# Using Natural Language Processing on the Free Text of Clinical Documents to Screen for Evidence of Homelessness Among US Veterans

**Adi V. Gundlapalli, MD, PhD, MS[1,2], Marjorie E. Carter, MSPH[1,2], Miland Palmer, MPH[1,2], Thomas Ginter, MS[1,2], Andrew Redd, PhD[1,2], Steven Pickard[1,2], Shuying Shen, MStat[1,2], Brett South, MS[1,2], Guy Divita, MS[1,2], Scott Duvall, PhD[1,2], Thien M. Nguyen, MS[3], Leonard W. D'Avolio, PhD[3], Matthew Samore, MD[1,2]**

**[1] VA Salt Lake City Health Care System, Salt Lake City, UT,**
**[2]University of Utah School of Medicine, Salt Lake City, UT;**
**[3]VA Boston Healthcare System, Jamaica Plain, MA**

**Abstract**

*Information retrieval algorithms based on natural language processing (NLP) of the free text of medical records have been used to find documents of interest from databases. Homelessness is a high priority non-medical diagnosis that is noted in electronic medical records of Veterans in Veterans Affairs (VA) facilities. Using a human-reviewed reference standard corpus of clinical documents of Veterans with evidence of homelessness and those without, an open-source NLP tool (Automated Retrieval Console v2.0, ARC) was trained to classify documents. The best performing model based on document level work-flow performed well on a test set (Precision 94%, Recall 97%, F-Measure 96). Processing of a naïve set of 10,000 randomly selected documents from the VA using this best performing model yielded 463 documents flagged as positive, indicating a 4.7% prevalence of homelessness. Human review noted a precision of 70% for these flags resulting in an adjusted prevalence of homelessness of 3.3% which matches current VA estimates. Further refinements are underway to improve the performance. We demonstrate an effective and rapid lifecycle of using an off-the-shelf NLP tool for screening targets of interest from medical records.*

**Introduction**

The free text of electronic medical notes is considered to be a rich source for health care operations and research. Information extraction and information retrieval methods using natural language processing (NLP) and machine learning have been successfully applied to electronic notes. The value of text data has been shown in several clinical and biomedical domains including bio-surveillance, adverse event detection and quality improvement[1-4].

With ever expanding medical databases, there is a need to bring information retrieval tools into the hands of all researchers. Be it for clinical operations, quality improvement or research, it is important to develop methods and tools that allow for rapid results using NLP to identify targets of interest. Often, this entails engaging the services of trained NLP scientists and programmers and working closely with them. Another paradigm is to deploy off-the-shelf NLP tools that can be used with minimal training and expertise.

Automated Retrieval Console v2.0 (ARC), an open source clinical information retrieval tool developed by Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC)[5], is an NLP tool that essentially retrieves 'documents like this one' based on a training set that contains sufficient numbers of positive and negative classifications [6,7]. ARC has been used by NLP and clinical researchers in several medical domains such as cancer and surgery[8,9], thus providing a mechanism for all researchers to access NLP tools in their domains.

Homelessness is a high-priority issue for all societies. Those returning from active combat (Veterans of wars) are known to be higher risk for homelessness[10] and so this non-medical 'diagnosis' is important to the US Department of Veteran Affairs (VA). Based on US Government estimates from data outside the VA, the estimates of homelessness among Veterans is nearly 1%[11]. Recent detailed studies using data on cohorts of Veterans returning from combat estimate the prevalence of homelessness among Veterans to be nearly 3.7%[12].

Current methods of identifying homeless Veterans are entirely based on administrative data such as ICD-9-CM codes for homelessness and specific clinic identifier codes that indicate that a Veteran is receiving homeless services through the VA (recent examples include[13-15]). The reliability and validity of using administrative data is domain specific and has been shown to be useful in several domains[16-20], though the validity of this approach for identifying homelessness has not been systematically studied. Validation of the administrative data by reviewing clinical documents for evidence of homelessness would serve as an ideal 'reference' standard; however, this is cost-prohibitive on a scale needed to make this meaningful as trained and experienced human reviewers require time and resources.

