The Effect of User Factors on Consumer Familiarity with Health Terms: Using Gender as a Proxy for Background Knowledge about Gender-Specific Illnesses

Alla Keselman [1,2], Lisa Massengale[1], Long Ngo[3], Allen Browne[1], and Qing Zeng[4]

[1] LHNCBC, National Library of Medicine, NIH, DHHS, Bethesda, MD
[2] Aquilent, Inc., Laurel, MD
[3] DSG, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[4] DSG, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

keselmana@mail.nih.gov

Abstract. An algorithm estimating vocabulary complexity of a consumer health text can help improve readability of consumer health materials. We had previously developed and validated an algorithm predicting lay familiarity with health terms on the basis of the terms' frequency in consumer health texts and experimental data. Present study is part of the program studying the influence of reader factors on familiarity with health terms and concepts. Using gender as a proxy for background knowledge, the study evaluates male and female participants' familiarity with terms and concepts pertaining to three types of health topics: male-specific, female-specific and gender-neutral. Of the terms / concepts of equal predicted difficulty, males were more familiar with those pertaining to neutral and male-specific topics (the effect was especially pronounced for "difficult" terms); no topic effect was observed for females. The implications for tailoring health readability formulas to various target populations are discussed.

Keywords: consumer health informatics; readability formulas; consumer health vocabularies.

1 Introduction

Studies suggest that individuals frequently have difficulties reading health texts, and that the readability of most consumer health websites are beyond the reading level of the average consumer [1]. Vocabulary complexity is one of the text factors that contribute to this difficulty [2]. An informatics tool that could evaluate vocabulary complexity of a health text and suggest consumer-friendly synonyms for "difficult" medical terms could help address this problem. The development of such a tool, however, is a challenging task. First of all, a definition of a "difficult" health term is required. Many general-purpose readability formulas estimate word "difficulty" in terms of their length [3]. This approach may not be appropriate for consumer health

domain, ridden with many short technical terms that are likely to be unfamiliar to lay health consumers (e.g., “myelin”, “apnea”). Moreover, reader’s ability to comprehend a text is affected by many factors that are located with the reader (e.g., prior knowledge, motivation), rather than with the text [4]. The effect of various reader factors on comprehension in consumer health texts domain may be somewhat different from other domains. For example, in many “general” domains individuals with high levels of educational attainment are more likely to comprehend texts with complex vocabulary. In the consumer health domain, however, experience with a particular disease may override “insufficient” education level.

We have previously developed a regression model for predicting consumers’ “familiarity likelihood scores” with health terms. The model relies on two sources of information: 1) empirical data from user studies evaluating “Consumer-Friendly Display” names for medical concepts [5] and (2) term frequency counts from consumer health corpora [6]. The algorithm assigns each consumer health term a predicted familiarity likelihood score from 0-1 range. Terms with scores in the 0.8-1 sub-range are categorized as “likely” to be familiar to health consumers, scores in the 0.5-0.8 sub-range are categorized as “somewhat likely” to be familiar, and scores in the 0-0.5 sub-range are categorized as “not likely” to be familiar. A validation study with 52 participants showed that model-based scores were indeed predictive of consumer recognition and understanding of health terms [7]. The validation study also pointed to two reader factors that could mediate familiarity with health terms: health literacy and English proficiency.

Present study continues to explore reader factors influencing familiarity with consumer health terms and concepts. General comprehension literature contains many testimonies of the effect of background knowledge on text comprehension [8]. Part of the positive effect of background knowledge on comprehension has to do with the fact that background knowledge broadens vocabulary knowledge [9]. Matching vocabulary complexity of consumer health materials to the level of background knowledge of potential readers may therefore improve readability. Web designers and writers rarely conceptualize consumer health audiences in terms of their background knowledge. Instead, audiences are usually defined in terms of some demographic and/or experiential factors (e.g., patients with a specific disease, women, seniors). A match between topic and reader characteristics, however, is likely to influence background knowledge, as individuals are more likely to have knowledge of issues that they have personally experienced and that are specific to their group. We may expect women to be more familiar than men with terminology pertaining to female-specific diseases, and diabetes patients to be more familiar than the general public with diabetes terminology. Identifying demographic factors that are likely to affect term familiarity may allow us to make the predictive model more sensitive by adjusting it to various target population groups.

The general hypothesis underlying this study is that readers’ background knowledge influences their familiarity with health terms and concepts. The specific hypotheses concern the effect of gender on consumer familiarity with terms related to gender-specific health issues. They are the following:

1. Participants’ gender will affect their familiarity with terms related to different health topics. Given comparable familiarity likelihood scores of the terms,

men will be more familiar with terms pertaining to male-specific and neutral health topics than with terms pertaining to female-specific topics. Similarly, women will be more familiar with terms pertaining to female-specific and neutral topics than with those pertaining to male-specific topics.

