A National Web Conference on Improving Health IT Safety Through the Use of Natural Language Processing to Improve Accuracy of EHR Documentation

Presented by:
Thomas Payne, M.D.
Li Zhou, M.D., Ph.D.

Moderated by:
Chris Dymek, Ed.D.
Agency for Healthcare Research and Quality

February 7, 2017
Agenda

• Welcome and Introductions
• Presentations
• Q&A Session With Presenters
• Instructions for Obtaining CME Credits

Note: After today’s Webinar, a copy of the slides will be emailed to all participants.
AHRQ’s Mission

To produce evidence to make health care safer, higher quality, more accessible, equitable, and affordable, and work within the U.S. Department of Health and Human Services and with other partners to make sure that the evidence is understood and used.
How AHRQ Makes a Difference

• AHRQ invests in research and evidence to understand how to make health care safer and improve quality.

• AHRQ creates materials to teach and train health care systems and professionals to catalyze improvements in care.

• AHRQ generates measures and data used to track and improve performance and evaluate progress of the U.S. health system.
Apply now for **Research Demonstration and Dissemination Projects** in clinical decision support:

- Scale and spread existing clinical decision support for patient-centered outcomes research  

- Develop new clinical decision support for patient-centered outcomes research  

The Division of Health IT is **actively seeking R01, R03, R18, and R21 applications** to study:

- Design, implementation, usability, and safe use of health IT  

- Use of health IT for patient-reported outcomes to improve quality  

- Utilizing Health Information Technology to Scale and Spread Successful Practice Models Using Patient-reported Outcomes  
The following presenters and moderator have no financial interests to disclose:

- Thomas Payne, M.D.
- Li Zhou, M.D., Ph.D.
- Chris Dymek, Ed.D.

This continuing education activity is managed and accredited by the Professional Education Services Group (PESG), in cooperation with AHRQ, AFYA, and RTI.

PESG, AHRQ, AFYA, and RTI staff have no financial interests to disclose.

Commercial support was not received for this activity.
Recent grant support for our presenters includes:

Dr. Thomas Payne:
AHRQ HS023631
AUR A100077

Dr. Li Zhou
AHRQ HS024264
AHRQ HS022728
Controlled Risk Insurance Company (CRICO)
Brigham Care Redesign Incubator and Startup Program
How to Submit a Question

• At any time during the presentation, type your question into the “Q&A” section of your WebEx Q&A panel.
• Please address your questions to “All Panelists” in the drop-down menu.
• Select “Send” to submit your question to the moderator.
• Questions will be read aloud by the moderator.
Learning Objectives

At the conclusion of this activity, the participant will be able to do the following:

1) Discuss the development and evaluation of an enhanced electronic note system that leverages voice recognition and Natural Language Processing (NLP) technologies to create electronic physician notes in the EHR.

2) Discuss the challenges of introducing speech recognition technology into existing medical culture and current clinician workflow, including user preferences and the quality of documents generated by this technology.

3) Explain the need for an automated error detection system using NLP for improving the accuracy and quality of speech recognition generated medical documents, and discuss the development and evaluation of such a system.
Improving Accuracy of Electronic Notes Using a Faster, Simpler Approach

Thomas H. Payne, M.D.
University of Washington
Objectives

- Review problems with current physician note-writing practices.
- See how these problems might be addressed using current technologies and commercial EHRs.
- Understand barriers to changing physician documentation practices and how to address them.
Disclaimer

- Analysis still underway
- Statistical testing not yet performed

*Final results may differ from those presented here.*
Terms Used in This Webinar

**Automatic speech recognition:** Using software to convert spoken speech into text.

Ways to use automatic speech recognition software:

- **Interactive:** The user speaks into a microphone and watches the screen as the voice is converted to text and the user corrects errors interactively.

- **Noninteractive:** The user creates a voice file containing the entire document by speaking into a telephone or voice recorder. Automatic speech recognition software converts the voice file to text in the background while the user is engaged in other activities. A transcriptionist or the user corrects errors.
What Hath We Wrought?
Other Problems With Notes

• Note bloat.

• Copy and paste is common, usually as a result of efforts to save time.

• Progress notes are finished so late that other team members may not see them until the next day.
Impression and Plan

y/o s/p cardiac transplant in 2002, admitted to Neurology for seizures, c/b traumatic intubation. Transferred to MICU service acutely hypotensive and hypoxic status post tracheostomy, now very agitation on the ventilator.

1) Neurology: Neuro following. Bedside 45 minute continuous EEG in attempt to capture possible seizure activity - results described above. Neuro exam largely unchanged. Will get MRI today to further assess extent of cerebral injury.

