Evaluation of Negation Phrases in Narrative Clinical Reports

Wendy W. Chapman PhD[1], Will Bridewell BS[2], Paul Hanbury BS[1], Gregory F. Cooper MD PhD[1,2], and Bruce G. Buchanan PhD[1,2]

[1]Center for Biomedical Informatics, [2]Computer Science/Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA

ABSTRACT

Objective: *Automatically identifying findings or diseases described in clinical textual reports requires determining whether clinical observations are present or absent. We evaluate the use of negation phrases and the frequency of negation in free-text clinical reports.*

Methods: *. A simple negation algorithm was applied to ten types of clinical reports (n=42,160) dictated during July2000. We counted how often each of 66 negation phrases was used to mark a clinical observation as absent. Physicians read a random sample of 400 sentences, and precision was calculated for the negation phrases. We measured what proportion of clinical observations were marked as absent.*

Results*: The negation algorithm was triggered by sixty negation phrases with just seven of the phrases accounting for 90% of the negations. The negation phrases received an overall precision of 97%, with "not" earning the lowest precision of 63%. Between 39% and 83% of all clinical observations were identified as absent by the negation algorithm, depending on the type of report analyzed. The most frequently used clinical observations were negated the majority of the time.*

Conclusion*: Because clinical observations in textual patient records are frequently negated, identifying accurate negation phrases is important to any system processing these reports.*

INTRODUCTION

A large portion of computerized patient data is stored as narrative reports. Narrative medical reports describe imaging observations, physical symptoms, and clinical assessments that can be useful for clinical research. Researchers in the medical informatics community have automatically extracted information contained in narrative reports for decision support [1], guideline implementation [2,3], detection and management of epidemics [4], and identification of patients eligible for research studies [5].

Automatic extraction or classification systems must identify whether observations in a report are present or absent. Handling negation in narrative clinical documents is still an open research topic on which little has been published. In this paper we evaluate which phrases are most frequently used to indicate the absence of a clinical observation. In addition, we examine the precision of the most commonly used negation phrases and measure how often clinical observations are negated.

BACKGROUND

This study was motivated by requirements of a system currently being developed at the University of Pittsburgh to identify patient subsets (IPS) for retrospective studies [5,6]. Given a superset of patients, the IPS system can aid researchers in identifying subgroups of patients of interest by using textual data from medical records. The IPS system uses a naïve Bayesian algorithm to create a probabilistic query that will identify subgroups of patients based on the discriminatory power of UMLS phrases and other keywords contained within the textual records.

The performance of the IPS system depends largely on accurately identifying whether UMLS phrases in a medical report are present or absent. We previously tested a simple negation algorithm called NegEx [7] which marks as absent those UMLS phrases that occur up to six words after a negation word. NegEx obtained 78% sensitivity and 84% precision on sentences selected from discharge summaries.

Here we extend the study of NegEx's performance by analyzing negation phrases not only in discharge summaries but also in other types of clinical reports. We believe the results of our study are applicable not only to the IPS system but to any algorithm attempting to identify negation within clinical reports.

METHODS

Below we describe the reports in which we analyzed negation. We also describe our algorithm for identifying relevant UMLS phrases, the negation phrases we tested, and the evaluations we performed.

Reports

All reports used in this study were extracted from the MARS (Medical Archival System, Inc) system [8] at the University of Pittsburgh Medical Center. The MARS system stores thirteen types of dictated clinical reports. We selected ten report types we felt represented the majority of all clinical reports stored on MARS. For every report type we selected all reports dictated during July 2000, totaling 42,160 reports (Table 1). The reports were dictated by more than two thousand different physicians.