2. The relationship between gender and topic may differ for terms with different predicted familiarity likelihood scores. Relatively common terms that are predicted as highly likely to be familiar may be equally familiar to both genders regardless of the topic. However, familiarity with terms predicted as unlikely to be familiar may be more affected by the gender-topic match.

Gender in this study was chosen as a proxy for better knowledge of gender specific health issues. The study was not concerned with the general effect of gender on the knowledge of health terminology and concepts.

Based on our previous findings, we also expected that regression model-based familiarity likelihood level would be predictive of consumers' actual term familiarity.

.

2 Methods

2.1 Participants

Convenience sample of 50 employees of the US National Library of Medicine was recruited for the study. Twenty five of the participants were males, and twenty five were females. All had adequate health literacy skills (scores in the 23-36 range out of 36, comparable average scores for both groups), according to Short Test of Functional Health Literacy in Adults (S-TOFHLA) [10]. Male and female groups had comparable educational levels. For each gender group, seven participants had high school level of education (possibly with some college work, but without college diploma), nine were college graduates, and nine had graduate degrees. Female participants were slightly younger than male participants (3 males and 8 females in the 18-25 year old category, 12 males and 9 females in the 26-39 year old category, 10 males and 7 females in the 40-59 year old category, and 1 female in the over 60 years old category).

2.2 Instrument

The survey instrument used in this study tested consumer familiarity with 27 health-related terms. Nine of these terms pertained to conditions that were prevalent among or specific to males (baldness and prostate cancer); nine terms pertained to conditions specific to females (menopause and pregnancy). The terms were extracted from consumer health websites on these four topics, linked to MedlinePlus consumer health portal of the US National Library of Medicine. The remaining nine terms were gender neutral terms extracted from MedlinePlus-linked consumer health website on the topic of gastroesophageal reflux disorder (GERD). From now on, for simplicity, we

will refer to the terms as “male”, “female” and “neutral.” These labels, however, refer to the topics of the texts from which the terms were extracted, rather than to the terms themselves.

The terms were selected to be comparable in “familiarity likelihood scores”, as computed by our regression model algorithm [5, 6]. In each group of nine terms (“male”, “female” and “neutral”), three terms were categorized as “likely” to be familiar to health consumers, three were predicted as “somewhat likely” to be familiar, and three were predicted as “not likely” to be familiar.

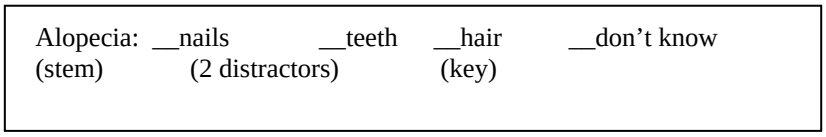

Fig. 1. Sample item from the familiarity test

The layout of the survey was modeled on the Short Assessment of Health Literacy for Spanish-speaking Adults (SAHLSA) [11], which in turn is based on the Rapid Estimate of Adult Literacy in Medicine (REALM) health literacy test for English speakers [12]. SAHLSA consists of 50 items, each with a “stem” or target term, “key” or semantically-related term, “distractor,” and a “don’t know” option to discourage guessing. The goal of SALHSA is to measure both reading ability and comprehension, and the task is to both correctly select and correctly pronounced the key answer option. Since we were interested in evaluating participants’ familiarity with health terms in written consumer health materials, the SALHSA requirement to pronounce the key answer was replaced with a second distractor (Figure 1).

Table 1. Familiarity instrument terms.

Familiarity Likelihood	“Male”	“Female”	Neutral
Likely to be familiar	baldness prostate prostatitis	folic acid osteoporosis menopause	asthma acid reflux biopsy
Somewhat likely to be familiar	scalp testosterone urethra	ovaries uterus prenatal	pulmonary fibrosis esophagus antacids
Not likely to be familiar	rogaine hematuria alopecia	perimenopause phytoestrogens blastocyst	heartburn sphincter internist

Two types of questions were developed for each term:

- Surface-level familiarity questions assessed the ability to match written health terms with basic relevant associated terms at the super-category, location or function level (eg, alopecia → hair) (Figure 1).
- Concept familiarity questions assessed the ability to associate written terms with brief phrases describing the meaning or “gist” (e.g., alopecia → hair loss).

The final instrument consisted of 54 questions (27 surface level familiarity questions and 27 concept familiarity questions). Table 1 presents distribution of items among topics and predicted difficulty scores.