2) Respiratory: Worsening hypoxia with bilateral infiltrates. BAL reveals mixed flora including neisseria, alaph-hem strep; will continue Vancomycin and Cefepime. Picture consistent with ARDS therefore continue LPS settings for vent. Of concern is his continued agitation in the setting of high cO2 requirements. Given his immobility, hypoxemia, sinus tachycardia, and overall agitation, the concern for pulmonary embolism is high at this time. We will start him on renal prophylaxis with bicarbonate and obtain a CTA to determine whether or not he has a PE. If the does, he should be anticoagulated with heparin, given his hx of HIT. Will have to start bivalirudin for anticoagulation.

3) ID: WBC 11.5 and continues to spike, but blood ox NGTD. At risk for infections considering immunosuppression for heart transplant and prolonged hospital stay. Will continue Vancomycin and Cefepime for now, although it is not clear if there is a PNA or simply ARDS at this time.

4) HTN: Continue Metoprolol to 75 mg po TID and Nipride 10mg PO daily. BP continues to be elevated in the setting of his agitation, as does heart rate. This may be multifactorial in nature, with dehydration, fevers, and pain contributing. It is notable that his agitation did not improve significantly to pm ativan and pain medications.

5) Anemia: Hct stable

6) s/p cardiac transplant: Cards, Dr., following. PT's Tacrolimus levels sporadik. Continually readjusting dosage. Today's level pending. Per pharmacy goal 8-10. Continue to hold TF while administering Tacrolimus.

7) Seizure: Will continue Keppra.

8) DVT prophylaxis: PT with hx of HITs, but not documented as occurring this hospitalization. Remains on SCDS and Teas.

9) Volume Status: Increased fluid intake over past 24 hours. Not +527ml, but BUN slightly worse at 33 from 26, suggesting some degree of dehydration; continue to hydrate liberally.

10) FEN: TF at goal rate. GI willing to place PEG when appropriate for long term care. Awaiting for improvement in WBC and low-grade fever before placing PEG.

11) Dispo: Needs continued ICU care - not ready for transfer to SNF. Need to readress current status with family.

12) CODE: FULL
Specific Aims

Specific Aim 1
To refine and implement a new voice-generated enhanced electronic note system (VGEENS), integrating voice recognition and transcription with natural language processing and links to the electronic medical record (EMR) to improve note creation efficiency and note accuracy.

Specific Aim 2
To evaluate VGEENS using a randomized trial with 30 internal medicine physicians in each arm to assess electronic note creation efficiency, note accuracy, and user satisfaction. Intervention physicians will use VGEENS, while control physicians will continue with note creation as they normally would.
Notes will be ‘dictated’ at the bedside or immediately after leaving it.
# Ways to Write Inpatient Progress Notes

<table>
<thead>
<tr>
<th>Method</th>
<th>Fast</th>
<th>Low cost</th>
<th>Incorporates EMR data</th>
<th>Encoded data</th>
<th>Well suited to rounding workflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGEENS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Keyboard</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Dictation</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Voice recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Specific Aim 2

60 consented inpatient physicians

Random assignment

30 use VGEENS documentation
Create 5-15 notes per day for 28 days

30 use usual documentation
Create 5-15 notes per day for 28 days

1. Minutes between when patient seen on rounds and note signed in EMR
2. Note quality, measured by Physicians Documentation Quality Instrument
3. Satisfaction of physician users
# Physicians Documentation Quality Instrument

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Score</th>
<th>Description of Ideal Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Up-to-date</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note contains the most recent test results and recommendations.</td>
</tr>
<tr>
<td>2. Accurate</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note is true. It is free of incorrect information.</td>
</tr>
<tr>
<td>3. Thorough</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note is complete and documents all of the issues of importance to the patient.</td>
</tr>
<tr>
<td>4. Useful</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note is extremely relevant, providing valuable information and/or analysis.</td>
</tr>
<tr>
<td>5. Organized</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note is well-formed and structured in a way that helps the reader understand the patient’s clinical course.</td>
</tr>
<tr>
<td>6. Comprehensible</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note is clear, without ambiguity or sections that are difficult to understand.</td>
</tr>
<tr>
<td>7. Succinct</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note is brief, to the point, and without redundancy.</td>
</tr>
<tr>
<td>8. Synthesized</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>The note reflects the author’s understanding of the patient’s status and ability to develop a plan of care.</td>
</tr>
<tr>
<td>9. Internally Consistent</td>
<td>Not at all 1 2 3 4 Extremely 5</td>
<td>No part of the note ignores or contradicts any other part.</td>
</tr>
</tbody>
</table>

**Total Score:**
Results
VGEENS System Successfully Developed

Specific Aim 1

- Used with commercial EHR
- Notes available in EHR Inbox within 5 minutes
- Secure
- Enhanced with text processing later in trial
- Downtime uncommon
Voice dictation files will then be securely transmitted to servers…
45 year old female with pulmonary sarcoidosis...

