Strategies to Address Residual/Unmeasured Confounding in VA Data

Francesca Cunningham, Pharm.D.
Associate Chief Consultant, PBM
Director, Center for Medication Safety (VAMedSAFE)
**Department of Veterans Affairs**

Review

- VA Healthcare System
- Databases Used for Surveillance in VA
- Other Data in VA
- Strategies /Tools to Mine Data

VA HEALTH CARE SYSTEM

- U.S. Veteran Population: Over 23 M
- Largest U.S. integrated health care system
  - Over 9 million enrollees
    - Approx. 6.5 M obtain health care
    - > 5 M obtain prescriptions
- Veteran Population (approx)
  - White – 83 %
  - Black Non–Hispanic –12%
  - Hispanic – 6 %
  - Women 10%

VA HEALTH CARE SYSTEM

- VA Medical Centers: 153
  - Affiliations with 107 medical schools
- VA Outpatient Clinics: Over 800
- VA Patients
  - Complex healthcare needs
    - Multiple co-morbidities
    - Mental health problems
    - Physically disabled
    - Socio-economic
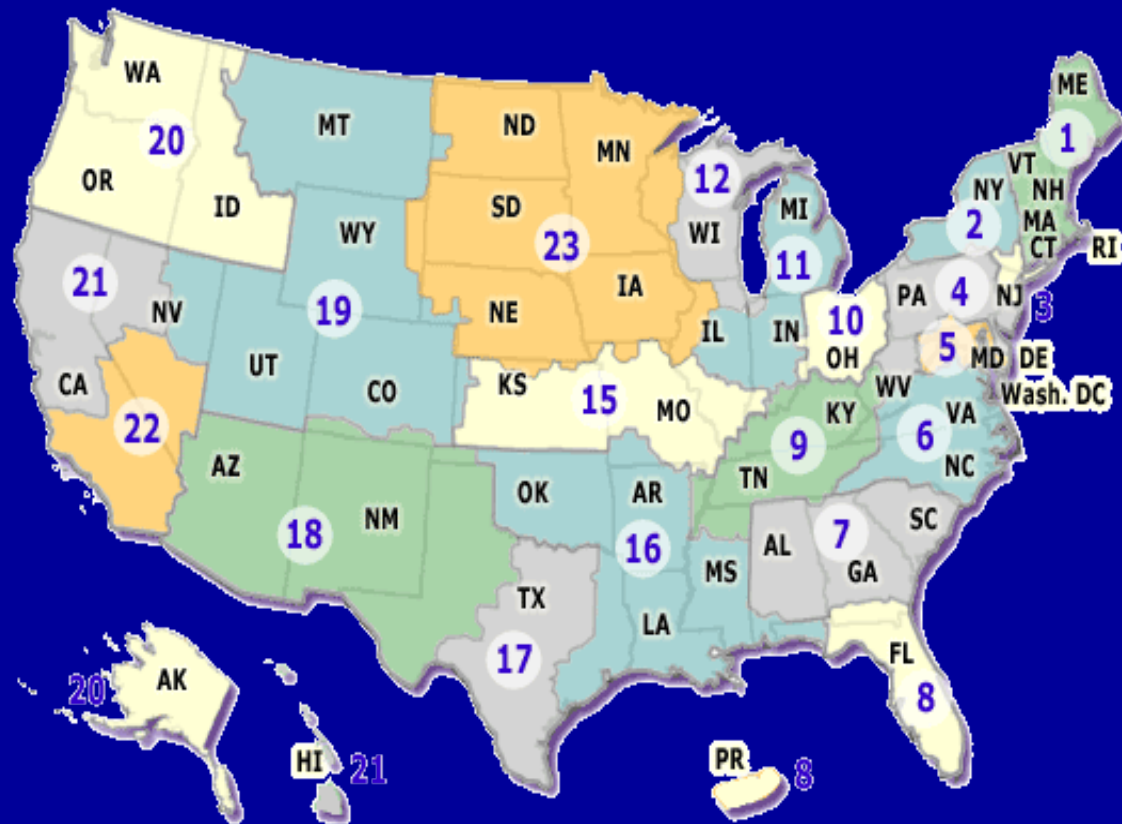
***Over 150 medical facilities
Currently distributed among 21 VISNS***

