

How certain are radiologists - do we need a standardized lexicon for describing certainty in radiology reports?

Paras Lakhani

Baskaran Sundaram

Christopher Roth

Adam Flanders

Thomas Jefferson University, Philadelphia, PA

Financial Disclosures

- All Authors: Nothing to Disclose

Purpose

- The goal of this study is to assess the degree of certainty of commonly used terms in radiology reports among radiologists.

Background

- Radiologists use a wide variety of terms to describe certainty in reports.
- Some examples include:
 - “Suspicious for”
 - “Suggestive of”
 - “May represent”
 - “Worrisome for”
 - “Cannot exclude”

Background

- Khorasani et al. surveyed radiologists and referring clinicians and found poor agreement in ranking 12 commonly used terms.¹
- However, it is difficult to rank a high number (12 terms) and get consistency.
- Standardization of terminology should improve consistency of reporting.

Background

- Panicek et al. proposed a 5 point standardized scale for describing certainty.

<u>Term</u>	<u>Estimated certainty (%)</u>
<i>Consistent with</i>	>90%
<i>Suspicious for/probable</i>	~75%
<i>Possible</i>	~50%
<i>Less likely</i>	~25%
<i>Unlikely</i>	<10%

Hypothesis

- When provided a 5-point scale similar to that proposed by Panicek et al., radiologists are more consistent in assigning terms to describe certainty.

Methods

- Performed in a large teaching hospital.
- The authors identified 26 radiology terms to convey certainty
- Surveyed 107 radiologists (57 attendings, 10 fellows, and 40 residents).
- Radiologists asked to categorize terms using a Likert scale:
 - 5) Very high probability or diagnostic of (>90%)
 - 4) High probability (~75%)
 - 3) Intermediate probability (~50%)
 - 2) Low probability (~25%)
 - 1) Very low probability (<10%).
- Mean and standard deviation values were obtained. A two-tailed t-test was used for statistical significance.

Results

- Survey response rate was 54/107 (50.5%).
- Terms scored as indicating the **highest (>90%)** probability of certainty included:

	Term	Score, SD
1.	"diagnostic of"	(4.95±0.21)
2.	"represents"	(4.88±0.32)
3.	"highly suspicious"	(4.81±0.39)
4.	"consistent with"	(4.60±0.66)
5.	"in keeping with"	(4.60±0.62)
6.	"highly suggestive"	(4.58±0.50)
7.	"highly concerning"	(4.53±0.59)
8.	"highly worrisome"	(4.53±0.55)
9.	"most likely"	(4.49±0.55)
10.	"compatible with"	(4.44±0.67).

Results

- Terms scored as indicating a **high (~75%)** probability of certainty included:

	<u>Term</u>	<u>Score, SD</u>
1.	"suspicious"	(3.84±0.43)
2.	"concerning"	(3.74±0.54)
3.	"probably"	(3.70±0.56)
4.	"likely"	(3.67±0.52)
5.	"worrisome"	(3.67±0.57)
6.	"suggestive"	(3.65±0.57)

Results

- Terms scored as indicating an **intermediate (~50%)** probability of certainty included:

	Term	Score, SD
1.	"may represent"	(3.07±0.46)
2.	"could represent"	(2.95±0.43)
3.	"equivocal"	(2.88±0.50)
4.	"possibly"	(2.77±0.61)
5.	"maybe"	(2.77±0.48)
6.	"question of"	(2.70±0.56)

Results

- Terms scored as indicating a **low (~25%)** or **very low (<10%)** probability of certainty included:

	Term	Score, SD
1.	"unlikely"	(1.84±0.37)
2.	"not excluded"	(1.56±0.74)
3.	"inconsistent with"	(1.33±0.48)
4.	"very unlikely"	(1.00±0.00).

Results

- Statistical Significance:
 - The differences between the responses as grouped above were statistically significant ($P < 0.001$).

Conclusions

- When provided a 5 point Likert scale, radiologists are fairly consistent in rating a wide variety of terms describing degree of certainty.
- This has implications for natural language processing systems, which can convert equivalent terms to those matching a standardized lexicon or rating system.



Thank you!