Although several risk factors for Veteran homelessness have been elucidated, there are currently few tools that are useful in predicting the risk of future homelessness among Veterans. References to risk factors for homelessness are often found only in the free text of medical records (unstructured data) written by providers in VA medical facilities and possibly precede the formal identification of Veterans as being homeless. This vital resource is not accessible to traditional methods of data mining and requires informatics-based methods such as NLP to reliably extract appropriate and relevant information for use in risk-stratification of Veterans for homelessness. To our knowledge, there are no current research studies involving mining of text data in the domain of homelessness research either within or outside the VA. With large numbers of troops returning from combat in the US, an urgent need exists to develop automated methods to screen for homelessness among Veterans, as early detection of homelessness leads to earlier intervention which will prevent many of the complications associated with long-term homelessness including increased morbidity and mortality [21-23].

The objective of this project was to develop natural language processing (NLP) algorithms to screen for homelessness among Veterans using clinical narratives in electronic clinical documents (unstructured data) using as few human/financial resources as possible and available open source off-the-shelf tools. The operational goal is to use these algorithms on large corpora of electronic notes to screen for homelessness and identify those individuals who are in need of essential preventive services, thereby supporting the mission of the VA.

## Methods

*Clinical Setting*

The VA has a large healthcare system and is the largest provider of healthcare services to homeless individuals in the US[24]. It encompasses more than 150 hospitals and delivers care to more than 6 million Veterans annually. As noted above, homelessness among Veterans is a significant public health problem.

*Selection of Corpus of Clinical Documents*

During the year 2009, a total of 2,229,983 Veterans (patients) were seen in multiple visits in nearly half of VA facilities throughout the US (two of four regions across the country); these visits generated a total of 60,921,956 clinical documents for an average of 27.3 documents per Veteran. From this corpus, a sample of 500 notes was selected with 'homeless' in the note title along with a control random sample of 500 notes from the rest of the corpus (Figure 1). The rationale for selecting notes with 'homeless' in the note title was to enrich the corpus for documents with evidence of homelessness. This over-sampling enriched the dataset with documents that contained references to homelessness in the free text in order for our algorithms to maximize discrimination, as otherwise the general prevalence of homelessness among Veterans is low.

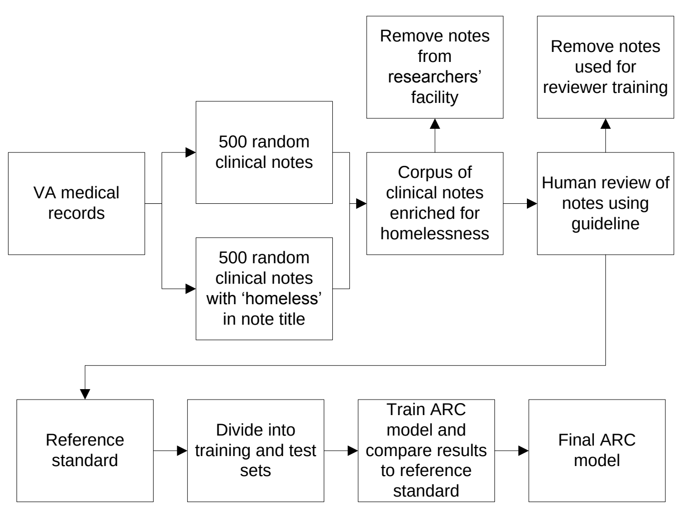


**Figure 1.** Flow chart of protocol followed to select clinical notes, develop reference standard and train the NLP tool (ARC, Automated Retrieval Console V2.0).

***Development of a Reference Standard for Document-Level Classification of Clinical Documents***

This set of 1,000 documents was evaluated by trained and experienced human reviewers. A written guideline was developed that reiterated US Department of Housing and Urban Development (HUD) definitions for homelessness[25]. In addition, a set of lexical terms was developed for psychosocial concepts related to homelessness based on the published literature, review of medical records of patients experiencing homelessness, and domain expertise [26].

Utilizing an annotation tool developed to simplify many annotation tasks, the documents were first pre-annotated with the lexical terms.[27] The reviewers then read through the text in the documents to determine if the terms were contextually appropriate. Finally, for each document, the reviewer annotated at least one instance or example of a word, term, and phrase or longer span of text (cue words) to classify the documents as containing information in one of three categories: 'confirmed evidence of homelessness'; 'possible or at risk of homelessness' or 'no evidence of homelessness'.

