A method to quantify residents' jargon use during counseling of standardized patients about cancer screening

Lindsay Deuster
Medical College of Wisconsin

Stephanie Christopher
Marquette University, stephanie.christopher@marquette.edu

Jodi Donovan

Michael Farrell
Medical College of Wisconsin

Follow this and additional works at: https://epublications.marquette.edu/cps_fac

Recommended Citation
Deuster, Lindsay; Christopher, Stephanie; Donovan, Jodi; and Farrell, Michael, "A method to quantify residents' jargon use during counseling of standardized patients about cancer screening" (2008). College of Professional Studies Faculty Research and Publications. 11.
https://epublications.marquette.edu/cps_fac/11
A Method to Quantify Residents’ Jargon Use During Counseling of Standardized Patients About Cancer Screening

Lindsay Deuster
Center for Patient Care and Outcomes Research, Medical College of Wisconsin, Milwaukee, WI

Stephanie Christopher
Center for Patient Care and Outcomes Research, Medical College of Wisconsin, Milwaukee, WI

Jodi Donovan
Center for Patient Care and Outcomes Research, Medical College of Wisconsin, Milwaukee, WI

Michael Farrell
Center for Patient Care and Outcomes Research, Medical College of Wisconsin, Milwaukee, WI

Internal Medicine, Pediatrics, and Population Health-Bioethics, Center for Patient Care and Outcomes Research, Milwaukee, WI
Abstract

Background
Jargon is a barrier to effective patient-physician communication, especially when health literacy is low or
the topic is complicated. Jargon is addressed by medical schools and residency programs, but reducing
jargon usage by the many physicians already in practice may require the population-scale methods used
in Quality Improvement.

Objective
To assess the amount of jargon used and explained during discussions about prostate or breast cancer
screening. Effective communication is recommended before screening for prostate or breast cancer
because of the large number of false-positive results and the possible complications from evaluation or
treatment.

Participants
Primary care internal medicine residents.

Measurements
Transcripts of 86 conversations between residents and standardized patients were abstracted using an
explicit-criteria data dictionary. Time lag from jargon words to explanations was measured using
“statements,” each of which contains one subject and one predicate.

Results
Duplicate abstraction revealed reliability $\kappa = 0.92$. The average number of unique jargon words per
transcript was 19.6 (SD = 6.1); the total jargon count was 53.6 (SD = 27.2). There was an average of 4.5
jargon-explanations per transcript (SD = 2.3). The ratio of explained to total jargon was 0.15. When
jargon was explained, the average time lag from the first usage to the explanation was 8.4 statements
(SD = 13.4).

Conclusions
The large number of jargon words and low number of explanations suggest that many patients may not
understand counseling about cancer screening tests. Educational programs and faculty development
courses should continue to discourage jargon usage. The methods presented here may be useful for
feedback and quality improvement efforts.

Keywords
Cancer Screening; Health Literacy; Breast Cancer Screening; Standardize Patient; Prostate Cancer
Screening

The interpersonal process is the vehicle by which technical care is implemented and on which its
success depends...why is it so often ignored in assessments of the quality of care?...the criteria
and standards that permit precise measurement of the attributes of the interpersonal process
are not well developed or have not been sufficiently called upon to undertake the task.

Avedis Donabedian¹
INTRODUCTION
Communication is said to be the “main ingredient” of medical care, but researchers have found communication problems across the health-care system. One of the most commonly mentioned is physicians’ use of jargon, which is defined as the “specialized language of a trade, profession, or similar group” that is “not likely to be easily understood by persons outside the profession”. In this paper we present a method designed to be suitable for quantifying clinicians’ jargon usage over a statewide population and use the method to describe internal medicine residents’ use of jargon during counseling about screening for prostate or breast cancer.

We chose to study communication about prostate and breast cancer screening because professional organizations recommend routine counseling about potential risks of screening. Counseling is recommended for prostate screening because there is no high-level evidence to suggest that prostate-specific antigen decreases morbidity or mortality. Counseling recommendations for breast cancer screening were instituted because of differences in the level of evidence for mammography in women >50 years of age versus 40–49.

