Hedging their Mets: The Use of Uncertainty Terms in Clinical Documents and its Potential Implications when Sharing the Documents with Patients

David A. Hanauer, MD, MS,[1] Yang Liu, MSI,[2] Qiaozhu Mei, PhD,[2] Frank J. Manion, MS,[3] Ulysses J. Balis, MD,[4] Kai Zheng, PhD[2,5]
[1]Department of Pediatrics; [2]School of Information; [3]Comprehensive Cancer Center; [4]Department of Pathology; [5]School of Public Health. Univ. of Michigan, Ann Arbor, MI

Abstract

In this study, we quantified the use of uncertainty expressions, referred to as 'hedge' phrases, among a corpus of 100,000 clinical documents retrieved from our institution's electronic health record system. The frequency of each hedge phrase appearing in the corpus was characterized across document types and clinical departments. We also used a natural language processing tool to identify clinical concepts that were spatially, and potentially semantically, associated with the hedge phrases identified. The objective was to delineate the prevalence of hedge phrase usage in clinical documentation which may have a profound impact on patient care and provider–patient communication, and may become a source of unintended consequences when such documents are made directly accessible to patients via patient portals.

Introduction

Uncertainties pervade medicine.[1] They originate in multiple sources ranging from clinicians' course of investigation for pinpointing the right diagnoses[2,3] to patients' pondering whether to start a medical treatment.[4] Expressions found in clinicians' written or oral communications describing such uncertainties are known as 'hedging,' herein defined as "any term or phrase that is ambiguous or lacks clear precision." Hedging may take multiple forms such as those related to probability (e.g., likely, probable, possible), frequency (e.g., often, occasionally, sometimes), or quantity (e.g., many, few, some). These concepts have also been described as being 'underspecified' in relation to their use in clinical practice guidelines.[5]

Numerous attempts have been made to study clinicians' use of hedging phrases, particularly with respect to probability expressions. Some early studies, for example, investigated how clinicians interpret the meaning of hedge expressions and how this interpretation might in turn affect intra-clinician communication and distributed medical decision-making.[6-14] Recent studies explored the interpretation of probability expressions in clinical documents such as pathology reports[15-17] and radiology reports.[18,19]

Similar investigations have also been conducted to study how patients and families interpret the meaning of hedge phrases.[12,20-25] These studies may have been driven by the medical establishment's growing acknowledgement of the importance of information transparency and patient autonomy. A recent survey study, for example, assessed if physicians and patients interpret hedge phrases differently. The results suggested that a great deal of variability exists with respect to how physicians and patients associate quantitative values to hedge phrases differently.[26]

It has been demonstrated that uncertainty expressions used by clinicians could result in negative patient perceptions during patient–provider encounters.[27,28] Nonetheless, such encounters are almost always conducted through face-to-face (or telephone) contact in which hedge phrases used could be timely clarified during the interaction. Written notes in patient charts, on the other hand, have been traditionally purposed to record data to support intra-clinician communications not intended to be directly viewable by patients. This premise could soon change, however, with an increasing recognition of the value of making clinician notes available to patients to improve information transparency. While most commercial

electronic health record (EHR) systems share primarily structured, non-narrative data through patient portals, the OpenNotes Project (http://www.myopennotes.org/) has been experimenting with the concept of letting patients and families have direct access to their data including clinician notes.[29-31] In such scenarios, clinicians' 'hedging' language is directly exposed to patients, which could result in both confusion and negative patient perceptions.

Therefore, while often justified due to the inherent uncertainties in medicine, the use of hedge phrases, combined with medical 'lingoes' speed-written (e.g., "mets" for describing cancer "metastases"), could create confusion among clinicians, and could become even more problematic when the communication is conveyed between healthcare providers and their patients. In this study, we sought to take a fresh, comprehensive look at clinicians' use of hedge phrases in an EHR. We first collected a wide range of hedge phrases drawn from the literature, and determined the frequency of these terms appearing in a corpus of clinical documents retrieved from an EHR. We then applied natural language processing (NLP) to identify key clinical concepts to assess how often they appeared in association with the hedge phrases. The results could provide valuable insights into understanding clinicians' use of uncertainty expressions in clinical documents and its potential implications when sharing such documents with patients.