VA Databases

- Patient Care Databases
  - MedSAS data sets
- Vital Status File
  - Mortality
- Prescription Databases
  - Outpatient Rx
  - Inpatient Rx

Common Data Elements

- Demographics
- ICD–9 CM
- Procedure
  - ICD9–CM
  - CPT4
- Prescriptions
- Laboratory
- Vitals

Other Data Elements

- Income
- Marital Status
- Education
- Priority Status
- Other
  - Theater Exposure

Other Data Elements

- Priority Status
  - Comprehensive care
  - Copay status
  - Access to Care
- Theater of War
  - Exposure to certain elements increase risk
    - Agent Orange –Vietnam
    - Toxins –Gulf War
    - IEDs – OEF/OIF/OND

Availability of Data to Address Residual Confounding

- VA National Survey Data
- Registries
- Corporate Data Warehouse
  - Health Factors
  - Chart Notes

Health Factors

- Health Factors
  - Physicians responses to clinical reminders
  - Multiple items
- Access
  - Local
  - Regional
  - National
- Example of Data
  - Outside Labs
  - Smoking
- Limitations
  - Many elements
  - Not standardized

Example of Health Factor Data

FACTOR
621 POSITIVE TOBACCO SCREENING
A CURRENT TOBACCO USER
AAA LIFETIME NON-SMOKER
ACS TOBACCO YES
ACTIVATE TOBACCO COUNSELING
ACTIVATE TOBACCO EDUCATION
ACTIVE SMOKER (TIDES)
ADVISED OF TOBACCO CESSATION STRATEGIES
BH WARD TOBACCO CESSATION GROUP
CARDIO SMOKER
CDM COPD PT GOAL TOBACCO CESS ESTAB
CDM COPD TOBACCO CURRENT USER
CDM COPD TOBACCO NON-USER
CHEWING TOBACCO USER
CHEWS TOBACCO
CIGAR SMOKER
CO-NO TOBACCO USE PAST 12 MO

How to Address Limitations

- Develop algorithm to standardize data
- Labor Intensive
  - Prior Example
    - 966 different terms
    - Standardized – 3 categories

Standardized Factors

	healthfactortype	sta3n	healthfactordatetime	healthfactorcategory	VISN	status
1	TOBACCO OFFER CESSATION MEDS	999	2012-10-19 14:30:00	REMINDER FACTORS	99	CURRENT SMOKER
2	TOBACCO OFFER CESSATION PROGRAM	999	2012-10-19 14:30:00	REMINDER FACTORS	99	CURRENT SMOKER
3	TOBACCO OFFER CESSATION MEDS	999	2013-03-22 08:30:00	REMINDER FACTORS	99	CURRENT SMOKER
4	TOBACCO OFFER CESSATION PROGRAM	999	2013-03-22 08:30:00	REMINDER FACTORS	99	CURRENT SMOKER
5	TOBACCO OFFER CESSATION PROGRAM	999	2012-10-23 09:00:00	REMINDER FACTORS	99	CURRENT SMOKER
6	TOBACCO OFFER CESSATION MEDS	999	2012-10-23 09:00:00	REMINDER FACTORS	99	CURRENT SMOKER
7	TOBACCO OFFER CESSATION MEDS	999	2013-02-15 14:30:00	REMINDER FACTORS	99	CURRENT SMOKER
8	TOBACCO OFFER CESSATION PROGRAM	999	2013-02-15 14:30:00	REMINDER FACTORS	99	CURRENT SMOKER
9	TOBACCO OFFER CESSATION MEDS	999	2013-03-05 16:00:00	REMINDER FACTORS	99	CURRENT SMOKER
10	TOBACCO OFFER CESSATION PROGRAM	999	2013-03-05 16:00:00	REMINDER FACTORS	99	CURRENT SMOKER