This analysis is part of a larger effort to develop communication assessment tools that will be suitable for use on a population scale, such as an institution, health plan, or entire state. Population-scale efforts will be important because communication problems can occur anywhere. Medical schools and residency programs discourage jargon during training, but it is unclear how long the benefits of training persist. Faculty development programs help practicing physicians, but these programs require more time, effort, and money than many physicians seem willing or able to commit. To be successful on a population scale, new communication assessment methods will need to be quantitatively reliable, concrete enough for use by quality improvement professionals, and able to function on a lean budget. To meet these requirements in previous analyses of communication behaviors, we have adapted a “communication quality indicator” approach from Quality Improvement. We also adapted methods from corpus linguistics (the study of language as found in collections of text used in the real world) in order to quantify jargon usage beyond the ordinal scale approaches used in previous research.

METHODS
Data Source
We used an explicit-criteria procedure to abstract transcripts of conversations between internal medicine residents and standardized patients portrayed to have a question about screening for prostate or breast cancer. Transcripts were made from tapes collected during four workshops in a Primary Care Internal Medicine residency program. The workshops were part of the educational curriculum. Residents were asked to give informed consent and were offered a chance to decline use of their tapes for research. Methods were approved by institutional review boards at Yale and the Medical College of Wisconsin.

Residents were taped in two standardized patient encounters before the didactic portion of the workshops. In one encounter, a 50-year-old man asked about prostate cancer screening; in the other encounter, a 43-year-old woman asked about breast cancer screening. The order of the two encounters was random for each resident. A fact sheet stated that the patient had no family history of cancer and
had had an unremarkable physical exam the week before, so the resident would not feel obliged to do a physical or take an extended history. Patients began with a short speech patterned after the following example:

I’m sorry I’m back so soon after my physical, but I had to leave so quickly that I didn’t get a chance to ask a question. I recently saw an advertisement about prostate [or breast] cancer screening, but I wasn’t sure if it was for me. What do you think?

To standardize the counseling task, patients were coached to avoid asking leading questions and to minimize the appearance of anxiety or confusion. This strategy allowed us to focus on jargon usage rather than residents’ ability to respond to questions (an important skill, but not the subject of this study).

Tapes were transcribed verbatim and proofread for accuracy by a board-certified internist (MF and Dr. Jeffrey Stein). The proofreaders deleted personally identifying text about the residents to reduce abstractor bias. To provide a content-related unit of duration, we used a sentence-diagramming procedure to parse transcripts into individual “statements,” which were defined as having one explicit or implied subject and one predicate. The approach was selected to be consistent with theories and research from cognitive psychology about the mental demands of holding several unfamiliar concepts in mind at the same time.\textsuperscript{26,27,28,29} Statements were also simpler for abstractors than the widely used utterance approach, which looks for individual concepts as well, but also parses speech at points such as 1 second pauses, tonal changes, speaker emphases, and some conjunctions.\textsuperscript{30} Examples of parsed text are shown in Figures 1 and 2, with each statement beginning at a number in pointed brackets \{\} and ending with a double slash //. A single slash/ or a quadruple slash //// is used to mark the middle or end of compound statements.

\textbf{Figure 1.} Example of simple use of the sentence parsing method and explanatory time lag calculation.
The final sample for analysis consisted of 86 transcripts (41 for prostate cancer screening and 45 for breast cancer screening).

Abstraction

Abstraction was facilitated by an explicit-criteria data dictionary that included jargon word lists and definitions to use with jargon explanations. Abstractors read the transcripts statement by statement and compared each word with the jargon lists. Transcripts were abstracted by two reviewers to assess reliability; one third were discussed further to ensure quality and consistency, following the suggestion by Feinstein.

Jargon word lists were developed using an automated procedure adapted from corpus linguistics. In brief, the procedure used a carefully structured seven-step protocol. Steps 1 through 4 were done by a spreadsheet without the need for subjective judgment.
A corpus document was constructed by merging all transcript files. Word frequency software (we use Textanz, Cro-Code, St. Petersburg, Russia) was used to extract a complete list of words, abbreviations, and two- or three-word combinations. Each of the words on the frequency list was converted to its lemma (root), and all words from the same lemma were grouped together. Words and lemmas were cross-indexed to remove “common words,” i.e., those words that were also listed on the composite list of familiar words listed earlier. The remaining words were cross-indexed against an electronic version of Stedman’s Medical Dictionary to identify words for the “highly specialized” list. All abbreviations were also listed as highly specialized with the exceptions of “U.S.” and “U.S.A.” The remaining words were collaboratively examined by the authors for assignment to either the “uncommon” list of jargon words, containing words some people may not recognize, or to the “common-but-confusing” list containing words that are common in English, but in the transcripts were used for an uncommon concept. Words that had been excluded in the third step were re-examined by the four authors to identify additional words for the common-but-confusing list.