Methods

A. Literature Review and List Compilation

We first conducted a literature search in PubMed to identify published studies that had described hedge phrases used in clinical documentation and communication. Keywords searched included "hedging," "uncertainty," "ambiguous," and "probability expressions" (and their alternative spellings), in combination with relevant terms such as "clinical," "pathology reports," and "medical records." All hedge phrases mentioned in the papers we found were extracted. In addition, we developed variants based on the phrases identified. For example, "can't rule out" was reported in two papers[15 16] while "unable to rule out" was never mentioned; the latter was deemed as a reasonable alternative and was added to our list.

B. Empiric Dataset

In this study, we analyzed a dataset consisting of 100,000 clinical documents retrieved randomly from our institution's EHR system. They belonged to hematology/oncology patients whose decedent status was double confirmed using our EHR and tumor registry. These documents represent a wide variety of note types ranging from admission notes to discharge summaries. All documents were de-identified prior to analysis using the MITRE Identification Scrubber Toolkit (MIST).[32] The use of these documents was reviewed and determined to be exempt as nonhuman subject research by our institutional review board.

C. Frequencies of Hedge Phrases

First, we split the 100,000 documents into a total of 3,863,418 sentences. Then, we identified those sentences containing at least one hedge phrase. Regular expressions were used to identify the maximum possible match of these phrases. In other words, the phrase "evidence of" would not be double counted if a longer phrase "no evidence of" is identified.

Because the word "may" was considered to be ambiguous beyond its hedging connotation ("May" can also be a month of the year), we excluded that term from our analysis when it was capitalized unless it was the first word of a sentence. Document metadata were included in the analysis to examine differences in hedge phrase usage across document type, clinical department, and data entry method. Percentages of hedge phrase usage between departments (reported in Table 4) were compared using a '2-sample test for equality of proportions with continuity correction' in *R* version 2.13.2 for Mac OS X.

D. Concept Identification using Natural Language Processing

We used the MetaMap software released by the National Library of Medicine in 2011 to determine the types of clinical concepts frequently associated with the hedging phrases.[33] [34] MetaMap is an NLP software program that identifies clinical concepts and maps them to other relevant concepts contained in the Unified Medical Language System (UMLS). The sentences containing at least one hedge term, as previously identified, were processed using MetaMap, and all phrases in the sentence that mapped to one or more UMLS concepts were identified and mapped to a specific semantic type. Only those phrases with a score of 1000, i.e., a perfect match, were retained to help ensure reliable results. For cases in which a phrase had more than one semantic type with a perfect score, both were retained.

Then, we excluded all mapped phrases that were beyond a window size of 4 with respect to the hedge phrase. This window size was chosen because it had been shown to be ideal in work related to assertion classification.[35] Finally, we summarized the frequency of the semantic types that were associated with the hedge phrases. Negation by MetaMap was not considered in this analysis since many of our terms were themselves negation terms (e.g., "no evidence for," "does not appear," "definitely not").

Results

A total of 31 papers were identified through the literature search,[6-20] [22-26] [35-45] which provided a list of 313 distinct hedge phrases (available upon request). An additional 28 terms were added, most of which represented slight variations from the phrases in the literature (e.g,, including 'not much of a chance' based on the published phrase 'not much chance'). Among the 313 phrases directly drawn from the literature, thirty-three appeared in 4 or more papers, shown in Table 1.

Most of the papers quantified the meaning of expressions through the use of surveys and interviews. Among them, we found a lot of variability in the approaches used to perform the quantification and/or in the manner in which the results were reported. An example showing the results of studying the word 'probable' is provided in Figure 1. This figure displays visual comparisons of the results and reporting approaches from each study, and also demonstrates the wide range of probabilities assigned by subjects.