2.3 Administration and Scoring

Participants first completed the demographic survey, followed by S-TOFHLA and familiarity survey, with surface-level items followed by concept familiarity items. Surface-level familiarity and concept familiarity scores were calculated separately, in the following way. First, for each type of familiarity, correct answers were assigned the score of 1, while incorrect answers were assigned the score of 0. Next, for each of the three categories of familiarity likelihood (likely, somewhat likely and not likely) within each of the three categories of topic ("male", "female" and "neutral"), the sum of the three answers was computed. Thus, for each type of familiarity score, there were 9 measurements for each subject. Each measurement represented a score for a difficulty level within a topic, and ranged from 0 to 3.

2.4 Statistical Analysis

We used linear mixed-effects models [13] to estimate the quantities of interest. We first checked the distribution of the Surface-Level Term Familiarity Score, and the Concept Familiarity Score. The distributions of these two outcome variables appeared to be reasonably normal. Since there were multiple measurements for each subject, the models took into account the within-subject correlation by treating the within-subject measurements using compound symmetry variance-covariance matrix structure. Linear contrasts were then used to obtain the linear combination of parameters of interest (e.g. the estimated mean difference between male and female score for male participants, or for female participants).

To model Surface-Level Term Familiarity, the independent variables Predicted Familiarity Likelihood Score (raw scores from the 0-1 range), Gender, Highest Education Level, Age, and Topic ("male", "female" or "neutral") were used as dependent variables in the linear mixed-effects model. Similarly, the same independent variables were used to model Concept Familiarity Score. As health literacy scores of all participants were in the "adequate" range as measures by S-TOFHLA, health literacy was not used as a variable in the analysis due to lack of variation. Education and age were included in the models as potential confounders, as well as to detect potentially meaningful trends for future studies.

3 Results

3.1 Overall Patterns

Both regressions found statistically significant effects (P <. 001) of predicted familiarity likelihood score on surface-level term familiarity and concept familiarity. Surface-level familiarity model also found statistically significant effect of topic, with participants appearing most familiar with neutral terms, followed by "male" and then "female" terms. The effect of female specific vs. neutral terms was -0.58 (P<.001), and the effect of male specific vs. neutral terms was -0.34 (P=.007). This may be due to the fact that while familiarity likelihood scores for the three topics were evenly distributed among the three familiarity likelihood categories, the raw scores were somewhat higher for "female" terms. No significant topic effects were found for the concept familiarity model.

3.2 Hypothesis 1: Gender Differences in Mean Familiarity Scores for Different Topics

Means and standard deviations of participants' surface level term and concept familiarity scores by gender and topic are presented in Table 2.

Table 2. Mean surface-level and concept familiarity scores

	Surface-Level Familiarity mean (SD)			**Concept Familiarity mean (SD)**		
Gender	**"Male" terms**	**"Neut" terms**	**"Female" terms**	**"Male" terms**	**"Neut" terms**	**"Female" terms**
Male (n=75)	2.17 (0.79)	2.45 (0.66)	1.85 (1.23)	2.18 (0.82)	2.26 (0.78)	1.68 (1.08)
Female (n=75)	2.57 (0.64)	2.64 (0.69)	2.36 (0.95)	2.56 (0.66)	2.41 (0.79)	2.33 (0.95)

n=75 refers to the number of observations used in the analysis (3 data points per participant) rather than participants

Male Participants' Performance. Male participants showed greater surface-level familiarity with "male" and "neutral" terms than with "female" terms (Table 2). The estimated mean difference between familiarity with "male" vs. "female" terms (corrected for differences in predicted familiarity likelihood scores for "male" and "female" terms, Age, and Education) via linear contrast from the linear mixed-effects model was 0.24 (SE=0.12, P=0.059). The estimated mean corrected difference between familiarity with neutral vs. "female" terms was 0.58 (SE=0.12, P<.001). Male participants also showed greater surface-level familiarity with neutral terms than "male" terms, mean corrected difference 0.34 (SE=0.12, P=0.007).

Concept familiarity was similarly greater for "male" and "neutral" terms than for "female" terms among male participants. The corrected difference between familiarity

with “male” vs. “female” concepts was 0.45 (SE=0.13, P<.001). The mean corrected difference between familiarity with neutral vs. “female” terms was 0.57 (SE=0.13, P<.001).

Female Participants’ Performance. No statistically significant effect of topic on surface-level and concept familiarity was found for female participants.