Table 1. Reports from July 2000

Report Type	Number of Reports	Number of Sentences
Progress notes (PN)	13,888	229700
Chest X-ray (CXR)	6,645	34489
Emergency department (ER)	4,533	193205
Surgical pathology (SP)	3,906	64265
Operative notes (OP)	3,376	80936
History & physical exams (HP)	3,260	150581
Computed tomography (CT)	2,597	29041
Discharge summaries (DS)	2,360	53701
Magnetic resonance imaging (MRI)	1,132	10480
Mammography (MAM)	463	2271
Total	42,160	848,669

Identifying relevant UMLS phrases

The IPS system's preprocessor identifies all UMLS phrases contained in the text by matching all strings in the text with UMLS string unique identifiers (SUI). Various methods exist for indexing documents with UMLS phrases (see [9] for an overview). Using simple string matching decreases our ability to identify relevant UMLS phrases that might be matched with more sophisticated methods (e.g., stemming) but reduces the noise created by false matches. For this study we limited UMLS phrases to those that represent clinical observations (as opposed to lab tests or surgeries performed, for example). All UMLS phrases with semantic types of finding, disease, symptom, fungus, congenital abnormality, acquired abnormality, lab result, injury, biologic function, physiologic function, mental process, mental dysfunction, cell dysfunction, anatomic abnormality, or experimental model of disease were identified in the reports. We deleted UMLS phrases that were either negation phrases (e.g., "no") or combinations of a negation phrase and an already existing UMLS phrase (e.g., "no rash"). We also matched the longest UMLS phrases available (e.g., "nonspecific viral rash" instead of "rash").

Negation phrases

Negation phrases used in this study came from three sources. First, we used all negation phrases triggered in the previous analysis of NegEx on discharge summaries. Second, we added negation phrases utilized by a system called SymText [10].Third, one of the authors (WWC) read a sampling of reports from each report type and added negation phrases that were not yet identifed.

Below we list three types of negation phrases, along with regular expressions (RE) in which the negation phrases were used in this study.

1) **Pre-UMLS phrases** – negation phrases that occur before a UMLS term:
RE: "**negation phrase** [1-5 words] *UMLS phrase*"
Example: "The patient **denies** any episodes of *chest pain*" -> chest pain – absent

2) **Post-UMLS phrases** – negation phrases that occur after a UMLS term:
RE: "*UMLS phrase* [1-5 words] **negation phrase**.
Example, "*Aspiration pneumonia* is **unlikely**" -> aspiration pneumonia – absent

3) **Pseudo-negation phrases** – phrases that resemble negation phrases but do not actually indicate negation. These phrases prevent a negation from being triggered and were all used with pre-UMLS regular expressions.
Example: "There has been **no change** in the patient's *jaw disease*" -> "jaw disease – present."

Evaluations

We performed three types of exploratory evaluations on the test reports. The first evaluation determined the most frequently used negation phrases. For every report type we calculated the number of times every negation phrase triggered NegEx.

In the second evaluation we calculated the precision (positive predictive value) of frequently triggered negation phrases. We randomly selected 400 sentences from the set containing the most frequently triggered negation phrases. Physicians judged whether the marked UMLS phrases in the sentences were present or absent. To be considered present, an observation must have been indicated as present or possible at the current visit. Observations explicitly said to be absent, observations from the patient's past history, and observations listed as future possibilities were considered absent at the current visit. We compared judgments made by the algorithm to those made by the physicians to calculate the precision of the negation phrases. Thus, precision is a measure of the correctness of the subset of UMLS phrases negated by the algorithm.

Four physicians were trained together for the judging task. The individual physicians then judged the UMLS phrases in the sentences. Each of four physicians analyzed 200 sentences. Reports were assigned so that every physician analyzed approximately 67 sentences in common with every other physician. Every sentence was read by two physicians. A fifth physician ruled on disagreements.

The third evaluation explored the nature of negation in different clinical reports. We measured the proportion of UMLS terms that were negated to indicate how much negation occurs in different types of reports. We also quantified how often the most frequently used UMLS terms were negated.

RESULTS

Below we describe results addressing the most frequently used negation phrases and the nature of negation in different clinical reports.

Most frequently used phrases

Of the 66 negation phrases compiled by the authors, 60 triggered negation of UMLS phrases in the July 2000 test set. The 14 most frequently used negation phrases (i.e., phrases that each accounted for more than 3% of the negations in any report type) are shown in Table 2. The most frequently occurring negation phrase in all report types was "no."

Although 60 negation phrases were triggered in the test set, just seven negation phrases accounted for 90% of the negations in the test reports (Figure 1).