Standardized Factors

	healthfactortype	sta3n	healthfactordatetime	healthfactorcategory	VISN	status
1	TOBACCO OFFER CESSATION PROGRAM	999	2013-05-30 08:00:00	REMINDER FACTORS	99	CURRENT SMOKER
2	LIFETIME NON TOBACCO USER	999	2012-10-22 10:00:00	TOBACCO	99	NEVER SMOKER
3	LIFETIME NON TOBACCO USER	999	2013-07-01 11:30:00	TOBACCO	99	NEVER SMOKER
4	LIFETIME NON TOBACCO USER	999	2012-11-26 09:30:00	TOBACCO	99	NEVER SMOKER
5	LIFETIME NON TOBACCO USER	999	2013-05-06 11:30:00	TOBACCO	99	NEVER SMOKER
6	LIFETIME NON TOBACCO USER	999	2013-08-05 11:00:00	TOBACCO	99	NEVER SMOKER
7	LIFETIME NON TOBACCO USER	999	2012-12-03 10:30:00	TOBACCO	99	NEVER SMOKER
8	LIFETIME NON TOBACCO USER	999	2013-05-20 11:00:00	TOBACCO	99	NEVER SMOKER
9	LIFETIME NON TOBACCO USER	999	2013-03-04 11:30:00	TOBACCO	99	NEVER SMOKER
10	LIFETIME NON TOBACCO USER	999	2013-03-07 14:30:00	TOBACCO	99	NEVER SMOKER
11	PREVIOUS TOBACCO USER	999	2013-05-16 14:00:00	TOBACCO	99	FORMER SMOKER
12	LIFETIME NON TOBACCO USER	999	2012-12-21 08:30:00	TOBACCO	99	NEVER SMOKER
13	PREVIOUS TOBACCO USER	999	2013-06-11 10:30:00	TOBACCO	99	FORMER SMOKER
14	PREVIOUS TOBACCO USER	999	2013-03-05 11:00:00	TOBACCO	99	FORMER SMOKER
15	PREVIOUS TOBACCO USER	999	2013-03-19 10:00:00	TOBACCO	99	FORMER SMOKER
16	TOBACCO OFFER CESSATION PROGRAM	999	2013-01-14 09:00:00	REMINDER FACTORS	99	CURRENT SMOKER
17	TOBACCO OFFER CESSATION MEDS	999	2013-01-14 09:00:00	REMINDER FACTORS	99	CURRENT SMOKER
18	TOBACCO OFFER CESSATION MEDS	999	2013-05-30 08:00:00	REMINDER FACTORS	99	CURRENT SMOKER
19	DAY TOBACCO LIFELONG NON-USER	999	2013-06-20 15:00:00	TOBACCO REMINDERS	99	NEVER SMOKER
20	QUIT TOBACCO <12 MONTHS AGO	999	2013-04-11 14:10:00	TOBACCO	99	FORMER SMOKER
21	LIFETIME NON TOBACCO USER	999	2013-04-25 19:53:00	TOBACCO	99	NEVER SMOKER
22	QUIT TOBACCO <12 MONTHS AGO	999	2013-01-04 11:00:00	TOBACCO	99	FORMER SMOKER
23	QUIT TOBACCO <12 MONTHS AGO	999	2013-04-15 09:00:00	TOBACCO	99	FORMER SMOKER
24	QUIT TOBACCO <12 MONTHS AGO	999	2013-06-25 13:00:00	TOBACCO	99	FORMER SMOKER

