Applying Multiple Methods to Access the Readability of a Large Corpus of Medical Documents

Danny Wu a, David Hanauer b, Qiaozhu Mei a
Patricia Clark c, Lawrence An c, Jainbo Lei d, Joshua Proulx e
Qing Zeng-Tritler e, Kai Zheng f

a. School of Information, University of Michigan, Ann Arbor, MI, USA
b. Department of Pediatrics, University of Michigan, Ann Arbor, MI, USA
c. Center for Health Communications Research, University of Michigan, Ann Arbor, MI, USA
d. Center for Medical Informatics, Peking University, Beijing, China
e. Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA
f. Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, USA
“Inviting patients to review these notes could improve understanding of their health, foster productive communication, stimulate shared decision making, and ultimately lead to better outcomes”

– Delbanco, et. al. *Open Notes*. 2010

**Potential disadvantages**

**Providers**
- More time on address concerns
- Inappropriate reactions

**Patients**
- See something shacks trust
- Time for examination diminished
What is the problem?

• Simply granting access is not enough!
• Patients are unable to read and understand the documents
  ➢ misinterpretation
  ➢ Confusion & anxiety
  ➢ Unintended consequences
How to approach the problem?

Garner M., Ning Z., & Francis J. 2010
Definition of Readability

“The quality of written language that makes it easy to read and understand”

– Merriam-Webster Dictionary

Quantify the readability by
• Number of difficult words
• Sentences length
• Dictionary coverage
“I had the pleasure of seeing Jennifer Smith in my clinic. Jennifer’s wound is continuing to make fantastic progress. It has closed down remarkably well. I understand that she is going to be leaving Arbor Hills this weekend. I am asked if she can go to twice a day wound changes. I feel this would be fine. She will continue with these.”

“Right hydroureter confirmed by retrograde pyelogram prior to stent placement
Right Mid-Ureter Ureteral Obstruction”
Research Questions

1. How different the readability of medical documents would be?
2. How the readability measures may perform?

<table>
<thead>
<tr>
<th>Measure</th>
<th>Doc Group 1</th>
<th>Doc Group 2</th>
<th>Doc Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Empirical Dataset

- EHR at U of M Hospital System
- 2.5 m clinical documents
- Decedent hematology/oncology patients in past 3 years
- Various types

Referral letters are special!
### Empirical Dataset

| Referral Letters | RL  | 76,012 | 50,000 |
| Non-referral documents | N-RL | 2,118,463 | 50,000 |

### Baseline

**Articles on Medline Plus**

926 (~ June 30, 2012)
Examples for 3 groups

Referral Letters

“As you recall, he is a eighty-six-year-old gentleman with a history of a significant cataract in his right eye who presented for re-evaluation of his cataract.”

Non-referral Letters

“CHF with ischemic cardiac myopathy and ejection and an ejection fraction of 35%. PVOD with bilateral carotid stenosis.”

Articles on Medline Plus

“The pattern of how you walk is called your gait. A variety of problems can cause an abnormal gait and lead to problems with walking.”
Text Features of the Groups

• Average document length
• Average sentence length
• Vocabulary Size
• Vocabulary Coverage
  – UMLS
  – Basic Medical English Dictionary (BMED)
<table>
<thead>
<tr>
<th>Surface Metric</th>
<th>RL</th>
<th>N-RL</th>
<th>MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average document length</td>
<td>623.6</td>
<td>495.5</td>
<td>124.6</td>
</tr>
<tr>
<td>Average sentence length</td>
<td>10.9</td>
<td>13.7</td>
<td>12.2</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>184,448</td>
<td>205,283</td>
<td>6,772</td>
</tr>
<tr>
<td>Vocabulary covered by UMLS</td>
<td>24.2%</td>
<td>22.7%</td>
<td>67.0%</td>
</tr>
<tr>
<td>Vocabulary covered by BMED</td>
<td>21.8%</td>
<td>19.1%</td>
<td>97.3%</td>
</tr>
<tr>
<td>Vocabulary covered by UMLS and BMED combined</td>
<td>33.7%</td>
<td>30.7%</td>
<td>99.5%</td>
</tr>
</tbody>
</table>
4 Readability formulas

Kim et al (2007)

\[
D_i = \sum \left( \frac{X_{ij}^{\text{test}} - X_{ij}^{\text{easy}}}{\text{STD}_{ij}^{\text{easy}}} \times W_{ij} \right) \times \frac{1}{\sum W_{ij}}
\]

Include health-specific samples!