Table 1. The top 33 hedge phrases most frequently discussed in the literature

Hedge Phrase	Citations	Hedge Phrase	Citations
possible	17	improbable	5
unlikely	14	always	5
likely	14	rare	5
probable	13	probably	5
certain	12	consistent with	5
never	10	possibly	5
sometimes	9	not unreasonable	4
rarely	8	doubtful	4
most likely	8	not certain	4
very likely	7	low probability	4
often	7	usually	4
almost never	6	suggestive	4
almost certain	6	diagnostic	4
perhaps	6	not likely	4
occasionally	6	low chance	4
frequent	6	likely	4
compatible with	6		

Among our list of 341 hedge phrases, 265 appeared in the corpus at least once. Examples of phrases that did not appear even once include 'not compatible with', 'more often than not', and 'moderate probability.' The 30 hedge phrases that appeared in the largest number of documents are listed in Table 2. Among the 30 most frequently used hedge phrases discovered in this study, only 10 were commonly reported in the literature. The rest were mentioned in 3 or fewer publications.

Table 3 lists the 20 most frequently occurring document types in our sample, representing 82.8% of all documents we analyzed. Nearly all (98.1%) of 'Admission History & Physical' documents contained at least one hedge phrase whereas the nursing notes (both standard and procedure notes) contained far fewer hedge phrases (31.5% and 13.5%, respectively). Other nursing notes not shown in the table also contained few hedge phrases. For example, only 14.4% of the 'Nursing Progress Note' documents contained a hedge phrase. In Table 4 are some of the clinical departments in our health system and the percentage of documents containing hedge phrases created by each department. Notes generated by the Department of Psychiatry used many hedge phrases whereas Nursing did not, consistent with the note types in Table 3.
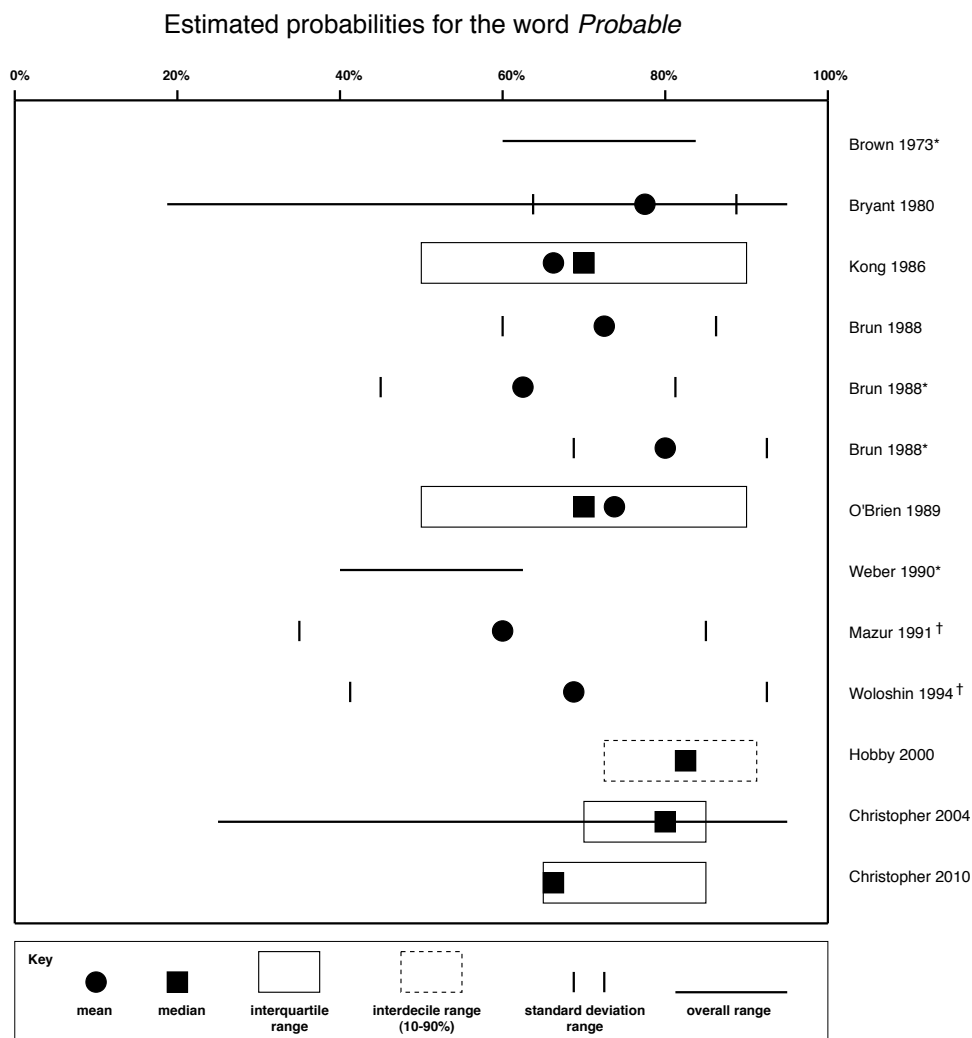

