Usability Evaluation of Family Physicians’ Interaction with the Comorbidity Ontological Modeling and ExecuTion System (COMET)

Samina R Abidi\textsuperscript{a}, Samuel Stewart\textsuperscript{b}, Michael Shepherd\textsuperscript{b}, Syed Sibte Raza Abidi\textsuperscript{b}

\textsuperscript{a} Medical Informatics Program, Faculty of Medicine
\textsuperscript{b} NICHE Research Group, Faculty of Computer Science
Dalhousie University, Halifax, Canada

August 22\textsuperscript{nd} 2013
Usability Evaluation

- Interaction between a CDSS and user is a highly personalized activity

- It is best to collect usability data directly from the users (Kushnirk, AW, Patel VL. 2004, Nielsen, J. 1994)
  - During active interaction with the system
  - While carrying out tasks that are representative of actual use

- Usability Problems (Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH, 2006)
  - Disrupts users’ clinical workflow
  - Cause time and money
  - Result in poor acceptance

- Usability data can provide a valuable opportunity to expose specific problems
<table>
<thead>
<tr>
<th>Usability Eval Methods</th>
<th>Description</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys/Questionnaires</td>
<td>To study preferences and subjective satisfaction of users through predetermined items</td>
<td>Ease of distribution and analysis</td>
<td>Low value in identification of emerging usability issues</td>
</tr>
<tr>
<td>Focus groups</td>
<td>A moderator guides a discussion with a group of users of the application</td>
<td>Data collected directly from end users</td>
<td>Low value in identification of emerging usability issues</td>
</tr>
<tr>
<td>Retrospective Interviews</td>
<td>Users are interviewed to find out about their experiences</td>
<td>Be better at obtaining detail information from users</td>
<td>Low value in identification of emerging usability issues in context of actual system use</td>
</tr>
<tr>
<td>Heuristic Evaluation</td>
<td>Evaluators examine the interface and judge its compliance with recognized usability principles. Can be used throughout development</td>
<td>Application of recognized and accepted principles</td>
<td>Dissociation from the end-users. Unreliable for identification of domain-specific problems</td>
</tr>
<tr>
<td>Cognitively Walkthrough</td>
<td>Task-specific usability inspection method. Evaluators walk through the selected tasks in step-wise fashion. Data about system response is reported</td>
<td>No need for end-users or fully functioning prototype for the study</td>
<td>Very tedious. Potential bias due to improper task selection</td>
</tr>
<tr>
<td>Think aloud protocols</td>
<td>Participant continuously verbalizes his/her thoughts while interacting with a system performing tasks that are representative of the actual. Allows the researchers to capture end-user's cognitive responses</td>
<td>Observation of these tasks and the verbalized thought processes help researchers to identify the underlying cognitive responses of the users to the technology</td>
<td>Thinking aloud may be distracting to some participants</td>
</tr>
</tbody>
</table>

Inquiry

Observation/Inspection

Testing
Background: COMET System

- Help gather relevant clinical data and provide evidence-based diagnostic and therapeutic assistance to family physicians (FP), while they manage patients with comorbid CHF and AF
- COMET is based on Canadian clinical practice guidelines (CPG) for CHF and AF
- Has been developed in collaboration with cardiologists at the QEII hospital in Halifax
- Designed to assist FP manage patients with less severe symptoms
COMET: Framework

- Health Knowledge Management Framework

1. **Generation** of CP (derived from CPG)
   - Knowledge Identification
   - Knowledge Synthesis

2. **Development** of a ‘unified’ ontologically modeled CP to handle comorbidities
   - Knowledge Formalization

3. **Alignment** of temporal and procedural aspects of computerized comorbid CP
   - Knowledge Alignment

4. **Execution** of computerized CP at point of care for clinical decision support
   - Knowledge Execution
COMET Functionalities

- **Diagnostic Assistance**
  - For CHF and Comorbid CHF-AF
  - Use of NYHA functional classification to identify patients with milder symptoms
  - Automated computation of Boston Criteria Score (Symptoms + Signs + CXR)
    - Better identification of patients for echo

- **Therapeutic Assistance**
  - Parallel therapies (ACEI/ARB, BB, Diuretics, Anticoagulation)
  - Uptitrations protocols nested within parallel treatment plans
    - Manage uptitrations simultaneously

- **Recommendation is supported by best evidence from CPG + strength of the evidence**
Preliminary Evaluation

1. **Evaluation of modeling of knowledge (ontology)**
   - Consistency & Conciseness using Pellet
     - Subsumption tests (Consistent, no contradictory information, no useless definitions)
   - Completeness
     - Adequate representational capacity to capture comorbid domain and procedural concepts

2. **Evaluation of functionality of COMET**
   - Ability of COMET to provide decision support for CHF and CHF-AF
     - Able to handle single disease and comorbid scenarios

3. **Evaluation of correctness of knowledge**
   - Domain experts: A cardiologist and two GPs
   - Three separate interview style informal sessions
   - Walk through the application, responses recorded
Research Objectives

• To assess how well the COMET system meets the functional goals and usability needs of family physicians (FP)
  - when it comes to the management of patients who have CHF and comorbid CHF-AF

• To receive end-user feedback in order
  - To identify potential areas of modifications that are needed to improve content, interface design, and tasks related usability
    - in order to enhance its overall usability
Study Design

• Background and post-interaction questionnaires &
• Think Aloud Protocol methods