Parent/Child (Relationship Type) Sarcoidosis (disorder) {31541009, SNOMED-CT}

Record note (~5 minutes)

"45 year old female with pulmonary sarcoidosis..."

Edit and sign note (~3 min)
### Specific Aim 2

#### Summary of Subjects, Outcome Data, and Results

<table>
<thead>
<tr>
<th></th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SUBJECTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consented</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Wrote ≥ 1 note</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OUTCOME DATA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded rounding time</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Notes written</td>
<td>709</td>
<td>1143</td>
</tr>
<tr>
<td>Satisfaction survey</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OUTCOME RESULTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing: Note available -</td>
<td>227</td>
<td>190</td>
</tr>
<tr>
<td>Rounding time (minutes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied: Highly or</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>moderately (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissatisfied: Moderately</td>
<td>40</td>
<td>6</td>
</tr>
<tr>
<td>or dissatisfied (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Note quality: Pending</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of Minutes From Midnight When Patients Seen on Rounds

Rounds, minutes after midnight

8 am 10 am 12 pm
INTERVENTION: Number of Minutes From Midnight When VGEENS Used

Use VGEENS soon after seeing patient

Number of subjects

Minutes from MN (min)

420 480 540 600 660 720 780 840 900 960 1020 1080 1140 1200 1260

8 am 10 am 12 pm
Number of Minutes From Midnight When VGEENS Notes Available

- 8 am
- 10 am
- 12 pm

Minutes After Midnight

Count of Trans min from MN

[Bar chart showing distribution of minutes after midnight with peaks at specific times]
Minutes After Midnight Notes
Transcribed, by Author

Colors represent different physicians

Count of Trans/min from MN

8 am 10 am 12 pm Minutes After Midnight

400 500 600 700 800 900 1000 1100 1200 1300

0 5 10 15 20 25 30 35

31 nulls
Time Between Rounds and Signed Note (Control)
Comparing VGEENS and Control Notes With Note Written Previous Day on Same Patient

VGEENS

Identification/chief complaint:
This is a 41-year-old woman with metastatic adenocarcinoma of unknown primary mostly involving the terminal ileum that started and worsening with pain.

Internal history:
General: A 41-year-old woman with a history of poorly controlled pain. She's been on a daily opioid regimen since her diagnosis of metastatic adenocarcinoma. She has been on various pain medications, including fentanyl, for pain management.

Significantly increased pain overnight. Her fentanyl PCA has been reduced to 50 mg to 75 mg/24 hours and is being increased as needed. She is able to walk with the assistance of a walker. The social worker is preparing to discharge the patient to her home later today.

Medications:
- Fentanyl PCA: 50 mg to 75 mg/24 hours
- Other pain medications as needed

Physical exam:
- Vitals: Temperature 36.5°C, heart rate 76 bpm, respiratory rate 20/min, blood pressure 120/80 mmHg
- Oxygen saturation 99% on room air
- Lung exam: Clear breath sounds bilaterally

Assessment and plan:
- Acute management of pain: Fentanyl PCA has been increased to 75 mg/24 hours. The patient will continue to receive Fentanyl PCA as needed.
- Discharge planning: The patient will be discharged to home today. She will continue to receive home care services.

Control

INPATIENT PROGRESS NOTE

HOSPITAL DAY: 07/16/17

IDENTIFICATION/CHIEF CONCERN: With depression noted after suicide attempt by ingestion

INTERNAL HISTORY:

ALLERGIES:
- None

PHYSICAL EXAM:
- Vital signs: Temperature 98.6°F, Heart rate 74 bpm, Respiratory rate 20/min, Blood pressure 120/80 mmHg
- Oxygen saturation 99% on room air

LABS: (Most recent results in bold range)

RESULTS FROM TODAY

RESULTS FROM YESTERDAY

ASSESSMENT & PLAN:

NOTE: Patient was discharged from the hospital earlier today due to improvement in his condition. He is being monitored closely and will be seen in follow-up appointments.

DISPOSITION: Patient is stable and will be discharged today. He will be followed up in the clinic in one week.
What Is Medical Informatics?
(attributed to Homer Warner)

- Technology: 10%
- Medicine: 10%
- Sociology: 80%
Lessons Learned

• Physicians may resist changing established note-writing habits.