The corpus was reviewed by 3 independent reviewers, an initial inter-rater agreement score (kappa for 3 raters) was calculated at 95% for the training set of 80 documents that were reviewed by all three reviewers. The three reviewers then independently reviewed all remaining documents and a separate adjudicator reviewed and resolved the small number of discordant classifications. This resulted in a corpus of human-reviewed reference standard for training and testing NLP tools (Figure 1).

*NLP tool: Automated Retrieval Console v2.0*

Automated Retrieval Console v2.0 (ARC) is a VA-developed NLP tool that essentially retrieves 'documents like this one' based on a training set that contains sufficient numbers of positive and negative classifications [6,7]. The ARC design is based on the hypothesis that supervised machine learning with robust enough feature sets is capable of delivering acceptable performance across a number of clinical information retrieval applications. ARC uses UIMA-based pipelines for NLP (cTAKES). For supervised machine learning, ARC integrates the open source Application Programming Initiative (API) exposed by the MAchine Learning for Language Toolkit (MALLET) [28]. It is developed in Java and is available as open-source software at http://maveric.org/mig/arc.html. Details of the feature sets and machine learning algorithms are described elsewhere [21,22]. For this project, ARCv2.0 was downloaded, installed and used to process documents as-is (off the shelf) with no modifications in the code or interface. As noted above, to date, most applications of ARC have been in medical domains.

*Development of ARC Training Models*

The reference standard of human reviewed and classified documents was partitioned to an M1 set that consisted of a set of documents with evidence of confirmed homeless versus all others which were considered negative; an M2 set of documents with evidence of confirmed and possible homelessness versus all others which were considered negative. Documents with no evidence of homelessness were also used for training. The metrics used for evaluating the performance of ARC were precision (fraction of retrieved documents that are relevant, a proxy for positive predictive value); recall (fraction of relevant documents retrieved, a proxy for sensitivity) and F-measure (harmonic mean of precision and recall).

*Training ARC on Progressively Larger Sets from the Reference Standard*

An initial set of 243 (28% of corpus) documents from the reference standard corpus and a second set of 243 + 402 (Total 645, 75% of corpus) documents were used to train ARC to classify documents as having either 'evidence of homelessness' or 'no evidence'. The training was performed in an iterative manner with different training models and algorithms internal to ARC that select appropriate feature types and their related values to maximize the signal while reducing noise[6]. The models used a combination of concepts consisting of parts of speech such as noun-phrases (this constituted the concept-level work flow), full sentences (sentence-level work flow) and paragraphs (paragraph level work flow) (Figure 2).

The noun-phrase, or concept-level, work flow trains on the annotated terms and the phrase around it. This is the most narrow feature set that can be used. The sentence-level work flow utilizes the annotated term and the sentence containing that term. The paragraph level work flow uses the paragraph that contains a given annotated term or terms. As the work-flow becomes less narrowly focused, ARC is able to utilize more features in the training of the models to find the best performing model for information retrieval. At the document-level, ARC is able to take the human classification of the document and utilize all features within the document to train the model. It is important to note that for this project, we used the training provided by ARC as-is and the cue words (described above) annotated by the human reviewers were not used as training features.

*Testing ARC on Remainder of the Reference Standard (Test Set)*

After optimum training was achieved using 75% of the reference standard corpus, ARC was tested on the remainder of the reference standard (217 documents). These documents had not been previously processed through ARC.

*Application of ARC to a Naïve Corpus and Error Analysis*

The best performing model of ARC was then applied to a random corpus of 10,000 documents from Veterans seen across the US in medical facilities during a 10-year time period.

The prevalence of homelessness among the Veterans whose documents were in this corpus was estimated by reviewing administrative data associated with these documents. Each document was associated with a visit and each visit was associated with ICD-9-CM diagnoses codes and unique 'stop' codes that are used in the VA healthcare system to indicate the type of clinic in which the patient was seen. We used a VA standard definition of homelessness using ICD-9-CM codes for homelessness (V60.0) and stop codes indicating receipt of homeless services (511, 522, 528, 529 and 530)[12,29].