For verification purposes, an amendment process allowed abstractors to propose corrections to the highly specialized, uncommon, and common-but-confusing word lists, provided that the word could be ratified by one other abstractor. When such a word was identified, an electronic search of previously abstracted transcripts was used to verify that the newly designated jargon word was not missed in previous abstractions.

Jargon Explanations
It was expected that the effectiveness of residents’ explanations would vary, so a trichotomous variable (definite/partial/absent) was developed. To be assigned a definite rating, the explanation statement had to refer to the jargon word directly and not use jargon itself. Explanations that used jargon were assigned a partial rating. For example in Figure 1 the explanation of “prostate-specific antigen” is rated as partial because it uses the word “marker” without an accompanying explanation. A partial rating could also be assigned for explanations that were not obviously linked to a specific word or for explanations that were questionable to the abstractors. Discrepancies between abstractors were automatically resolved by assignment to the partial explanations category.

Calculations and Analyses
The “explanation ratio” is a measure of the burden of unexplained jargon words over the entire transcript. This quantifying approach is grounded in cognitive psychology research suggesting that the mental effort necessary for puzzling over unexplained concepts inhibits comprehension of information that is subsequently encountered. The explanation ratio is calculated using the formula

\[
\text{Explanation ratio} = \frac{\text{Number of explained words}}{\text{Total count of jargon words in transcript}}
\]

where the numerator is equal to the number of jargon words that are never explained or that are more than two statements earlier than the explanation, plus 0.5 for each jargon word that only follows a partial explanation. For example, in Figure 2 the explanation ratio for “mammogram” is 1+4, or 0.25.
The “explanatory time lag” was a measure of informational content (in statements) presented between the first usage of the jargon word and its first full or partial explanation. In Figure 1 the time lag for “prostate-specific antigen” is 1 statement, while in Figure 2 the time lag for the word “mammogram” is 22 statements.

The explanation ratio and explanation time lag calculations were both done by a spreadsheet (Microsoft Excel, Redmond, WA) without subjective interpretation. Statistical analyses were performed using JMP software (SAS Institute, Cary, NC). One-way ANOVA and chi-squared tests were used as appropriate for variable type. Inter-abstractor reliability for jargon words was calculated using Cohen’s method.

RESULTS

Participant characteristics are shown in Table 1. The interviews averaged 10.1 min (range, 2 to 21.9) and 146.4 statements per transcript. Inter-abstractor reliability was $\kappa = 0.92$. To evaluate the feasibility of our methods we tracked time and expenses, and project that quality improvement projects in the future will be achievable with a budget under $50 per clinician evaluated.

<table>
<thead>
<tr>
<th>Table 1. Participant Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. responding</strong></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>25–29 years</td>
</tr>
<tr>
<td>30–37 years</td>
</tr>
<tr>
<td><strong>Year in residency</strong></td>
</tr>
<tr>
<td>1st</td>
</tr>
<tr>
<td>2nd</td>
</tr>
<tr>
<td>3rd or 4th **</td>
</tr>
</tbody>
</table>
| *Forty-seven residents responded to the age question  
**Fourth year residents were from the medicine-pediatrics program|

Jargon Words

Abstraction and the preceding seven step process in the protocol allowed us to populate the jargon word list with 350 unique jargon words across 86 transcripts. The average number of unique jargon words per transcript was 19.6 (SD 6.1). Most words were used more than once so that the total jargon count averaged 53.6 words per transcript (SD 27.2). There were no significant differences in jargon count by resident gender or year in residency. The five most frequent jargon words included for each type of cancer are listed in Table 2.
Jargon Explanations

All but one of the transcripts included at least one jargon explanation. There was an average of 4.5 explanations per transcript (SD 2.3). Prostate screening transcripts averaged more explanations per transcript than mammography transcripts (5.9 versus 3.2 respectively, p < .0001). The five most frequently explained jargon words for each type of cancer are listed in Table 3.