• The VGEENS approach can make creating progress notes faster if voice recorded at bedside or soon after.

• Notes created using voice may contain less text carried forward from prior notes and may be more accurate.

• Features popular with physicians: carrying forward plan, ‘checklist’ information, minimizing editing requirements.
Successes

• We developed and deployed a new system, integrated with a commercial EHR, to create inpatient progress notes within 5 minutes.

• If used at the bedside or soon thereafter, notes are available much sooner for others to view.

• Notes created using voice may contain less text carried forward from prior notes.

• We have a method to apply decision support based on progress note content within minutes.
Challenges

- Physicians may resist changing established note-writing habits.
- On average, satisfaction was greater with usual (control) method of note writing, perhaps because popular VGEENS features weren’t available until late in the controlled trial.
Lessons From a First-Time Principal Investigator

• Developing a system to be used in producing and conducting an RCT in 2 years is ambitious, but possible.

• People are enthusiastic about joining an interesting project. Diverse disciplines help!

• Pick a problem you know is important and understand deeply.

• Think about the next project from Day 1 (automated editing, NLP tools, measuring note accuracy, etc.)
Our Team

Co-investigators
Andrew White
Meliha Yetisgen
Tom Gallagher

Computing engineer
Andrew Markiel

Research team
Amelia Chappelle
Jennifer Zech

Collaborators
David Alonso
Xinran (Leo) Liu
Ross Lordon
Kevin J. Lybarger
Mari Ostendorf
Trevor Steinbach
This project was supported by grant number R21HS023631 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the Department of Health and Human Services.
Contact Information

Thomas Payne

tpayne@u.Washington.edu
Improving Health IT Safety Through the Use of Natural Language Processing to Improve Accuracy of EHR Documentation

Li Zhou, M.D., Ph.D.
Harvard Medical School
Brigham and Women’s Hospital
• Accurate medical documents are critical for safe patient care and effective inter-provider communication.

• Errors in medical documents can lead to medical errors in patient care, some of which cause injury or even death.

• ~5 million errors per year are tied to wrong medications; 1 in 4 medication errors involves a pair of drugs whose names look alike or sound alike.
  ► Altenol vs. Atenolol
  ► Lyrica vs. Lamictal

http://www.nbcnews.com/id/37386398/ns/health-health_care/t/look-alike-sound-alike-drugs-trigger-dangers/#.WA0LZOUrLIU
• **Non-word errors** (e.g., *Humulog* for *Humalog*)
  - Free-text entries, typed notes

• **Real-word errors** (i.e., the word is spelled correctly but is contextually wrong, such as *there* for *their*)
  - Speech Recognition (SR) generated text
Background: Non-Word Errors Detection and Correction

- Spelling errors in free-text EHRs

<table>
<thead>
<tr>
<th></th>
<th>Notes (n=315)</th>
<th>Free-text allergy entries (n=2626)</th>
<th>Free-text med orders (n=2743)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error rate</td>
<td>0.5%</td>
<td>4.5%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Clinical terms</td>
<td>28.2%</td>
<td>65.5%</td>
<td>78.0%</td>
</tr>
<tr>
<td>Real-word errors</td>
<td>3.8%</td>
<td>1.8%</td>
<td>0%</td>
</tr>
</tbody>
</table>

- We developed a spell checker in our MTERMS NLP system

<table>
<thead>
<tr>
<th></th>
<th>MTERMS Spell Checker</th>
<th>Aspell Default</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Notes</td>
<td>Allergies</td>
</tr>
<tr>
<td>Precision</td>
<td>71.1</td>
<td>96.2</td>
</tr>
<tr>
<td>Recall</td>
<td>81.0</td>
<td>92.7</td>
</tr>
<tr>
<td>F measure</td>
<td>75.7</td>
<td>94.4</td>
</tr>
<tr>
<td>Accuracy</td>
<td>78.1</td>
<td>88.2</td>
</tr>
</tbody>
</table>

Background: Speech Recognition and Real-Word Errors

- Previous studies were limited in scope and sample size
- Error rates by word range from 1.5%-15%
- High error rate by report (23%) were found in radiology reports
- 76% of radiologists believed error rate by report < 10%