We also compared negation phrases triggered in radiology reports to those triggered in non-radiology reports. Non-radiology reports contained examples of all 60 triggered negation phrases. Radiology reports contained a subset of 41 negation phrases. Negation phrases uniformly missing from radiology reports but present in non-radiology reports typically fell into one of three categories: patient reporting of symptoms (e.g., "denies" and "declines"); knowledge about a patient's current or past disease state (e.g., "not had" and "never developed"); and findings from physical exam (e.g., "not feel"). Figure 2 shows the breakdown of the most frequently used negation phrases in radiology reports.

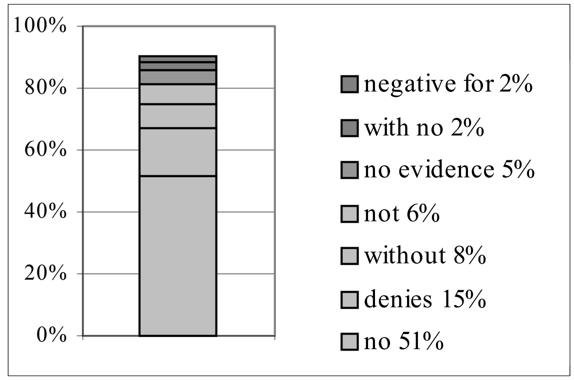

Figure 1. Negation phrases accounting for the majority of negations in the test set.

Table 2. Most Frequently Used Negation Phrases

	ER	PN	HP	DS	CXR	CT	OP	SP	MAM	MRI
no (n=62436)	51.2%	52.9%	52.7%	41.7%	63.2%	55.0%	39.7%	50.4%	63.2%	43.4%
denies (n=17845)	21.0%	13.2%	16.4%	4.9%			1.1%	0.1%		
without (n=9538)	5.6%	11.1%	8.6%	9.3%	0.6%	6.1%	6.0%	4.1%	13.3%	2.2%
not (n=7591)	4.9%	7.6%	5.8%	8.9%	1.3%	1.0%	15.8%	16.4%	0.3%	3.7%
no evidence (n=5488)	3.0%	2.9%	2.5%	5.2%	26.7%	23.4%	16.9%	5.6%	13.6%	29.5%
with no (n=3009)	2.2%	2.1%	2.2%	4.6%	3.5%	3.4%	4.4%	6.8%	0.6%	
negative for (n=2979)	4.1%	0.7%	2.5%	3.5%		0.1%	0.4%	1.6%		3.2%
denied (n=1576)	1.4%	0.7%	1.2%	4.8%			0.6%	0.3%		
to rule out (n=932)	0.6%	0.5%	0.8%	2.4%	0.1%	0.0%	3.1%	1.3%		
no significant (n=820)	0.7%	0.6%	0.5%	0.9%	0.1%	1.2%	1.1%	2.2%		5.7%
w/o evidence (n=397)	0.2%	0.2%	0.1%	0.5%	0.6%	3.8%	1.2%	0.2%		3.2%
no new (n=368)	0.1%	0.4%	0.4%	0.2%	0.3%	1.6%	0.1%		3.1%	
no abnormal (n=105)	0.0%	0.1%	0.1%	0.0%		1.2%	0.0%	0.1%		2.1%
no suspicious (n=55)		0.1%	0.0%	0.0%		0.1%		0.1%	5.3%	2.5%

* Cells represent the percentage of negations in the report type triggered by the negation phrase. For example, 51.2% of all negations in ER reports were triggered by "no." Empty cells indicate the negation phrase did not trigger in the report type.

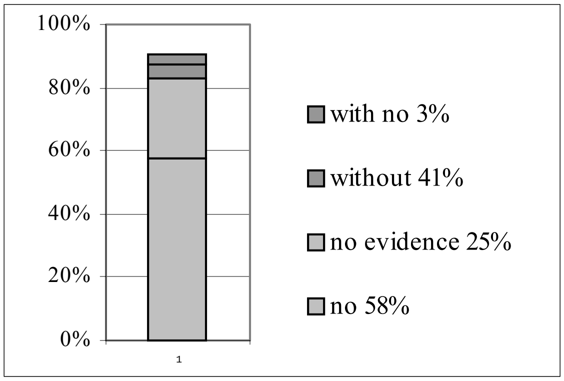

Figure 2. Negation phrases accounting for the majority of negations in radiology reports.