Figure 1. The estimated probabilities for the word "probable" as reported in the literature. Those marked with an asterisk (*) obtained probabilities from non-clinicians in a non-medical context; studies marked with a cross (†) involved patients; the rest queried clinicians for their probability estimates. The paper by Brun presented results from multiple studies, 3 of which are shown in the graph to demonstrate the variability reported.

Table 2. The top 30 hedge phrases most frequently appearing in the study corpus

Hedge Phrase	Total number of Documents	Hedge Phrase	Total number of Documents
may	24,036	could be	4,739
possible*	23,002	most likely*	4,625
likely*	21,307	appear	4,301
positive	21,126	necessary	4,220
several	14,737	seems	3,882
no evidence of	13,283	probably*	3,836
evidence of	12,350	frequent*	3,580
most	12,293	never*	3,419
consistent with*	11,189	many	3,368
unremarkable	10,374	sure	3,368
few	9,769	suggest	3,328
usual	6,174	apparently	3,302
think	5,352	occasionally*	3,269
possibly*	5,350	possibility of	3,046
potential	5,116	diagnostic*	2,813

* Phrases most frequently discussed in the literature.

Table 3. The top 20 most frequent document types, sorted by decreasing frequency of hedge phrases

Document Type	Total number of Documents	Proportion with ≥ 1 Hedge Phrase (%)
Admission History & Physical	1,879	98.08
Discharge Summary	4,753	93.18
Emergency Department Note	2,441	89.55
Letter	6,197	89.43
Consult – Inpatient – New	1,875	88.16
Note*	7,501	85.76
Consult – Inpatient	1,720	85.58
Letter/Note – Return Visit	18,509	85.53
Letter/Note – New Patient	2,558	82.13
Endoscopy Report	1,053	81.86
Progress Note	15,198	80.87
Social Work Note	1,865	79.03
Operative Report	2,412	78.57
Consult – Inpatient- Follow-up	4,462	77.63
Nutrition Note	2,199	74.17
Procedure Note	1,548	57.95
Phone Note	3,730	44.69
Physical Therapy Inpatient Note	1,121	38.45
Nursing Note	1,772	31.49
Nursing Procedure note	3,266	13.49

* 'Note' represents a generic document type initially used in the EHR before more specific types were created.

Table 5 shows the top 30 UMLS semantic types associated with a hedge phrase identified by MetaMap. Many of these high level categories are clinically relevant including 'Disease or Syndrome,' 'Finding,' and 'Sign or Symptom.' Figure 2 displays a network graph showing the relationships between the most common semantic types identified by MetaMap and the most frequent hedge phrases associated with

those semantic types. The term 'likely' was associated with 9 of the top 10 semantic types, followed by 'may' and 'possible' which were each associated with 8 of them. Table 6 provides examples of hedge phrases associated with clinical concepts identified with our approach, drawn from our document corpus.

In a prior study using a subset of the corpus, we showed that there are linguistic property differences in documents that were typed versus dictated/transcribed that could influence the performance of NLP systems.[46] The current study included 47,665 documents that were typed by clinicians and 19,882 documents created via dictation/transcription. The remainder of the documents were added from other source systems but we could not ascertain how they were created based on the metadata. Among the typed notes, 66.52% had at least one hedge phrase, whereas 86.81% of the dictated notes had at least one phrase, and this difference was highly significant (p < 0.001).