An error analyses was performed by reviewing all documents that were flagged as positive by administrative data (admin) and ARC. Using explicit guidelines based on the US Housing and Urban Development (HUD) definition of homelessness [25], the flagged documents were classified by two trained human reviewers as 'confirmed evidence of homelessness'; 'possible or at risk of homelessness' or 'no evidence of homelessness' following the protocol used for generating the original reference standard described above.

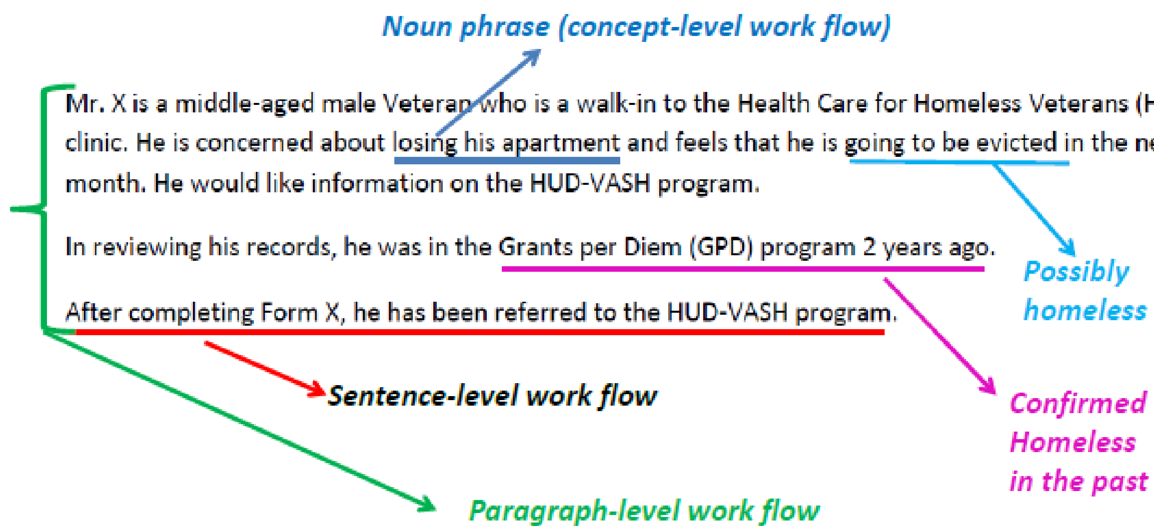


**Figure 2.** Training the NLP tool (Automated Retrieval Console v2.0) using either noun phrases, full sentences, complete paragraphs or the entire text of the document. In this synthetic document generated for illustrative purposes, the grants per diem program (GPD) and the Housing and Urban Development Veterans Affairs Supportive Housing (HUD-VASH) refer to specific programs for homeless Veterans that provide funding for temporary housing.

Due to the large number of negative documents expected in this corpus of 10,000 documents, the documents that flagged negative by administrative data and ARC were not reviewed manually by humans. Thus, the metric reported from the error analyses is precision or positive predictive value (true positive/ (true + false positives)).

**Results**

*Reference Standard corpus*

After removing the 80 initial training documents and, to eliminate bias, 58 more that were noted to be from the medical facility where the research was being conducted, there were a total of 862 documents in the final corpus of reference standard documents. Of these, 388 (45%) were classified by human review as having 'confirmed evidence' of homelessness, 34 (4%) were 'possible or at risk of' homelessness and 440 (51%) of the documents were noted to have 'no evidence' of homelessness. The 'confirmed' and 'possible or at risk of' were combined to generate the 'homeless evidence' set.

*Training ARC: Results on First Set of 243 Documents*

The simplest and least computationally intense Models M1 and M2 using only the noun-phrase concept level workflow were used to train ARC. As shown in Table 1, the results were sub-optimal, though with M2, the precision increased to 91% at the expense of the recall. This indicated the need for a larger training set and also a larger annotation context.