Precision of phrases

To test precision, four hundred sentences were randomly selected. The test set contained 813 UMLS observations that were marked as absent by the algorithm. Physicians comprising the gold standard disagreed on twelve sentences that were subsequently ruled on by a fifth physician (author GFC). Overall, 97% (791/813) of the UMLS concepts marked as absent were also judged to be absent by the gold standard (precision = 97%). Thirteen of the fourteen most frequently used negation phrases were represented in the sample set of 400 sentences (Figure 3).

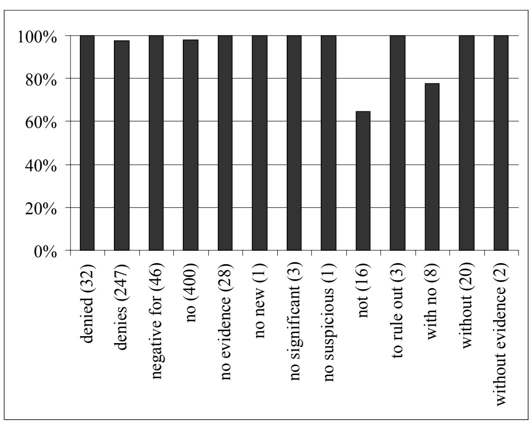

Figure 3. Precision of thirteen negation phrases represented in 400 sentences. Numbers in parentheses are counts of occurrence.

Negation in different clinical reports

We measured the proportion of UMLS phrases marked absent in the reports, shown in Figure 4.

An examination of the five most frequently used UMLS phrases for each report type yielded 20 unique phrases. The 20 phrases were identified between 14 (*nodule*) and 7,667 (*lesion*) times in the entire test set. Below we list the 20 phrases, along with the percentage of the time the phrases were negated in the test set. Seventeen of the 20 phrases were marked absent by NegEx the majority of time.

•*Distress*	99%	•*ulcer*	81%
•*pneumothorax*	98%	•*lesion*	80%
•*fracture*	95%	•*pleural effusion*	76%
•*febrile*	92%	•*nausea*	76%
•*edema*	91%	•*pneumonia*	75%
•*mass*	86%	•*infection*	75%
•*birth defect*	86%	•*mutation*	67%
•*shortness of breath*	85%	•*pain*	50%
•*chest pain*	85%	•*nodule*	14%
•*vomiting*	85%	•*retroflexion*	1%

DISCUSSION

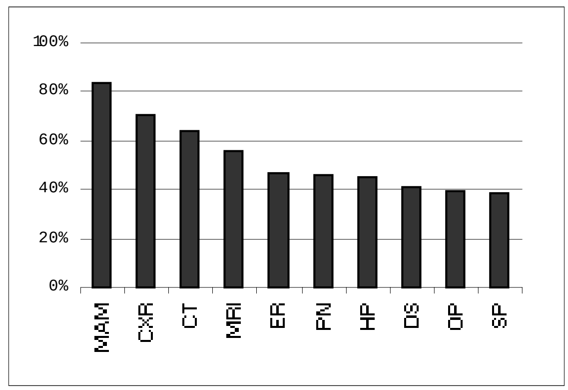

Figure 4. Percentage of UMLS phrases marked absent in reports (number negated concepts in reports / total number marked concepts in reports).

As shown in Figures 1 and 2, in medical reports a small number of negation phrases appear to be used the majority of the time to indicate the absence of clinical observations. Because we do not measure which UMLS concepts were not negated due to a missing negation phrase, we can not claim complete coverage of negation phrases. Still, evidence from our previous study and human effort used to compile negation phrases for this study support the claim that "no," "denies," "without," and "not" are the most frequently used phrases to indicate the absence of clinical observations.