Table 4. Hedge phrase usage by clinical departments

Group	Clinical Department	Total number of Documents	Proportion with ≥ 1 Hedge Phrase (%)	P value for proportion test
1	Psychiatry	291	93.47	
	Emergency Medicine	3,737	91.06	0.20
	Neurology	1,833	90.67	0.67
2	Pediatrics	4,463	83.87	2.6×10^{-12}
	Internal Medicine	35,943	83.51	0.56
	Family Medicine	2,177	83.28	0.80
3	Obstetrics and Gynecology	1,049	79.03	3.9×10^{-3}
	Dermatology	1,057	76.44	0.17
	Ophthalmology	851	75.91	0.83
4	Orthopedic Surgery	603	70.65	2.9×10^{-2}
	General Surgery	9,954	70.60	1.0
	Neurosurgery	1,005	68.76	0.24
5	Anesthesiology	1,290	65.35	0.09
6	Physical Medicine & Rehabilitation	5,967	56.39	4.2×10^{-9}
7	Nursing	6,817	33.93	$< 2.2 \times 10^{-16}$

P values in rightmost column are based on proportion test between the row in which the value appears and the preceding row. Departments with similar percentages are grouped according to shading.
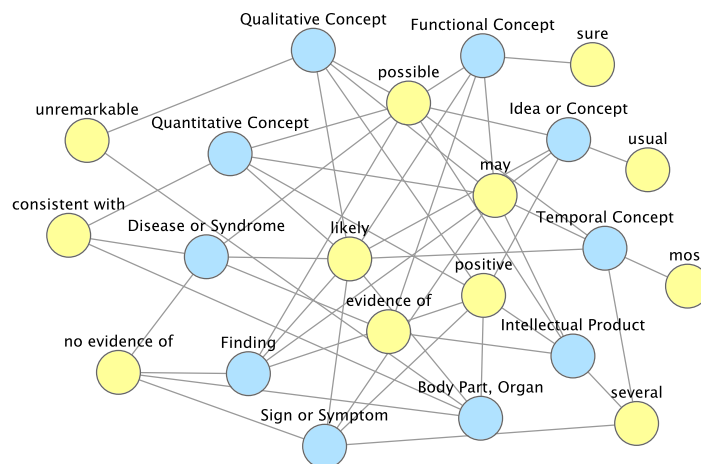

Figure 2. Network graph depicting the association between the 10 most common semantic types (blue nodes) and 12 of the most frequently used hedge phrases (yellow nodes).

Table 5. UMLS semantic types associated with hedge phrases, sorted by decreasing frequency

UMLS Semantic Type	Freq.	UMLS Semantic Type	Freq.
Functional Concept	45,637	Mental Process	7,589
Qualitative Concept	39,159	Conceptual Entity	7,238
Disease or Syndrome	34,785	Health Care Activity	6,998
Finding	33,024	Spatial Concept	6,851
Idea or Concept	22,607	Diagnostic Procedure	6,221
Temporal Concept	20,149	Organic Chemical, Pharm. Substance	5,186
Sign or Symptom	13,718	Body Location or Region	5,022
Quantitative Concept	13,380	Medical Device	4,591
Body Part, Organ, or Organ Component	12,422	Amino Acid, Protein, Immunol. Factor	4,579
Intellectual Product	11,120	Inorganic Chemical	4,475
Pathologic Function	10,902	Clinical Attribute	4,255
Patient or Disabled Group	10,254	Organism Function	3,845
Activity	10,085	Body Substance	3,232
Therapeutic or Preventive Procedure	9,608	Laboratory Procedure	3,163
Neoplastic Process	8,576	Geographic Area	3,023

Table 6. Examples of complete hedge phrases drawn from the corpus. All hedges phrases are in bold. Concepts identified by MetaMap (within the window of +/- 4) are italicized.

It is also **possible**, although **less likely**, that the *synechia* formed following nasal trauma...
I can find **little evidence** that the *thyroid* is the source of her cancer.
…hand pain and *numbness* which **may be associated** with *arthritis* with some nerve entrapment.
Possible but **inconclusive** source of *melena*.
The etiology of the patient's *fatigue* is **not clear** at this *time*.
He had normal range of *affect* and with **perhaps** some *mild anxiety*.
… **somewhat unlikely** that a *shunt occlusion* or migration would be the cause of her altered mental status
He continues to complain of insomnia and *jitteriness*, which I **suspect** is due to *prednisone*

Discussion

It is not surprising that clinicians' frequent use of hedge phrases has continued into the era of EHR. With the availability of a large-size corpus, our study was able to further quantify several nuances in hedge phrase utilization. For example, the results show that physicians express uncertainties far more frequently than nurses. While hedge phrases are present in a variety of documents, they appear in distinct rates.