*Training ARC: Results on Second Set of 645 Documents*

Using a larger portion of the reference standard (645 documents and concept-level workflow) also resulted in suboptimal performance (Table 1). When the work flow for ARC was switched to document level with use of all possible features as determined by internal ARC algorithms, the test characteristics improved dramatically: the precision was noted to be 95%, the recall was 95% and the F-measure was 95.

*Results of Testing ARC on remainder of the Reference Standard (Test Set)*

The remainder of the reference standard documents (N= 217) were processed through ARC using the best performing model based on document level annotation. Performance metrics were: Precision 94%, Recall 97%, F-Measure 96.

**Table 1.** Results of **training** the NLP tool (ARC) using either concept or document level work flow and models described in the text. The training corpus consisted of either 243 or 645 documents of the reference standard.

Model M1: set of documents which were confirmed homeless vs. negative; Model M2: set of documents which were confirmed **and** possible homeless vs. negative

| Work Flow | | Model | Precision | Recall | F-Measure |
|---|---|---|---|---|---|
| **Concept-Level Work Flow Using <u>243 documents</u> for Training** | Noun-phrase | M1 | 11 | 63 | 19 |
| | | M2 | 3 | 91 | 5 |
| **Concept-Level Work Flow Using <u>645 documents</u> for Training** | Noun-phrase | M1 | 7 | 20 | 1 |
| | | M2 | 1 | 30 | 2 |
| | Sentence | M1 | 5 | 57 | 9 |
| | Paragraph | M1 | 5 | 59 | 9 |
| | | M2 | 5 | 53 | 9 |
| **Document-Level Work Flow Using <u>645 documents</u> for Training** | All possible features using ARC internal algorithms | M2 | **95** | **95** | **95** |

*Screening for Homelessness using ARC on a Random Corpus of 10,000 Documents*

This corpus of 10,000 documents represents 9924 unique patients (Veterans), of whom 93% are male. With regard to race, 62% are white, 20% are American Indian or Alaskan Native, 14% are black or African American, and the remaining 4% are either Asian, Mexican American, Pacific Islander, or ‘unknown’. The majority were in the age group 40-69 years of age. The documents were from visits over an 8 year period between 2002 and 2010. The distribution of demographics and note titles in this corpus reflects the general VA patient population.

Based on a subset of documents classified by both reviewers, the inter-rate reliability (kappa statistic) for the error analyses was noted to be 97% between the two reviewers.

Administrative codes for homelessness were positive for 62 documents from this corpus (prevalence of homelessness of 62 per 9924 Veterans, 0.6%). This matches the estimates developed by the US Government in 2009[11]. The 3 most common note titles associated with an administrative diagnosis of homelessness were social work notes, mental health notes and homeless program notes. Human review of all 62 documents (error analysis) revealed a precision of 87% (Table 2). This would imply an adjusted prevalence of homelessness in this cohort of 0.5%.

ARC flagged 463 documents as positive for homelessness implying a prevalence of homelessness of 4.7% in this cohort of 9924 unique Veterans. The most common note titles flagged by ARC were mental health, social work and addendum (which are usually short notes that are added or attached to a main note). The error analysis by human review of these 463 documents noted a precision of 70% for homelessness (with 324 documents having either confirmed evidence or at risk of homelessness). This would indicate an adjusted prevalence of homelessness of 3.3% in this random set of 9924 Veterans, which matches recent estimates of homelessness among Veterans of 3.7%[12].

**Table 2.** Results of processing a random corpus of 10,000 documents from 9924 Veterans by the NLP tool (ARC v2.0) and associated error analyses for determining precision for homelessness by human review. The most common document titles in the entire corpus were nursing notes, primary care notes and addendum

| **Determination of Homelessness By** | **Positive Flags (Implied prevalence of homelessness)** | **Precision by Human Review (Implied prevalence of homelessness)** | **Common Document Titles of all Positives** |
|---|---|---|---|
| Using Administrative codes as described in the text | 62<br>(0.6%) | 87%<br>(0.5%) | Social work, mental health, homeless program notes |
| Using ARC NLP algorithms | 463<br>(4.7%) | 70%<br>(3.3%) | Social work, mental health, addendum |