Negation phrases that triggered NegEx showed high precision. However, our sample size of 400 test sentences was low and probably underrepresented negation as exhibited in the entire set of reports. The negation phrase with the lowest precision was "not." Determining whether "not" applies to a nearby UMLS phrase results from difficulty in determining the scope of the negation phrase. Consider the following sentence: "This is not the source of the infection." The concept being negated by "not" is the *source*, not the *infection* itself. A simple negation algorithm such as NegEx does not take into account syntactic clues or lexical information that help determine the scope of the negation phrase.

Our results showed that clinical observations are negated more often in radiology reports than in other types of reports. Radiology exams often list non-significant findings to assure the physician of their absence. However, clinical findings in various other report types were also negated by our algorithm a substantial portion of the time. A clinician's ability to narrow a long differential list to a single or a few probable diagnoses is evidenced by ruling out findings that support competing diagnoses. We would therefore expect to see a large portion of clinical findings explicitly mentioned as absent.

In fact, most frequently occurring UMLS were negated the majority of the time (see Table 3). These results quantitatively highlight the importance of accurately detecting negation in free-text clinical reports: Because a large portion of all clinical findings mentioned in textual reports are negated, accurately identifying whether clinical observations are present or absent is critical to accurately extracting information from the reports.

A major limitation of this study is not assessing the sensitivity of the negation phrases used by NegEx. Determining the completeness of our list of negation phrases could be accomplished by having physicians read random samples of sentences from the reports that statistically represent the entire set of reports. In this way, we could determine negation phrases that are used in the reports but not currently implemented by NegEx. Armed with such knowledge, we could extend NegEx to represent the additional negation phrases. We plan to pursue a study of sensitivity.

Also not addressed here are false negatives made by NegEx (i.e., UMLS phrases that should be marked as absent but are marked as present). Our previous study [7] measured sensitivity and revealed deficiencies in the simple algorithm. Syntactic information regarding the scope of the negation phrase would certainly help identify UMLS phrases farther away from the negation phrase that should be negated.

Conclusion

Negated clinical observations are frequent in clinical reports, and identifying whether the observations are present or absent is crucial to representing the information described in the report. Therefore, any system indexing clinical observations in narrative reports should address negation. Addressing negation involves identifying phrases that indicate negation within the reports. Our results show that a small set of negation phrases account for a large portion of negation occurring in clinical reports. Despite the high precision of most of the frequently occurring negation phrases, to be robust a negation algorithm triggered by the negation phrases should also apply syntactic and lexical knowledge.

Acknowledgments

We thank the physicians who read the reports: Michael Caruso, MD, Rebecca Crowley, MD, Christopher Mast, MD, and Andrew Post, MD. Wendy Chapman is supported by NLM training grant T15 LM07059. This project was in part supported by NLM grant G08 LM06625.

References

1. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. J Am Med Inform Assoc. 2000;7:593-604.
2. Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. Proc. AMIA Symp. 1999;:256-60.
3. Fiszman M, Haug PJ. Using medical language processing to support real-time evaluation of pneumonia guidelines. Proc. AMIA Symp. 2000;:235-9.
4. Hripcsak G, Knirsch CA, Jain NL, Stazesky RC, Pablos-mendez A, Fulmer T. A health information network for managing inner city tuberculosis: bridging clinical care, public health, and home care. Comput Biomed Res. 1999;32:67-76.
5. Aronis JM, Cooper GF, Kayaalp M, Buchanan BG. Identifying patient subgroups with simple Bayes'. Proc AMIA Symp. 1999;:658-62.
6. Cooper GF, Buchanan BG, Kayaalp M, Saul M, Vries JK. Using computer modeling to help identify patient subgroups in clinical data repositories. Proc AMIA Symp. 1998;:180-4.
7. Bridewell W, Chapman WW, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform. 2001 (submitted)
8. Yount RJ, Vries JK, Councill CD. The Medical Archival System: an information retrieval system based on distributed parallel processing. Information Processing Management, 1991;27:379.
9. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. J Am Med Inform Assoc. 2001;8:80-91.
10. Koehler, SB. SymText: A natural language understanding system for encoding free text medical data. Ph.D. Dissertation, University of Utah, 1998.