<table>
<thead>
<tr>
<th>Author</th>
<th>Doc Type</th>
<th>Sample Size</th>
<th>Error Rate</th>
</tr>
</thead>
</table>
| Devine* 2000 | A discharge summary and a progress note | 12 physicians | IBM software: 7.0% to 9.1%
| | | | Dragon: 14.1% to 15.2% |
| Kanal* 2001 | 72 radiology reports | 6 participants | IBM MedSpeaker: 10.3%
| | | | Significant errors: 7.8% |
| Zick* 2001 | 47 ED charts | 2 physicians | Dragon NaturallySpeaking: 1.5%
| | | | Errors/chart: 2.5 |
| Quint 2008 | 265 radiology reports | - | 22% of reports contained significant errors |
| Basma 2011 | 615 radiology reports | - | 23% of SR reports contained major errors
| | | | 4% in conventional dictation transcriptions |

* Error Rates: total number of errors divided by the total number of words in the report.
Background: Our SR Error Study in ED

- Retrospectively analyzed 100 emergency physician notes during Dec 2012.
  - Generated via a front-end SR system (Dragon® Medical 10.0)
  - Further edited and signed by the physicians
- 71% of notes contained errors; 1.3 errors per note; 9 errors per 1000 words.
- 15% contained one or more clinically significant errors.
- Physicians signed their notes with known errors, indicating proofreading the entire medical note to search for errors is time consuming.

Surveyed 114 Dragon Users and 50 eScription Users at Brigham and Women’s Hospital, Boston
Aim 1: Conduct error analysis to estimate the prevalence and severity of SR errors.

Aim 2: Develop NLP methods for automated error detection.

This presentation reports our errors analysis in back-end SR generated documents at different processing stages.
Methods

• Stratified random sample of 169 dictated notes using transcription services (back-end SR)
  ▶ 79 from Brigham and Women’s Hospital (24 operative notes and 55 office notes)
  ▶ 40 discharge summaries from North Shore Medical Center, Boston
  ▶ 50 from the University of Colorado Hospital (35 discharge summaries and 15 operative notes)

• Four processing stages
  ▶ Original audio file dictated by the provider (AO note)
  ▶ Note generated by SR engine of the vendor transcription service (SR note)
  ▶ Note edited by a professional medical transcriptionist (MT note)
  ▶ Final note reviewed and signed by a clinician (SN note)

• Three-level annotation schema
  ▶ General error types
  ▶ Semantic error types
  ▶ Clinically significant errors

• Manual review to create gold standard
Measures

• **Length of time** to dictate a note
• **Turnaround time** for each note version
• **Differences** in the SR note, MT note, and SN note from the gold standard
• **Error rate**: number of errors divided by the number of words
• **Percentage of each error type** by overall errors
• Percentage of notes with at least one *clinically significant error*
• **Repeated** these analyses for SR, MT, and SN notes; for each note type; and across all notes
## General Error Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
</table>
| Insertion    | One or more words was added to the transcription                            | AO: There is distal biliary obstruction observed  
SR: There is no distal biliary obstruction observed |
| Deletion     | One or more words was deleted from the transcription                        | AO: CHADS2 VASC score 4  
SR: score |
| Enunciation  | An error due to a mispronunciation or failure to enunciate on the part of the speaker | AO: to find a homeopathic provider  
SR: defined homeopathic provider |
| Suffix error | The root word is correct, but there is an incorrect, added, or omitted suffix | AO: mental status worsened  
SR: mental status worsens |
| Dictionary error | An error due to the target word not being present in the SR system’s dictionary | AO: driving a Camry and hit another car  
SR: driving an Academy and hit another car |
| Spelling error | The transcriptionist made a spelling error when editing the output of the SR system | AO: we counseled him on risk of infection  
MT: we counseled hom on risk of infection |
| Homonym error | One word has been substituted for another identically pronounced word        | AO: serial high resolution anoscopy  
SR: cereal high resolution anoscopy |
| Nonsense error | A substitution that is so far off that it cannot be determined which (if any) category it falls under | AO: follow up in 3 to 5 days  
SR: neck veins are evaluated |
| Prefix Error  | The root word is correct, but there is an incorrect, added, or omitted prefix | AO: Inadequate evaluation to exclude neoplasia  
SR: Adequate evaluation to exclude neoplasia |
| Number error | Any error involving a number, whether it is written as a digit (“3”) or as a word (“three”) | AO: the patient is a 17-year-old female  
SR: the patient is a 70-year-old female |
| Punctuation Error | A period, comma, or other punctuation mark was added where it should not have been | AO: at discharge she had no flank tenderness  
SR: at discharge, She had no flank tenderness |
### Semantic Error Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>General English</td>
<td>Any English words that do not fit into the categories below</td>
<td>AO: which she would otherwise forget SR: which she would otherwise for gas that</td>
</tr>
<tr>
<td>Stop Word</td>
<td>Common English words (we are using the list defined at <a href="http://ranks.nl/stopwords">http://ranks.nl/stopwords</a>)</td>
<td>AO: intermittent pain under the right breast SR: intermittent pain in the right breast</td>
</tr>
<tr>
<td>Medication</td>
<td>Medication names and dose information</td>
<td>AO: initiated on Lamotrigine therapy SR: initiated on layman will try therapy</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Any words that are part of a specific medical diagnosis</td>
<td>AO: Dengue SR: DKA</td>
</tr>
<tr>
<td>Lab</td>
<td>Includes lab test names and lab test results</td>
<td>AO: TSH of 26.7 SR: TSH of 22nd 6.7</td>
</tr>
<tr>
<td>Imaging Test</td>
<td>Imaging exam names/types and exam results</td>
<td>AO: nonobstructive on CT imaging SR: nonobstructive on imaging</td>
</tr>
<tr>
<td>Procedure</td>
<td>Procedure names and descriptions</td>
<td>AO: CA ligament was released on the leading edge SR: CA ligament was released operating edge</td>
</tr>
<tr>
<td>Physical exam</td>
<td>Any information directly related to the physical exam (ht/wt, HR, BP, etc.) and any associated values</td>
<td>AO: T 36.7 degrees SR: T3-T7 disease</td>
</tr>
<tr>
<td>Patient/provider info.</td>
<td>Any words involving patient/provider metadata, such as the patient’s name, doctors’ names, patient MRN, etc.</td>
<td>AO: SURGEON: [surgeon’s actual name] SR: SURGEON: Stathis stairs</td>
</tr>
<tr>
<td>Date</td>
<td>Any dates, including those that are written with words (January 1, 2017) or written with numbers (01/01/2017)</td>
<td>AO: 10/10/2016 SR: 10/10/2000</td>
</tr>
<tr>
<td>Symptom</td>
<td>Any symptom or description of symptoms</td>
<td>AO: very mild althralgias SR: very also arthralgias</td>
</tr>
<tr>
<td>???</td>
<td>When ??__?? (or similar) is left in the note, or when something is completely nonsense</td>
<td>AO: no foreign material was identified MT: no foreign ??__?? was identified</td>
</tr>
</tbody>
</table>
Preliminary Results