**Discussion**

This project demonstrates the feasibility of rapidly assembling a pipeline for information retrieval using electronic clinical documents to answer the question: *“Are there documents in this corpus that are like the reference standard?”* We were able to extract and generate a reference standard of ‘positive’ and ‘negative’ documents in a reasonable period of time. ARC was downloaded and used ‘off the shelf’ with no modifications to the code or the interface. With minimal input from the developers of ARC, our team was able to install, run and re-run models for training ARC, identify the best performing model and then apply that model to a test set of documents. All steps were completed in approximately 3 months. ARC has been successfully used by several groups for information retrieval in medical domains. *To our knowledge, this is the first application of ARC in identifying a non-medical diagnosis such as homelessness from electronic medical records.*

There are several lessons learned that are applicable to our project and for others in the field. The resources and time taken to develop a lexicon, a reference standard, and the learning curve to use ARC were modest. This exemplifies the original intent of ARC to bring NLP to a wider audience.

The performance of the NLP tool (ARC) on the training and test sets of the reference standard varied according to the model used. Contrary to our intuition, the model that used only noun phrases (concept-level workflow) alluding to direct evidence of homelessness performed extremely poorly with precision in the single digits. Progressively broadening the feature set used for training was also only marginally beneficial until the training included all features within the document. At that point, we noted a dramatic rise in precision, recall and F-measure. It is possible that the concept-level workflow was not suited to this particular task.

The performance of ARC with no further modifications on a random corpus of VA text documents was suboptimal with a precision of 57%. However, as the documents flagged by ARC are known to be of the highest yield for homelessness (based on administrative data and other results not shown), it is very likely that fine tuning internal ARC models or adding a module that is informed by the error analyses will improve the performance. If the documents identified from the error analysis with one or more risk factors are included as 'possible' evidence of homelessness, then the precision improves to 70%. This raises the issue of training on an enriched corpus and then using ARC on a 'real world' corpus of documents in which prevalence of homelessness is known to be low. We are investigating alternative training and test set designs to optimize real world accuracy problems introduced by low prevalence targets.

It is interesting to note that in reviewing the administrative code flags for homelessness from VA databases of the visits associated with these documents, the prevalence of homelessness was less than 1%. This is likely to be falsely low and reveals the inherent problem with relying on administrative data alone. From the error analyses on the positives flagged by ARC, it is evident that there are more Veterans with evidence of homelessness than indicated by ICD codes alone. This 'adjusted' prevalence of homelessness (3.3%) based on our results more closely matches other VA estimates[12]. Thus administrative codes are likely not sufficient to estimate true prevalence of homelessness among Veterans. It is important to note that the Veterans from our document corpus may be different with regard to their recent combat status to the cohort described in the Department of Veterans Affairs report [12] and thus gross comparisons of the prevalence of homelessness may be challenging.

The time and resources needed to develop large reference standards and perform effective error analyses for training NLP tools are often substantial. Though this project was not designed to address the optimum size of reference and training sets, these are important in developing and implementing IR pipelines. Small, practical and easily developed data sets would be ideal with a trade-off between precision and recall.

We acknowledge several limitations. The classification of homelessness was performed at the individual 'document' level. The patient-level diagnosis of homelessness was not ascertained for either the reference standard or random corpus. We cannot exclude the possibility that the patients had a diagnosis of homelessness either before or after the index document. This may be an important factor in determining if the patients whose documents had evidence of risk factors by error analyses were noted to be homeless at any other time point. As the prevalence of homelessness is relatively low in the corpus, there are more than 9500 negative documents. It is likely that there are false negatives in the data corpus. A human review of a small sample would likely not yield meaningful results and the time/resources required to appropriately review a large segment of this corpus are prohibitive. We plan to address the important issue of false negatives in a future project. The estimates of prevalence of homelessness are based on precision only; thus these should be considered more as lower bounds.

The precision we achieved for this task was modest at best. In reality though, NLP classifiers may not need to achieve near perfect performance characteristics. As with most surveillance systems, if we are able to narrow down the documents needed to be reviewed with one pass of an automated algorithm with reasonable performance, we can then apply second algorithms to further improve performance.

Further iterations of ARC algorithms to screen large corpora of patient documents are underway; including further customizing the feature set ARC utilizes to train models. One immediate possibility is to include the set of cue words annotated by humans as a custom feature set for training. Other enhancements to ARC in later versions would also likely provide mechanisms for improving performance.

