of notes with SBDH documentation. (Feller 2018)
- Annotators read the entire length of each note
 - Notes labeled for the presence of 29 social SBDH risk factors
 - 3 SBDH status types: 'Never', 'Historical', and 'Current'
 - 76 notes were double annotated (Kappa 0.736 across all SBDH)
 - 327 distinct note types were included



Corpus Characteristics

- Inclusion Criteria #1
 - 167 (of 1064) individuals had 3 or more annotated notes
- Inclusion Criteria #2
 - 108 individuals also had a longitudinal history of >180 days
 - Median History: 772 days (Mean 997)
- Cohort Characteristics
 - 398 patients 70% male
 - Average age 50 (std. 12 yrs)
 - 8% had HIV (32/398)



Question #1 Do SBDH change over time?

Encounter:



N = 84 patients

```
Question #1
Do SBDH change over time?
```

Encounter:



N = 92 patients

Question #1 Do SBDH change over time?



Encounter:

N = 76 patients

Question #1 Do SBDH change over time?

Percentage of patients who transitioned from one SBDH state to another

Alcohol Use	28%		
Housing Status	25%		
Drug Use	20%	Note 1 Note 2	Note 3
Condom Use	15%	\mathbf{O}	U
History of STIs	13%		
Sexual Orientation	4%		

Question #2 What is the timescale of changes in SBDH?

Median

defines the time point where 1/2 of the population has experienced a SBDH change

Drugs: 571 days Housing: 464 days Alcohol: 446 days Sexual Orientation: N/A



Question #3 Do changes reflect data quality issues?

Yes; common non-sensical transitions



Question #3 Do changes reflect data quality issues?

Yes; among multiple annotated notes observed within the same day or same week, there are frequent changes

Intra-Day Changes		Intra-Week Changes	
Alcohol Use	25%	Alcohol Use	32%
Drug Use	24%	Drug Use	24%
Housing Status	13%	Housing Status	16%

Future Research

- 1. Research is needed to examine whether changes in SBDH documentation reflect true changes in the patient state versus documentation errors.
 - This could be accomplished by obtaining additional data like validated psychometric instruments or laboratory tests for substance abuse.
- Computational approaches should assess whether using multiple data sources within the EHR can overcome the challenges that data quality issues pose for automated inference of SBDH

Conclusions

Information extraction approaches should not treat SBDH as static

• Each patient's SBDH status must be reevaluated on a regular basis

Data quality issues were frequently observed and must be accounted for by information extraction approaches

 Aggregating multiple data points may provide a more accurate representation of a patient's SBDH status



Thank you!

Email me at: djf2150@columbia.edu

Funded by National Library of Medicine (T15 LM007079) and National Institute of General Medical Sciences (R01 GM114355)



Related Work:



Previous approaches for extracting SBDH

- Healthcare providers predominantly document determinants of sexual health in free-text notes (Walsh 2014, Chen 2011)
- NLP has been successfully applied to extract a broad range of SBDH (
 - Smoking status (Uzuner 2008; 1st i2b2 NLP shared task)
 - Substance Abuse (Yestigen 2017, Melton 2015, Chen 2015, Carrell 2015)
 - Adverse Childhood Events (Bejan 2017)
 - Social support (South 2017, Dorr 2019)
 - Homelessness (Melton 2016, Gundlapalli 2015, Oreskovic 2017, Bejan 2017)
 - Chronic Stress (Dorr 2019)
 - Gender Identity (Romano 2017, Roblin 2016)