3.3 Hypothesis 2: The Effect of Predicted Familiarity Likelihood Scores on the Relationship between Gender and Topic (for Male Participants)

We hypothesized that the relationship between gender and topic may differ for terms with different predicted familiarity likelihood scores. As the overall effect of topic on familiarity was not significant for female participants, the analysis for Hypothesis 2 was conducted for male participants only. Table 3 presents the relationship between predicted familiarity likelihood (based on the ranges used in our previous work, see Introduction) and actual familiarity scores for the three topics for male participants. Examination of the data suggests that the greatest difference in mean familiarity scores between “male” vs. “female” terms and “male” vs. “neutral” terms lies at the level of “difficult” terms predicted not likely to be familiar. A linear mixed-effects model was used to estimate the mean corrected difference between “male” and “female” terms (1.32, SE=0.19, P<0.0001), and “male” and “neutral” terms (-0.44, SE=0.19, P=0.02).

Table 3. Male participants’ performance by term difficulty level

	Surface-level term familiarity Mean (SD)			**Concept familiarity Mean (SD)**		
Familiarity likelihood	“Male” terms	“Neut” terms	“Female” terms	“Male” terms	“Neut” terms	“Female” terms
Likely (n=25)	2.28 (0.54)	2.88 (0.33)	2.44 (0.58)	2.20 (0.70)	2.68 (0.56)	1.84 (0.80)
Somewhat (n=25)	2.60 (0.58)	2.40 (0.65)	2.80 (0.50)	2.64 (0.64)	2.16 (0.75)	2.64 (0.49)
Not likely (n=25)	**1.64 (0.91)**	**2.08 (0.70)**	**0.32 (0.56)**	**1.72 (0.84)**	**1.96 (0.84)**	**0.56 (0.65)**

The pattern was similar for concept familiarity. The estimated corrected mean difference between “male” and “female” terms was 1.16 (SE=0.21, P<0.0001), and between “male” and “neutral” terms was –0.14 (SE=0.21, P<0.0001).

4 Discussion

The study supported our notion that while a primarily frequency-based algorithm for estimating consumer familiarity with health terms has significant predictive power, some reader factors may also carry predictive weight. Including these reader factors into the regression model algorithm can potentially make the model more powerful. In particular, this study pointed to the effect of gender on familiarity with health terms that pertain to gender-specific topics.

The findings of the study also suggest that the relationship between gender and knowledge of terminology related to gender-specific health topics may be less straightforward than we had expected. As expected, men were more likely to be familiar with "male" and "neutral" terms thank with "female" terms of comparable predicted difficulty. Also as expected, the relationship mostly existed at the level of low frequency terms, predicted to be largely unfamiliar. The findings for the female participants, however, were unexpected, as no difference was found in women's familiarity with terms pertaining to different topics. One possible explanation for this is that women are more likely than men to play the role of family caregivers and therefore be familiar with health issues that are not directly relevant to them [14]. We should also keep in mind that this study only tested a small set of terms related to four gender specific health issues, presented to the participants out-of-context, rather than within a passage. Finally, the relationship may be somewhat obscured by the ambiguity of the concept of gender-specific terminology. While all the terms used in the study pertained to gender-specific health issues, some of them denoted anatomical structures and attributes that were common to both males and females (e.g., "scalp", "urethra"). It is conceivable that the effects of the study would be stronger and would generalize to female participants, had we chosen a different definition of "gender-specificity."

In the present study, gender was used as a proxy for background knowledge. The findings support the idea that background knowledge and experience are likely to affect individuals' familiarity with health-related terminology. A follow-up study could validate these findings by including a direct measure of background knowledge and correlating it with gender. Other (perhaps more clinically promising) proxies of background knowledge that deserve research attention are health status, diagnosis and time since diagnosis. For example, patients with chronic illnesses and experience with managing their health status are likely to be more knowledgeable about terms and concepts related to their condition than the newly diagnosed. This, in turn, will have implications for setting the optimal terminological and conceptual complexity of e-health websites targeting various specific populations.

The ultimate goal of our research agenda is to develop an algorithm that could predict readability and comprehensibility of consumer health materials for individuals. Part of the challenge lies in identifying the factors that are likely to affect readability and term familiarity. While the specific hypotheses of the present study addressed the effect of gender, level of education was also included in the regression analysis. The lack of education effect is counter-intuitive, may be due to the limited value range of the variable (high school to graduate school), and perhaps warrants a more thorough investigation. Another variable that is likely to affect term familiarity is health literacy. Studying the effect of health literacy is methodologically difficult,

because most existing tests (e.g., S-TOFHLA) have low ceiling and are not sensitive enough to detect health literacy variations in a typical convenience sample. Yet another part of the challenge lies in accurately estimating the size of the effect of various factors, and then incorporating these effect sizes into the formula. This task requires collecting the data on large samples of participants, using a wide range of health terms.