There are practical operational uses for information retrieval tasks involving homelessness. The immediate objective is to develop reliable and efficient processes to screen documents of Veterans for homelessness. The healthcare system can then determine whether those Veterans who screen positive are receiving appropriate services. Our ultimate goal is to develop predictive models to identify those at risk of homelessness so that services can be offered to those individuals to prevent homelessness.

**Conclusion**

Natural language processing based tools can be used to screen for non-medical diagnoses such as homelessness from electronic medical records. We have shown the effective training and testing of an off-the-shelf tool (Automated

Retrieval Console v2.0) using a modest sized document corpus of 1000 documents, limited resources, and a reasonable time frame. With no further modifications, the application of ARC on a random corpus of documents yielded a precision of 70% and an adjusted prevalence of homelessness among this cohort of Veterans of 3.3%. Further refinements are underway to develop ARC models that could be used by operational partners and researchers to better understand homelessness among Veterans and offer appropriate prevention services.

**Acknowledgments**

The research reported here was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service (HIR 10-002). Dr. Adi Gundlapalli is the principal investigator, and a staff physician, at VA Salt Lake City Health Care System. Resources and administrative support were provided by the VA Salt Lake City Health Care System (IDEAS Center). We gratefully acknowledge the insights gained through discussions and collaborations with our homeless service community partners in Salt Lake City, Utah: Fourth Street Clinic (Mr. Monte Hanks); The Road Home (Ms. Michelle Flynn and Ms. Michelle Vasquez), Volunteers of America (Ms. Jessica Fleming, Ms. Jamie Jones) and the State of Utah (Ms. Kathleen Moore). We would like to thank our colleagues at the National Center on Homelessness Among Veterans (Drs. Dennis Culhane, Steven Metraux and Jamison Fargo) for their discussions and advice. We also thank Jorie Butler, PhD, Warren Pettey, MPH, CPH, Sarah Craig, BS for their help with this project.

## Disclaimer

The views expressed in this article are those of the author(s) and do not necessarily represent the views of the Department of Veterans Affairs.