Our study is ongoing; final results may differ from those presented here.
## Results

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length of notes (words)</strong></td>
<td>558</td>
<td>524</td>
<td>102</td>
<td>1230</td>
</tr>
<tr>
<td><strong>Dictation time (minutes)</strong></td>
<td>5</td>
<td>4.5</td>
<td>0.4</td>
<td>31.5</td>
</tr>
<tr>
<td><strong>Turnaround time</strong> – Time between completion of dictation and upload to EHR system (hours)**</td>
<td>3.5</td>
<td>1</td>
<td>2 minutes</td>
<td>38.8</td>
</tr>
<tr>
<td><strong>Clinician review time</strong> - Time between upload to EHR system and clinician signing of note (days)**</td>
<td>4.2</td>
<td>1</td>
<td>0</td>
<td>42</td>
</tr>
</tbody>
</table>
## Overall Error Rates and General Types

<table>
<thead>
<tr>
<th></th>
<th>Total Errors n (%)</th>
<th>Errors – General Types n (%)&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Deletion</th>
<th>Insertion</th>
<th>Enunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discharge Summaries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>3892 (9.6)</td>
<td>1395 (35.8)</td>
<td>1031 (26.5)</td>
<td>655 (16.8)</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>195 (0.5)</td>
<td>87 (44.6)</td>
<td>36 (18.5)</td>
<td>35 (18.0)</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>163 (0.4)</td>
<td>74 (45.4)</td>
<td>29 (17.8)</td>
<td>29 (17.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Office Notes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>1588 (6.6)</td>
<td>539 (33.9)</td>
<td>306 (19.3)</td>
<td>431 (27.1)</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>96 (0.4)</td>
<td>29 (30.2)</td>
<td>13 (13.5)</td>
<td>36 (37.5)</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>32 (0.1)</td>
<td>6 (18.8)</td>
<td>7 (21.9)</td>
<td>12 (37.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Operative Notes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>1233 (4.8)</td>
<td>376 (30.7)</td>
<td>401 (32.8)</td>
<td>167 (13.7)</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>120 (0.5)</td>
<td>47 (39.2)</td>
<td>29 (24.2)</td>
<td>18 (15.0)</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>96 (0.4)</td>
<td>42 (43.8)</td>
<td>25 (26.0)</td>
<td>15 (15.6)</td>
<td></td>
</tr>
<tr>
<td><strong>All Notes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>6703 (7.5)</td>
<td>2310 (34.5)</td>
<td>1738 (25.9)</td>
<td>1253 (18.7)</td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>411 (0.6)</td>
<td>163 (39.7)</td>
<td>78 (19.0)</td>
<td>89 (21.7)</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>291 (0.3)</td>
<td>122 (41.9)</td>
<td>61 (21.0)</td>
<td>56 (19.2)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup> n = number of errors; % = total number of errors divided by the total number of words in the notes.