When thinking about tailoring health materials to individuals' characteristics, it is important to distinguish between stable and transient features of readers and users. Transient features are those that exhibit significant fluctuations as a function of time and context. These may include the users' mood, level of fatigue and current blood glucose level. Stable features are those that can only be changed with a significant investment of time and effort (if at all), including for example presence of a chronic disease that requires continuous management, caregiver status, and age. While both factors may affect reading comprehension, it is presently unrealistic for us to talk about tailoring messages to accommodate transient states. Instead, we can focus on stable characteristics that constitute membership in the targeted audience group for the health materials in question. For example, if our hypothesis of the relationship between background knowledge and terminological knowledge is true, we can further hypothesize that patients' knowledge about their disease increases with time since diagnosis. We can then envision two versions of a website dedicated to providing information about a specific disease. One version would be for the newly diagnosed, the other for individuals who have lived with this diagnosis for a period of time. Vocabulary complexity could then constitute one of the differences between the two versions. Similarly, a website providing support to caregivers can have some information specifically tailored for male and female caregivers. Findings of our study (if confirmed by subsequent research) would suggest that the information for female caregivers could accommodate more complex terminology with less detriment to comprehension.

This study looked at the effect of gender on two types of consumer familiarity with health terms: surface-level familiarity and concept familiarity. While the findings for both types of familiarity were similar, we should not assume that these results would generalize to all contexts. Consumer health term and concept familiarity has more inherent complexity than the present survey captures. Historically, health literacy studies do not distinguish among different levels of familiarity, from associating the term with a broad health area it belongs to, to deep understanding of the underlying concept. The ability to associate the term with a related term or use it in a sentence correctly is often viewed as an indicator of understanding the underlying concept. However, the relationship between surface level familiarity and conceptual knowledge may be non-linear. In a previous study we have shown that conceptual knowledge may lag behind terminological familiarity, and that the gap may be greater for frequent terms that are more likely to be familiar [7]. Our current algorithm was not specifically designed to predict conceptual knowledge. Further work is knowledge assessment is necessary for optimize the algorithm for predicting understanding.

In summary, this paper presents a step in a research program, intended to accumulate knowledge for developing a formula for predicting readability of health materials for various consumer groups. Follow-up work should address the limitations of this study by increasing the scope of terms in the study, including additional

individual factors, defining the continuum of term/concept familiarity and developing methodology for assessing various stages of familiarity. Findings of such research program can be used in the design of tools for assisting consumers with information seeking and comprehension of health materials.

Acknowledgments. This research was supported by the Intramural Research Program and the Association Fellowship of the US National Library of Medicine, US National Institutes of Health (AK, LM, AB) and NIH grant R01 LM007222-05 (LN, QZ). The authors thank Tony Tse and Guy Divita for their comments on the earlier versions of this manuscript.

References

1. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22-29;287(20): 2691-2700.
2. Berland GK, Elliott MN, Morales LS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001 May 23-30;285(20): 2612-2621.
3. Flesch JR, Kincaid C. Flesch-Kincaid Readability Formula. Boston: Houghton Mifflin. 1965.
4. Guthrie TJ, Wigfield A, Metsala JL, Cox KE. Motivational and Cognitive Predictors of Text Comprehension and Reading Amount. Scientific Studies of Reading 1999; 3 (3): 231-56
5. Zeng Q, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. Proc AMIA Symp 2005: 859-63.
6. Zeng Q, Kim E, Crowell J, Tse T. A text corpora-based estimation of the familiarity of health terminology. Proc ISBMDA 2005: 184-92.
7. Keselman A, Tse T, Crowell J, Browne A, Ngo L, Zeng Q. Assessing Consumer Health Vocabulary Familiarity: An Exploratory Study. Proc MEDNET 2006.
8. Langer JA. Examining Background Knowledge and Text Comprehension Reading Research Quarterly 1984; 19 (4) : 468-81
9. Anderson RC, Freebody P. Vocabulary knowledge. In J.T. Guthrie (Ed.), Comprehension and teaching: Research reviews. Newark: International Reading Association. 1981.
10. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1): 33-42.
11. Lee S-YD, Bender DE, Ruiz RE, Cho YI. Development of an easy-to-use Spanish health literacy test. Health Serv Res. 2006;41(4):1392-1412.
12. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med. 1993;25(6): 391-5.
13. Laird MN, Ware HJ. (1982). Random-Effects Models for Longitudinal Data. Biometrics 38: 963-974.
14. Bull MJ. Interventions for women as family caregivers. Annu Rev Nurs Res. 2001;19:125-42