Among the semantic types detected using MetaMap, many of them are clinically relevant. There were nearly 35,000 instances in which a hedge phrase was associated with a 'Disease or Syndrome.' Such instances must be carefully communicated to patients as it has been demonstrated that "in situations where there is substantial uncertainty, extra vigilance is required to ensure that patients are given the tools and information they need to participate in cooperative decision making about their care."[47]

It should be noted that disclosing uncertainty to patients, rather than concealing it, can yield better results. It has been shown that educating patients regarding the limitations of medicine (such as the inherent uncertainty in diagnoses and prediction of disease progression) can improve clinician–patient communication as well as increase patient involvement and understanding of their illness.[48] For example, a study conducted in a general medicine clinic found that physicians who expressed uncertainty and educated patients about it received higher patient satisfaction scores.[49] In another study setting, similar results were reported that clinicians' willingness to disclose and openly discuss uncertainty with patients

led to better patient satisfaction.[50] Nonetheless, with a growing interest in making clinical data more accessible to patients via patient portals, clinical documents containing hedge phrases may now be directly exposed to patients without adequate clarification, which could result in unintended adverse consequences. Specific strategies, therefore, need to be developed to better handle hedge phrases in patient-facing scenarios. Such strategies may include providing computerized alerting mechanisms to detect certain hedge phrase usage that may be problematic, or including contextual educational materials that provide explanations to patients about the cause of the uncertainty and its implications.[42 44 51-54] Involving patients in developing strategies for handling these commonly used phrases will be important.

This study has several limitations. First, we quantified the multitude of hedge phrase usage without manually reviewing each instance to verify what the context was and whether the use was appropriate. For example, we did see that the word "may" also appeared in non-ambiguous contexts related to its meaning of 'permission' such as "I explained that he may call back at any time with further questions." Adding a qualitative component to the study wherein patients and physicians interpreted a selection of hedge phrases from our corpus would provide additional data about the use and meaning of the phrases. Furthermore, our analysis was conducted based on data collected in a single institution and we were therefore unable to study local differences in how hedge phrases are used or accepted which limits the generalizability of our findings. Finally, the association of a semantic entity and a hedge phrase was determined approximately; the hedge phrase appearing within 4 words of a semantic entity might not have necessarily been used to modify the very entity. Further work with manual verification would help to improve the precision of the analysis.

Conclusion

We empirically evaluated clinicians' use of hedge phrases in their clinical documentation by examining the frequency of these phrases and their association with clinical concepts identified through natural language processing. We found that hedge phrases are used in a substantial proportion and variety of clinical documents, and many are associated with clinically relevant concepts. The use of hedge phrases in clinical notes has the potential for unintended consequences as patients begin to gain direct access to their notes through patient portals.

Acknowledgements

References

1. **Wellbery C**. The value of medical uncertainty? *Lancet* 2010;375(9727):1686-7.
2. **Guenter D**, Fowler N, Lee L. Clinical uncertainty: helping our learners. *Can Fam Physician* 2011;57(1):120-5.
3. **Swoboda DA**. Negotiating the diagnostic uncertainty of contested illnesses: physician practices and paradigms. *Health (London)* 2008;12(4):453-78.
4. **Fraenkel L**. Uncertainty and patients' preferred role in decision making. *Patient Educ Couns* 2011;82(1):130-2.
5. **Codish S**, Shiffman RN. A model of ambiguity and vagueness in clinical practice guideline recommendations. *AMIA Annu Symp Proc* 2005:146-50.
6. **Bryant GD**, Norman GR. Expressions of probability: words and numbers. *N Engl J Med* 1980;302(7):411.