<sup>2</sup> n = number of errors; % = number of errors of a specific type divided by the total number of errors.
Error Rates across All Note Types and Stages

Error rate

- SR note
- MT note
- Signed note

Note processing stage

- Discharge summaries
- Operative notes
- Office notes

Error rate

0% 2% 4% 6% 8% 10% 12%
## Errors by Semantic Type

<table>
<thead>
<tr>
<th></th>
<th>General English</th>
<th>Medication</th>
<th>Diagnosis</th>
<th>Procedure</th>
<th>Symptom</th>
<th>Lab</th>
<th>Physical Exam</th>
<th>Imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discharge Summaries (75)</strong></td>
<td>SR 3255 (83.6)</td>
<td>127 (3.3)</td>
<td>76 (2.0)</td>
<td>13 (0.3)</td>
<td>60 (1.5)</td>
<td>50 (1.3)</td>
<td>30 (0.8)</td>
<td>25 (0.6)</td>
</tr>
<tr>
<td></td>
<td>MT 124 (63.6)</td>
<td>14 (7.2)</td>
<td>9 (4.6)</td>
<td>2 (1.0)</td>
<td>7 (3.6)</td>
<td>2 (1.0)</td>
<td>14 (7.2)</td>
<td>2 (1.0)</td>
</tr>
<tr>
<td></td>
<td>SN 105 (64.4)</td>
<td>6 (3.7)</td>
<td>9 (5.5)</td>
<td>2 (1.2)</td>
<td>7 (4.3)</td>
<td>2 (1.2)</td>
<td>14 (8.6)</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td><strong>Office Notes (55)</strong></td>
<td>SR 1305 (82.2)</td>
<td>41 (2.6)</td>
<td>49 (3.1)</td>
<td>26 (1.6)</td>
<td>38 (2.4)</td>
<td>12 (0.8)</td>
<td>7 (0.4)</td>
<td>7 (0.4)</td>
</tr>
<tr>
<td></td>
<td>MT 79 (84.1)</td>
<td>1 (1.0)</td>
<td>4 (4.2)</td>
<td>2 (2.1)</td>
<td>2 (2.1)</td>
<td>1 (1.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>SN 28 (87.5)</td>
<td>0 (0.0)</td>
<td>2 (6.3)</td>
<td>0 (0.0)</td>
<td>1 (3.1)</td>
<td>1 (3.1)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td><strong>Operative Notes (39)</strong></td>
<td>SR 947 (77.4)</td>
<td>4 (0.3)</td>
<td>19 (1.6)</td>
<td>93 (7.6)</td>
<td>7 (0.6)</td>
<td>1 (0.1)</td>
<td>4 (0.3)</td>
<td>1 (0.1)</td>
</tr>
<tr>
<td></td>
<td>MT 81 (67.5)</td>
<td>0 (0.0)</td>
<td>4 (3.3)</td>
<td>15 (12.5)</td>
<td>2 (1.7)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td></td>
<td>SN 72 (75.0)</td>
<td>0 (0.0)</td>
<td>3 (3.1)</td>
<td>9 (9.4)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (1.0)</td>
</tr>
<tr>
<td><strong>All Notes (169)</strong></td>
<td>SR 5507 (82.2)</td>
<td>172 (2.6)</td>
<td>144 (2.2)</td>
<td>132 (2.0)</td>
<td>105 (1.6)</td>
<td>63 (0.9)</td>
<td>41 (0.6)</td>
<td>33 (0.5)</td>
</tr>
<tr>
<td></td>
<td>MT 284 (69.1)</td>
<td>15 (3.7)</td>
<td>17 (4.1)</td>
<td>19 (4.6)</td>
<td>12 (2.9)</td>
<td>3 (0.7)</td>
<td>15 (3.7)</td>
<td>3 (0.7)</td>
</tr>
<tr>
<td></td>
<td>SN 205 (70.4)</td>
<td>6 (2.1)</td>
<td>14 (4.8)</td>
<td>11 (3.8)</td>
<td>9 (3.1)</td>
<td>3 (1.0)</td>
<td>15 (5.2)</td>
<td>3 (1.0)</td>
</tr>
</tbody>
</table>