## References


1. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of medical informatics.* 2008:128-144.
2. Chapman WW. Closing the gap between NLP research and clinical practice. *Methods of information in medicine.* 2010;49(4):317-319.
3. Jha AK. The promise of electronic records: around the corner or down the road? *JAMA : the journal of the American Medical Association.* Aug 24 2011;306(8):880-881.
4. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *Journal of the American Medical Informatics Association : JAMIA.* Sep-Oct 2011;18(5):544-551.
5. McGuire J, Rosenheck RA, Kasprow WJ. Health status, service use, and costs among veterans receiving outreach services in jail or community settings. *Psychiatr Serv.* Feb 2003;54(2):201-207.
6. D'Avolio LW, Nguyen TM, Goryachev S, Fiore LD. Automated concept-level information extraction to reduce the need for custom software and rules development. *Journal of the American Medical Informatics Association : JAMIA.* Sep-Oct 2011;18(5):607-613.
7. D'Avolio LW, Nguyen TM, Farwell WR, et al. Evaluation of a generalizable approach to clinical information retrieval using the automated retrieval console (ARC). *Journal of the American Medical Informatics Association : JAMIA.* Jul-Aug 2010;17(4):375-382.
8. D'Avolio LW, Nguyen TM, Farwell WR, et al. Evaluation of a generalizable approach to clinical information retrieval using the automated retrieval console (ARC). *Journal of the American Medical Informatics Association.* July 1, 2010 2010;17(4):375-382.
9. Vo E, Davila JA, Hou J, et al. Differentiation of ileostomy from colostomy procedures: Assessing the accuracy of current procedural terminology codes and the utility of natural language processing. *Surgery.* 8// 2013;154(2):411-417.
10. Balshem H, Christensen V, Tuepker A, Kansagara D. A Critical Review of the Literature Regarding Homelessness Among Veterans. In: US Department of Veterans Affairs, ed. *A Critical Review of the Literature Regarding Homelessness Among Veterans*. Washington (DC)2011.
11. US Housing and Urban Development and US Department of Veterans Affairs. Veteran Homelessness: A Supplement to the 2009 Annual Homeless Assessment Report to Congress. In: US Housing and Urban Development and US Department of Veterans Affairs, ed. Washington DC: US Government; 2011.
12. Department of Veterans Affairs Office of Inspector General. Homeless Incidence and Risk Factors for Becoming Homeless in Veterans. In: VA Office of Inspector General, ed. Washington DC: VA Office of Inspector General,; 2012.
13. Harris AH, Reeder RN, Ellerbe L, Bowe T. Are VHA administrative location codes valid indicators of specialty substance use disorder treatment? *Journal of rehabilitation research and development.* 2010;47(8):699-708.
14. Austin J, McKellar JD, Moos R. The influence of co-occurring axis I disorders on treatment utilization and outcome in homeless patients with substance use disorders. *Addictive behaviors.* Sep 2011;36(9):941-944.
15. Edens EL, Mares AS, Tsai J, Rosenheck RA. Does active substance use at housing entry impair outcomes in supported housing for chronically homeless persons? *Psychiatr Serv.* Feb 2011;62(2):171-178.
16. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. *Medical care.* Sep 1998;36(9):1324-1336.
17. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. *Journal of clinical epidemiology.* Apr 2007;60(4):397-409.
18. Roumie CL, Mitchel E, Gideon PS, Varas-Lorenzo C, Castellsague J, Griffin MR. Validation of ICD-9 codes with a high positive predictive value for incident strokes resulting in hospitalization using Medicaid health data. *Pharmacoepidemiology and drug safety.* Jan 2008;17(1):20-26.
19. Banerjea R, Findley PA, Smith B, Findley T, Sambamoorthi U. Co-occurring medical and mental illness and substance use disorders among veteran clinic users with spinal cord injury patients with complexities. *Spinal cord.* Nov 2009;47(11):789-795.
20. Tracy LA, Furuno JP, Harris AD, Singer M, Langenberg P, Roghmann MC. Predictive ability of positive clinical culture results and International Classification of Diseases, Ninth Revision, to identify and classify noninvasive Staphylococcus aureus infections: a validation study. *Infection control and hospital*

*epidemiology : the official journal of the Society of Hospital Epidemiologists of America.* Jul 2010;31(7):694-700.

**21.** White MC, Tulsky JP, Dawson C, Zolopa AR, Moss AR. Association between time homeless and perceived health status among the homeless in San Francisco. *Journal of community health.* Aug 1997;22(4):271-282.

**22.** O'Connell JJ, Mattison S, Judge CM, Allen HJ, Koh HK. A public health approach to reducing morbidity and mortality among homeless people in Boston. *Journal of public health management and practice : JPHMP.* Jul-Aug 2005;11(4):311-316.

**23.** Hibbs JR, Benner L, Klugman L, et al. Mortality in a cohort of homeless adults in Philadelphia. *The New England journal of medicine.* Aug 4 1994;331(5):304-309.

**24.** Statement of Peter H. Dougherty, Director, Homeless Veterans Programs, Department of Veterans Affairs before the Senate Committee on Appropriations, May 1, 2008. 2008.

**25.** U.S. Department of Housing and Urban Development. Federal Definition of Homelessness. 2011; http://portal.hud.gov/hudportal/HUD?src=/topics/homelessness/definition. Accessed December 3, 2011, 2011.

**26.** Gundlapalli A, Redd A, Carter M, et al. Validating a Strategy for Psychosocial Phenotyping Using a Large Corpus of Clinical Text. *Submitted for Publication.* 2013.

**27.** South B, Shen S, Leng J, Forbush T, Duvall S, Chapman W. A Prototype Tool Set to Support Machine-Assisted Annotation. Paper presented at: BioNLP 20122012; Montreal, Canada.

**28.** *MALLET: A Machine Learning for Language Toolkit* [computer program]. Amhert, MA: University of Massachusetts; 2002.

**29.** Blow F, McCarthy J. Care for Veterans with Psychosis in the Veterans Health Administration, FY10, 12th Annual National Psychosis Registry Report. In: US Department of Veterans Affairs, ed2010.