Table 2. Most Commonly Used Jargon Words in Counseling

<table>
<thead>
<tr>
<th>Condition</th>
<th>Jargon word</th>
<th>Instances</th>
<th>Percent of total for condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer screening</td>
<td>Mammogram</td>
<td>558</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>Screen</td>
<td>245</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>Tissue</td>
<td>117</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>Biopsy</td>
<td>101</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>False positive</td>
<td>55</td>
<td>2.7</td>
</tr>
<tr>
<td>Prostate cancer screening</td>
<td>Prostate</td>
<td>837</td>
<td>32.7</td>
</tr>
<tr>
<td></td>
<td>Screen</td>
<td>246</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>Symptom</td>
<td>109</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Rectal</td>
<td>109</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Biopsy</td>
<td>108</td>
<td>4.2</td>
</tr>
</tbody>
</table>

*Jargon words are lemmatized for this table (e.g., “Screen” includes screen, screened and screening)

**Total number of jargon words in all mammography transcripts = 2,050. Total number of jargon words in all prostate transcripts = 2,558

The average explanation ratio was 0.15 (SD 0.11), meaning that the remaining 85% of jargon was neither explained in the transcript nor explained before two more statements were presented. The explanation ratio was slightly better for prostate transcripts (0.18) than for breast cancer transcripts (0.12, p = 0.0125), but there were no significant differences by resident gender or year in residency.

Time Lag Between Jargon Words and Explanations

The average time lag was 8.4 statements (SD 13.4), meaning that a patient would have been exposed to eight concepts while puzzling over an unexplained word. The lag was greater in mammography (11.6) as compared with prostate transcripts (4.6 statements, p = 0.004). There were no significant differences by resident gender or year in residency.

Sensitivity Analysis Without “Prostate” and “Mammogram”

Two issues prompted us to do a post hoc sensitivity analysis after deleting instances of the word “prostate” and words beginning with “mammogra-,” with the exception of compound words such as “prostate-specific antigen.” The first was that these two words together accounted for almost a third of the total number of words (1,395 of the 4,608 or 30.3%) across the 86 transcripts. Second, many of the opening statements from the standardized patients included one of these words. We used a search
routine to delete these words, so that the average total jargon count decreased to 37.3 words per transcript, the expected relative decrease of 30.3%. The number of transcripts without an explanation increased to 5 of 86 total transcripts (5.8% of total). The average number of explanations per transcript decreased 17% to 3.7. There was a 10% increase in the explanation ratio to 0.16. The average time lag from the first usage of the word to its explanation decreased by 0.08% (0.006 statements).

DISCUSSION

Jargon is frequently a barrier to effective communication, especially when discussing a complicated topic like the risks associated with screening. The two purposes of this paper were to describe residents’ use of jargon during counseling about prostate and breast cancer screening and to introduce a new method for quantifying jargon usage that can be used in population-based samples. In the analysis we found that jargon words were common, explanations were rare, and many explanations lagged well behind usage of the words that they were supposed to explain. The analysis was limited by use of a small sample of residents from a single program, which limits generalizability to other residency programs or clinicians in practice. Additional research should be conducted in more broadly representative samples of physicians before wider implementation of the method.

From the patient perspective, our approach may overestimate jargon burden because of the low explanation ratio of the word “prostate” and words beginning with “mammogra-” (Table 3). The high prevalence of these words is understandable since they were central to the topic of conversation, but the words’ low explanation ratio could have been affected by the standardized patient’s necessary use of the words to begin conversation. When these words were removed for the sensitivity analysis described in the results section, however, the jargon count was still high, and the explanation ratio was still quite small. In addition, the finding that 60% and 17% of residents choose to explain the word mammogram and prostate, respectively, (Table 3) suggests that some residents still thought it was necessary to explain these words even though they had already been used by the patient.