7. **Kenney RM**. Between never and always. *N Engl J Med* 1981;305(18):1097-8.
8. **Kong A**, Barnett GO, Mosteller F, Youtz C. How medical professionals evaluate expressions of probability. *N Engl J Med* 1986;315(12):740-4.
9. **Mapes RE**. Verbal and numerical estimates of probability in therapeutic contexts. *Soc Sci Med* 1979;13A(3):277-82.
10. **Nakao MA**, Axelrod S. Numbers are better than words. Verbal specifications of frequency have no place in medicine. *Am J Med* 1983;74(6):1061-5.
11. **O'Brien BJ**. Words or numbers? The evaluation of probability expressions in general practice. *J R Coll Gen Pract* 1989;39(320):98-100.
12. **Ohnishi M**, Fukui T, Matsui K, Hira K, Shinozuka M, Ezaki H, et al. Interpretation of and preference for probability expressions among Japanese patients and physicians. *Fam Pract* 2002;19(1):7-11.
13. **Timmermans D**. The roles of experience and domain of expertise in using numerical and verbal probability terms in medical decisions. *Med Decis Making* 1994;14(2):146-56.
14. **Toogood JH**. What do we mean by "usually"? *Lancet* 1980;1(8177):1094.
15. **Christopher MM**, Hotz CS. Cytologic diagnosis: expression of probability by clinical pathologists. *Vet Clin Pathol* 2004;33(2):84-95.
16. **Christopher MM**, Hotz CS, Shelly SM, Pion PD. Interpretation by clinicians of probability expressions in cytology reports and effect on clinical decision-making. *J Vet Intern Med* 2010;24(3):496-503.
17. **Galloway M**, Taiyeb T. The interpretation of phrases used to describe uncertainty in pathology reports. *Patholog Res Int* 2011;2011:656079.
18. **Khorasani R**, Bates DW, Teeger S, Rothschild JM, Adams DF, Seltzer SE. Is terminology used effectively to convey diagnostic certainty in radiology reports? *Acad Radiol* 2003;10(6):685-8.
19. **Hobby JL**, Tom BD, Todd C, Bearcroft PW, Dixon AK. Communication of doubt and certainty in radiological reports. *Br J Radiol* 2000;73(873):999-1001.
20. **Biehl M**, Halpern-Felsher BL. Adolescents' and adults' understanding of probability expressions. *J Adolesc Health* 2001;28(1):30-5.
21. **Fuller R**, Dudley N, Blacktop J. Risk communication and older people-understanding of probability and risk information by medical inpatients aged 75 years and older. *Age Ageing* 2001;30(6):473-6.
22. **Mazur DJ**, Merz JF. Patients' interpretations of verbal expressions of probability: implications for securing informed consent to medical interventions. *Behav Sci Law* 1994;12(4):417-26.
23. **Shaw NJ**, Dear PR. How do parents of babies interpret qualitative expressions of probability? *Arch Dis Child* 1990;65(5):520-3.
24. **Woloshin KK**, Ruffin MTt, Gorenflo DW. Patients' interpretation of qualitative probability statements. *Arch Fam Med* 1994;3(11):961-6.
25. **Mazur DJ**, Hickam DH. Patients' interpretations of probability terms. *J Gen Intern Med* 1991;6(3):237-40.
26. **Foppa M**, Schneider de Araujo B, Macari A, Reichert R, Goldim JR. Limitations in the use of qualitative terms to inform diagnoses. *Arch Intern Med* 2011;171(14):1291-2.
27. **Blanch DC**, Hall JA, Roter DL, Frankel RM. Is it good to express uncertainty to a patient? Correlates and consequences for medical students in a standardized patient visit. *Patient Educ Couns* 2009;76(3):300-6.
28. **Ogden J**, Fuks K, Gardner M, Johnson S, McLean M, Martin P, et al. Doctors expressions of uncertainty and patient confidence. *Patient Educ Couns* 2002;48(2):171-6.
29. **Delbanco T**, Walker J, Darer JD, Elmore JG, Feldman HJ, Leveille SG, et al. Open notes: doctors and patients signing on. *Ann Intern Med* 2010;153(2):121-5.
30. **Vogel L**. OpenNotes Project "levels the playing field" between doctors and patients. *CMAJ* 2010;182(14):1497-8.
31. **Walker J**, Leveille SG, Ngo L, Vodicka E, Darer JD, Dhanireddy S, et al. Inviting patients to read their doctors' notes: patients and doctors look ahead: patient and physician surveys. *Ann Intern Med* 2011;155(12):811-9.