1 n = number of errors; % = number of errors of a specific type divided by the total number of errors
Clinical Information Errors in SR Notes

Error rate

Discharge summaries
Operative notes
Office notes

- Medication
- Diagnosis
- Lab
- Procedure
- Symptom
## Clinical Information Errors Across Note Stages

<table>
<thead>
<tr>
<th>Total Errors</th>
<th>Clinical Information Errors n (%)</th>
<th>General English Errors n (%)</th>
<th>Other Errors n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SR</strong></td>
<td>6703</td>
<td>691 (10.3)</td>
<td>5507 (82.2)</td>
</tr>
<tr>
<td><strong>MT</strong></td>
<td>411</td>
<td>84 (20.4)</td>
<td>284 (69.1)</td>
</tr>
<tr>
<td><strong>SN</strong></td>
<td>291</td>
<td>61 (21.0)</td>
<td>205 (70.44)</td>
</tr>
</tbody>
</table>

Other errors include patient and provider information, dates and ???.

- 40% of SR notes, 7% of MT notes, and 5% of SN notes contain at least one clinically significant error.
## Content Rearranging and Stylistic Changes

<table>
<thead>
<tr>
<th></th>
<th>Medical Transcriptionist</th>
<th>Clinician</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rearranged text</td>
<td>17.4%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Made stylistic changes</td>
<td>91.7%</td>
<td>43.1%</td>
</tr>
<tr>
<td>Added information</td>
<td>N/A</td>
<td>29.7%</td>
</tr>
<tr>
<td>Deleted information</td>
<td>N/A</td>
<td>24.8%</td>
</tr>
</tbody>
</table>

% = number of notes where changes were made divided by total number of notes.
Error Examples

- **TB vaccine vs. TD vaccine.**
- **Staining vs. standing.**
  - “Continues to have daily standing” (pt’s menstruation).
  - SR, transcriptionist, and signed note all missed this error.
- **Menorrhagia and gluten allergy** were missed by SR and transcriptionist, and remained omitted on the signed note.
- SR and transcriptionist missed the name of the drug and listed as ??__??
  - The drug was *celecoxib*. The SR and transcriptionist notes did not record it. The signed note listed drug as *naproxen*!
Discussion: Productivity and Quality

• Back-end dictation service had a relatively quick turnaround time and a low error rate.

• While many errors were generated by SR, most (~94%) were corrected by the medical transcriptionist manually.
  ▶ The addition of a human editing an SR-generated note is invaluable.

• 20% of EHR-related malpractice cases were due to incorrect information in the EHR.¹
  ▶ Without the MT revising the notes, clinically significant errors could have had a negative impact on patient care and potentially caused legal issues.

Discussion: Error Checking

• Errors still left in the signed note suggest that some providers may not review their dictated notes thoroughly or at all.
  ▶ 7% of signed notes contained a blank space the transcriptionist marked as ??___??

• If physicians use SR directly, they may have to spend a considerable amount of time correcting the SR-generated text.
  ▶ Although the errors might be less than our results, since the SR can be trained by the individual physician.

• Automated error detection may help improve the accuracy of dictated documents.
Ongoing and Next Steps

• Conduct error analysis for front-end user of SR.
  ► Clinical observations
  ► Simulations

• Build a knowledge base.
  ► Confusion sets
  ► Error frequencies
  ► Error patterns

• Develop automated methods to detect SR errors.
  ► Statistical methods (noisy channel models, co-occurrence statistics), machine learning, and knowledge-based methods.
Acknowledgements

• This study is funded by AHRQ R01HS024264.
• SR Research Team
  ► David Bates
  ► David Mack
  ► Evgeni Kontrient
  ► Foster Goss
  ► Jessica Huynh
  ► Kenneth Lai
  ► Leigh Kowalski
  ► Marie Meteer
  ► Ray Doan
  ► Shiri Hassid
  ► Suzanne Blackley
  ► Warren Acker
• BWH Dragon Steering Committee, BWH Health Informatics Services, Partners eCare Clinical Informatics Team, and Partners Clinical Quality & Analysis Team
Contact Information

Li Zhou

LZHOU2@PARTNERS.ORG
How to Submit a Question

• At any time during the presentation, type your question into the “Q&A” section of your WebEx Q&A panel.
• Please address your questions to “All Panelists” in the drop-down menu.
• Select “Send” to submit your question to the moderator.
• Questions will be read aloud by the moderator.
Obtaining CME/CE Credits

If you would like to receive continuing education credit for this activity, please visit

http://hitwebinar.cds.pesgce.com/eindex.php