**Table 3. Most Commonly Explained Jargon Words in Counseling**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Jargon word*</th>
<th>n</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer screening</td>
<td>Mammogram</td>
<td>27</td>
<td>(60.0)</td>
</tr>
<tr>
<td></td>
<td>False Positive</td>
<td>16</td>
<td>(35.6)</td>
</tr>
<tr>
<td></td>
<td>Biopsy</td>
<td>11</td>
<td>(24.4)</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>8</td>
<td>(17.8)</td>
</tr>
<tr>
<td></td>
<td>Screen</td>
<td>5</td>
<td>(11.1)</td>
</tr>
<tr>
<td>Prostate cancer screening</td>
<td>Prostate-specific antigen</td>
<td>31</td>
<td>(75.6)</td>
</tr>
<tr>
<td></td>
<td>PSA</td>
<td>28</td>
<td>(68.3)</td>
</tr>
<tr>
<td></td>
<td>Benign prostatic hypertrophy</td>
<td>20</td>
<td>(48.8)</td>
</tr>
</tbody>
</table>
Regardless of the size of the problem, we are concerned about the challenge that jargon presents for patients with limited health literacy or English skills, and for patients who are ethnically or culturally different from their clinicians. Patients with high health literacy may have sufficient understanding of words like “prostate” and therefore might not need an explanation, but clinicians should be cognizant that patients may use words that they do not entirely understand and consider taking a conservative approach of providing an explanation when there is doubt about likely understanding.

Several research questions remain. The finding that most residents explained at least one jargon word suggests that they may already be aware that misunderstanding of certain words is possible; thus, studies should investigate whether clinicians tend to overestimate their patients’ vocabularies or if there is some other reason why jargon is often unexplained. A statewide study of counseling after newborn screening currently in progress will allow us to quantify the association between jargon usage and comprehension in that population.

In the time since this manuscript was first submitted, two other research groups have developed methods to quantify jargon usage. In one study of primary care visits, Castro et al. operationalized jargon as either “clinical or technical terms with only one meaning listed in a medical dictionary (e.g., hemoglobin A1c)” or “clinical terms used in health care contexts with distinct meanings in lay contexts (e.g., your weight is stable)” (p. s86). In another study, Keselman et al. developed a “predictive familiarity model” based on input to a multiple regression formula from text frequencies found in Reuter’s news reports, queries to an Internet health search engine, queries to a general Internet search engine, and further input from 41 laypersons. The model was used to categorize 45 words as either “likely,” “somewhat likely,” or “unlikely” to be familiar to laypersons. This line of research is a positive development. Future research comparing the models, their generalizability, costs, and their relationship to care outcomes will advance the field further.

In our view, the most important area for future jargon research is the development of metrics that can be applied on a population-wide scale with the ultimate goal of reducing jargon usage and improving clinicians’ explanations. To be successful on these large scales, methods will need to be quantitatively reliable, concrete enough for use by quality improvement professionals with typical training, and able to function on a lean budget. We have adapted a methodological approach from Quality Improvement because of that field’s track record for improving other complex clinician behaviors over large populations. The main goal of this project was to extend our previous work into a method that will reliably and affordably quantify jargon usage and explanations for use in population-based quality improvement projects.

In summary, the high prevalence of jargon words and low prevalence of jargon explanations suggests that there may be problems with communication and patient comprehension during discussions about cancer screening. Our explicit-criteria method is concrete enough for use by quality improvement professionals and customizable for different clinical topics. Since it is not possible to draw conclusions about individuals based on population-based data, we advise clinicians to approach each patient as an
individual, regardless of the patients’ apparent health literacy, and to be conservative about word choice, explanation, and assessment of understanding. Reduced use of jargon and improved explanations by individual clinicians or over entire populations are likely to result in increased communication effectiveness and patient participation in care.

Notes

Acknowledgments

Lindsay Deuster was supported by a training grant from the National Heart, Lung and Blood Institute (T35-HL72483-24). Dr. Farrell was supported in part by grant K01HL072530 from the National Heart, Lung, and Blood Institute. An earlier version of this paper was presented as a poster at the 2007 annual meeting of the Society of General Internal Medicine. The authors are grateful to Dr. Jeffrey Stein for assistance with educational workshops and transcript proofreading, and to Dr. Richard Frankel for very helpful critique and suggestions about the manuscript.

Conflict of Interest

None disclosed.

References


17. Donovan J, Deuster L, Christopher SA, Farrell, MH. Residents’ precautionary discussion of emotions during communication about cancer screening, poster at 2007 annual meeting of the Society for General Internal Medicine, Toronto, Ontario, Canada.


27. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1994;101(2):343–52.


32. Stedman TL. Stedman’s medical dictionary. 28Philadelphia: Lippincott Williams & Wilkins; 20061 v. (various pagings).