\[
W_{ij} = \frac{|\bar{X}_{ij}^{\text{difficult}} - \bar{X}_{ij}^{\text{easy}}|}{\text{STD}_{ij}^{\text{easy}}}
\]

KFGL (1975)

\[
0.39 \times \left( \frac{\text{words}}{\text{sentences}} \right) + 11.8 \times \left( \frac{\text{syllables}}{\text{words}} \right) - 15.59
\]

SMOG (1969)

\[
1.0430 \times \sqrt{\frac{\text{polysyllables}}{\text{sentences}}} \times 30 + 3.1291
\]

GFI (1952)

\[
0.4 \times \left( \frac{\text{words}}{\text{sentences}} + 100 \times \left( \frac{\text{complex words}}{\text{words}} \right) \right)
\]
# Results

*Use ANOVA in STATA, a < 0.05*

<table>
<thead>
<tr>
<th>Readability Measures</th>
<th>RL (50,000)</th>
<th>N-RL (50,000)</th>
<th>MP (936)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al.</td>
<td>-0.3 ± 0.2</td>
<td>-0.4 ± 0.2</td>
<td>0.1 ± 0.2</td>
</tr>
<tr>
<td>FKGL</td>
<td>9.4 ± 1.3</td>
<td>9.1 ± 1.9</td>
<td>8.3 ± 1.8</td>
</tr>
<tr>
<td>SMOG</td>
<td>12.3 ± 1.1</td>
<td>11.9 ± 1.5</td>
<td>10.9 ± 1.5</td>
</tr>
<tr>
<td>GFI</td>
<td>13.2 ± 1.5</td>
<td>12.9 ± 2.2</td>
<td>11.2 ± 2.2</td>
</tr>
</tbody>
</table>

Between -1 and 1

Number of years of education
# Results

<table>
<thead>
<tr>
<th>Readability Measures</th>
<th>RL (50,000)</th>
<th>N-RL (50,000)</th>
<th>MP (936)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al.</td>
<td>-0.3 ± 0.2</td>
<td>-0.4 ± 0.2</td>
<td>0.1 ± 0.2</td>
</tr>
<tr>
<td>FKGL</td>
<td>9.4 ± 1.3</td>
<td>9.1 ± 1.9</td>
<td>8.3 ± 1.8</td>
</tr>
<tr>
<td>SMOG</td>
<td>12.3 ± 1.1</td>
<td>11.9 ± 1.5</td>
<td>10.9 ± 1.5</td>
</tr>
<tr>
<td>GFI</td>
<td>13.2 ± 1.5</td>
<td>12.9 ± 2.2</td>
<td>11.2 ± 2.2</td>
</tr>
</tbody>
</table>

* Use ANOVA in STATA, a < 0.05
# Results

<table>
<thead>
<tr>
<th>Readability Measures</th>
<th>RL (50,000)</th>
<th>N-RL (50,000)</th>
<th>MP (936)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al.</td>
<td>-0.3 ± 0.2</td>
<td>-0.4 ± 0.2</td>
<td>0.1 ± 0.2</td>
</tr>
<tr>
<td>FKGL</td>
<td>9.4 ± 1.3</td>
<td>9.1 ± 1.9</td>
<td>8.3 ± 1.8</td>
</tr>
<tr>
<td>SMOG</td>
<td>12.3 ± 1.1</td>
<td>11.9 ± 1.5</td>
<td>10.9 ± 1.5</td>
</tr>
<tr>
<td>GFI</td>
<td>13.2 ± 1.5</td>
<td>12.9 ± 2.2</td>
<td>11.2 ± 2.2</td>
</tr>
</tbody>
</table>

* Use ANOVA in STATA, $a < 0.05$
Results

- Kim et al.
- Flesch-Kincaid Grade Level (KFGL)
- Simple Measure of Gobbledygook (SMOG)
- Gunning Fog Index (GFI)
Discussion

1. Readability of medical documents needs to be improved

2. Traditional measures were unable to reflect readability by number of years of education (College student? 2 more years?)

3. Traditional measures had counter-intuitive results (RF > N-RF)
Discussion

1. Readability of medical documents needs to be improved

2. Traditional measures were unable to reflect readability by number of years of education (College student? 2 more years?)

3. Traditional measures had counter-intuitive results (RF > N-RF)
Discussion

1. Readability of medical documents needs to be improved

2. Traditional measures were unable to reflect readability by number of years of education (College student? 2 more years?)

3. Traditional measures had counter-intuitive results (RL > N-RL)
Future Work

1. Documents from a multiple sources
2. Engage real patients to validate these computational results
3. Identify nuances across different subtypes
Thank you!