32. **Aberdeen J**, Bayer S, Yeniterzi R, Wellner B, Clark C, Hanauer D, et al. The MITRE Identification Scrubber Toolkit: design, training, and assessment. *Int J Med Inform* 2010;79(12):849-59.

33. **Aronson AR**. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001:17-21.

34. **Aronson AR**, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 2010;17(3):229-36.

35. **Uzuner O**, Zhang X, Sibanda T. Machine learning and rule-based approaches to assertion classification. *J Am Med Inform Assoc* 2009;16(1):109-15.

36. **Beyth-Marom R**. How Probable is Probable?A Numerical Translation of Verbal Probability Expressions. *Journal of Forecasting* 1982;1:257-69.

37. **Brown TA**, Shuford EH. Quantifying Uncertainty Into Numerical Probalities for the Reporting of Intelligence. Rand Report R-1185-AR. July 1973.

38. **Brun W**, Teigen KH. Verbal Probabilities: Ambiguous, Context-Dependent, or Both? *Organizational Behavior and Human Decision Processes* 1988;41:390-404.

39. **Druzdel MJ**. Verbal Uncertainty Expressions: Literature Review. Technical Report CMU-EPP-1990-03-02, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburg, PA, May 1989.

40. **Elkin PL**, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, et al. A controlled trial of automated classification of negation from clinical notes. *BMC Med Inform Decis Mak* 2005;5:13.

41. **Farkas R**, Vincze V, Móra G, Csirik J, Szarvas G. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning: Shared Task*. Uppsala, Sweden, 2010:1-12.

42. **Light M**, Qiu XY, Srinivasan P. The Language of Bioscience: Facts, Speculations, and Statements in Between. In: Hirschman L, Pustejovsky J, editors. *HLT-NAACL 2004 Workshop: BioLINK 2004, Linking Biological Literature, Ontologies and Databases*, 2004:17-24.

43. **McGlone MS**, Reed AB. Anchoring in the interpretation of probability expressions. *Journal of Pragmatics* 1998;30(6):723-33.

44. **Szarvas G**. Hedge Classification in Biomedical Texts with a Weakly Supervised Selection of Keywords. *Proceedings of ACL-08*, 2008:281-89.

45. **Weber EU,** Hilton DJ. Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance* 1990;16(4):781-89.

46. **Zheng K**, Mei Q, Yang L, Manion FJ, Balis UJ, Hanauer DA. Voice-dictated versus typed-in clinician notes: linguistic properties and the potential implications on natural language processing. *AMIA Annu Symp Proc* 2011;2011:1630-8.

47. **Parascandola M**, Hawkins J, Danis M. Patient autonomy and the challenge of clinical uncertainty. *Kennedy Inst Ethics J* 2002;12(3):245-64.

48. **Henry MS**. Uncertainty, responsibility, and the evolution of the physician/patient relationship. *J Med Ethics* 2006;32(6):321-3.

49. **Gordon GH**, Joos SK, Byrne J. Physician expressions of uncertainty during patient encounters. *Patient Educ Couns* 2000;40(1):59-65.

50. **Johnson CG**, Levenkron JC, Suchman AL, Manchester R. Does physician uncertainty affect patient satisfaction? *J Gen Intern Med* 1988;3(2):144-9.

51. **Agarwal S**, Yu H. Detecting hedge cues and their scope in biomedical text with conditional random fields. *J Biomed Inform* 2010;43(6):953-61.

52. **Lewin BA**. Hedging: an exploratory study of authors' and readers' identification of 'toning down' in scientific texts. *Journal of English for Academic Purposes* 2005;4(2):163-78.

53. **Velldal E**. Predicting speculation: a simple disambiguation approach to hedge detection in biomedical literature. *J Biomed Semantics* 2011;2 Suppl 5:S7.

54. **Vincze V**, Szarvas G, Farkas R, Mora G, Csirik J. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* 2008;9 Suppl 11:S9.