Chart
Use of Natural Language Processing

- Uses
  - Enhance data extraction
  - Validation
  - Surveillance
- Tools
  - Off the Shelf
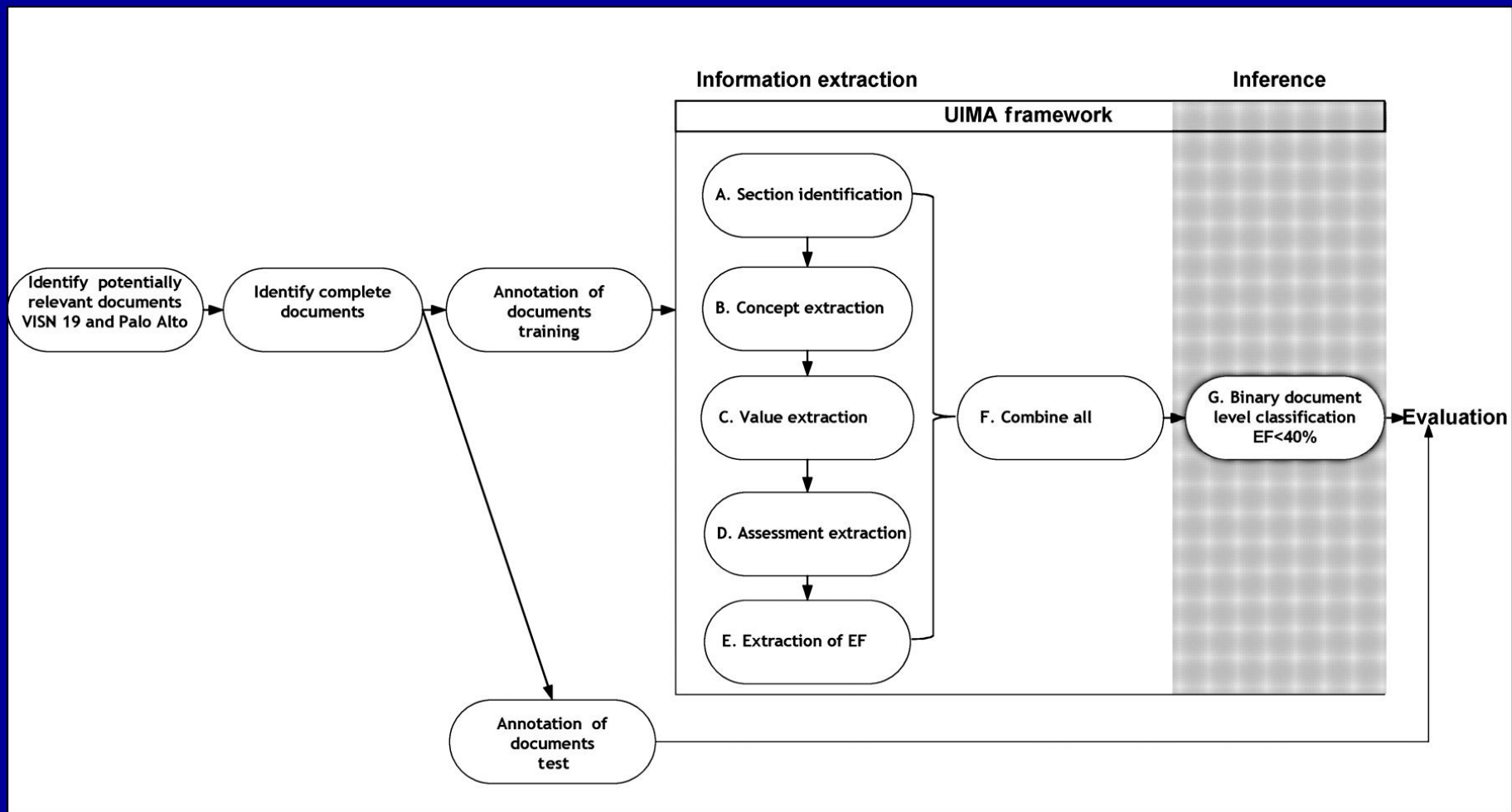 - Internally Developed
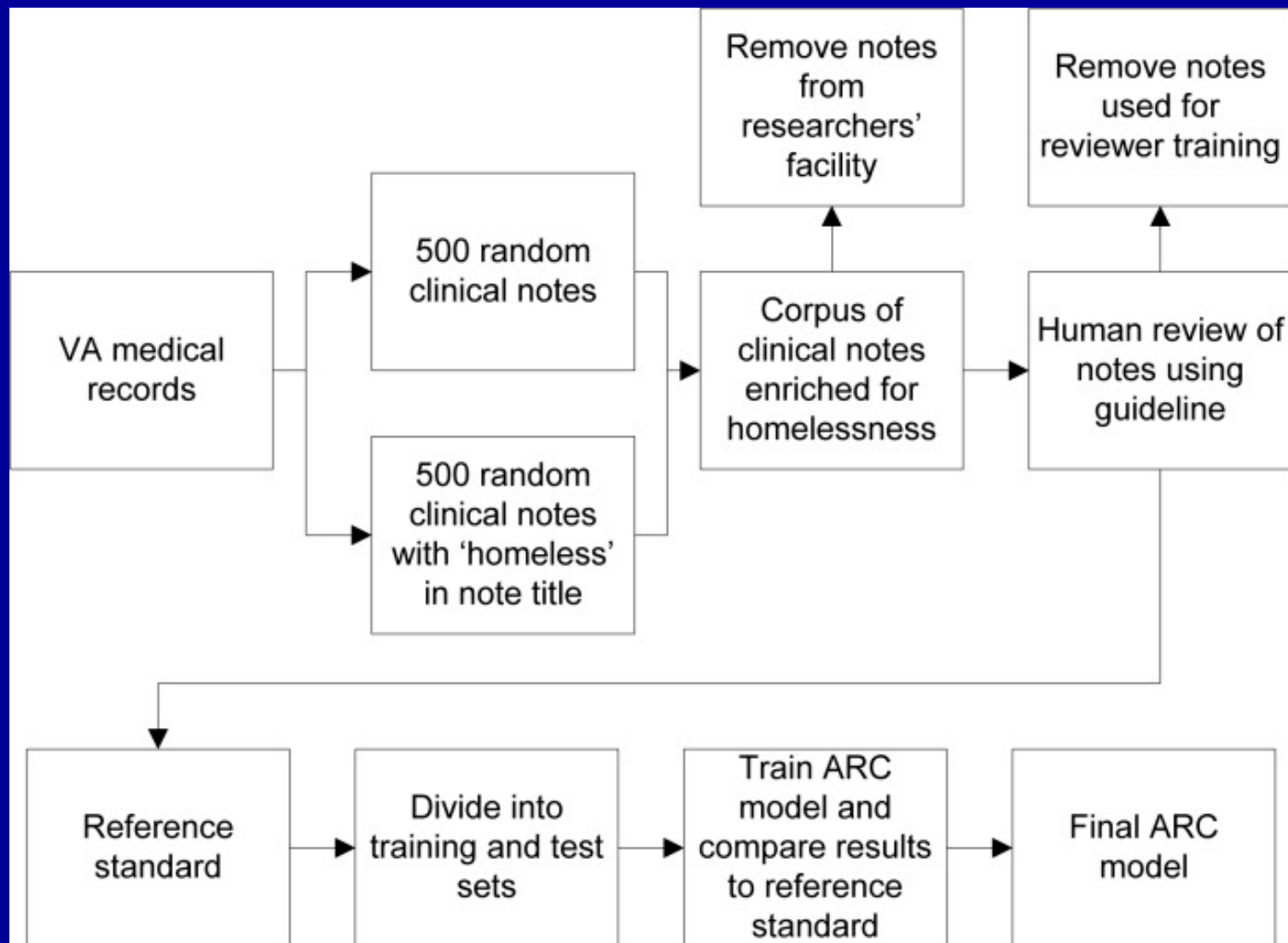
NLP Use

- History
  - Other diagnosis
- Social History
  - Smoking status
  - Alcohol use
- Diagnosis
  - Confirm diagnosis
  - Categorize severity
- Medication
  - Confirm medications
  - Other (OTC meds)

NLP Tools

- Tools can be used to identify diagnosis or assist in severity assessment
  - Characterizing severity of HF via ejection fraction
- Tools can be used to identify variables for unmeasured/residual confounding
  - Smoking or alcohol use
- Tools can be used to identify homelessness
- Other

Information extraction

Inference

UIMA framework

Identify potentially relevant documents VISN 19 and Palo Alto → Identify complete documents → Annotation of documents training → A. Section identification → B. Concept extraction → C. Value extraction → D. Assessment extraction → E. Extraction of EF → F. Combine all → G. Binary document level classification EF<40% → Evaluation

Annotation of documents test

From: Garvin et. al. J Am Med Inform Assoc 2012;19:859–866.

From:Gundlapalli et al. AMIA Annu Symp Proc 2013:537–546

Summary

- VA databases are rich but other factors pertinent to Veteran patients should also be assessed for certain studies

- Important to obtain external data to account for potential residual confounding in certain studies

- Health Factors provide a good source for extra variables

- NLP – excellent tool for extracting needed data

QUESTIONS