• Based on Usability framework
  – Comprehensive and grounded in theories and methodologies
  • From cognitive sciences and usability engineering
Recruitment Procedure

- **Ethics**
  - From Dalhousie’s HSREB

- **Recruitment**
  - NS FP
  - Addresses obtained from NS CPS FP’s directory
  - Mail and e-mail
    - 155 recruitment notices
    - Study Web-site – URLs e-mailed
  - Direct phone calls
  - Direct visits to FP offices

- **Venue**
  - FP office
    - All interactions took place of FP’s laptop or desktop computers

- **Sample Size**
  - 10 FP
  - “Up to 75% usability problems can be uncovered within first 5 users, and up to 85% by the 8th users, after which yield drops” (Kushnirk & Patel (2004) and Vizri (1992))

- **Inclusion Criteria**
  - All licensed FP in NS
  - Have access to internet

- **Exclusion Criteria**
  - Not licensed in Canada
  - No internet in the office
Data Collection & Survey Instruments

2 methods of data collection

Questionnaires

Pre-interaction
(Background)

Purpose: to collect personal and professional data

Post-interaction

Purpose: To have an overall assessment of the COMET system’s ability to meet the functional goals and usability need of FP

Screen design layout and navigation

Learnability

Think Aloud Protocols

Purpose: To record direct observations along with end user thought processes during actual interactions with the system

Task related usability

Overall satisfaction and recommendation

Content adequacy, understandability and helpfulness

Task related usability

Overall satisfaction and recommendation
# Study Explained

# Informed Consent

# Pre-interaction questionnaire

# 5-10 min training session

# 3 rep scenarios explained and given

# Participants’ interaction with COMET

# Post-Interaction Questionnaire

## Procedure

### Part 1 - Please fill in the following information:

1. **Age:**

2. **Gender:**

3. **Years since you graduated from medical school:**

4. **Your current medical record system is**
   - Paper-based
   - Electronic
   - Combination

5. **If your current medical record system is electronic, does it have any decision support capabilities?**
   - No
   - Yes

6. **If your answer is Yes to question number 5, then which of the following decision support capabilities does your electronic medical record system have? (Check all that apply)**
   - Identifying possible drug interactions
   - Alerts and prompts to ensure regular screening of the patient
   - Alerts and prompts to facilitate therapeutic care planning
   - Alerts and prompts to facilitate diagnosis
   - Alerts and prompts to ensure compliance with clinical practice guidelines
   - Other decision support capability: please specify ___________

7. **Do you seek clinical practice guidelines to aid patient care?**
   - No
   - Yes

8. **Do you seek patient care information on-line?**
   - No
   - Yes

9. **How would you best characterize your practice? (Please check only one)**
   - Solo primary care practice
   - Solo specialty care practice
   - Primary care group or partnership
   - Single specialty group or partnership
   - Multi-specialty group or partnership (including staff or group model HMOs)

10. **How comfortable do you feel using computers in general?**
    - Very comfortable

---

The entire Procedure took approx 60 min

---

Participants’ Think Aloud. Screen and thoughts/comments are recorded manually
Case Scenarios

3 Representative case Scenarios
1. Pt. has CHF
2. Pt. has comorbid CHF-AF
3. Inconclusive Diagnosis

Follow the therapeutic plans/referral from here

Reflect essential steps and associated outcome(s)

Each step/outcome combo represents functional goals and usability need of a FP

Developed using CPG/med literature by the researcher
## 1. Learnability

**Learnability: How easy is it to learn and use COMET**

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree 1</th>
<th>Moderately agree 2</th>
<th>Neither agree or disagree 3</th>
<th>Moderately disagree 4</th>
<th>Strongly disagree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## 3. Content

**Content helpfulness, usefulness and understandability**

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree 1</th>
<th>Moderately agree 2</th>
<th>Neither agree or disagree 3</th>
<th>Moderately disagree 4</th>
<th>Strongly disagree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Questionnaires**

- 37 items in all
- Ranges from 10 (Task related usability to 5 (overall satisfaction)
- Items scored on 5 point Likert scale 1-5
- All items are positively worded
Results
Study Demographic & Background Data

Mean age = 43 years
Age ranges from 35 to 53 years
Mean Years since graduation = 16.4
Years of graduation ranges from 4 to 29 years

The demographics revealed a group of users that have EMRs in particular and computers in general highly integrated into their practice
Qualitative Analysis of Think Aloud Protocols Data & Results

Each protocol was read line by line by at least 2 researchers

A comprehensive list of all the issues encountered by the users along with the frequency with which they were encountered was generated

<table>
<thead>
<tr>
<th>No.</th>
<th>Screen</th>
<th>Usability and Content Issues in Think Aloud Protocols</th>
<th>No of Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Demographic Information</td>
<td>Entering the date of birth takes too long, especially that most heart failure patients are quite old</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I see many patients every day, entering the demographic information takes too long</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It should be hooked with my EMR, so that the demographic information enters the tool directly and I don’t have to do that again</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It’s too time consuming, I see about 30 patients per day</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>History and Physical Exam</td>
<td>High blood pressure should be an option in the drop down menu for the blood pressure</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Why isn’t the bradycardia option in the drop down list for the pulse rate</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The normal option should be blocked when you have already selected an abnormal sign or symptom</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Its all cardiologists’ stuff, I never do JVP and Carotid Pulse in my practice. I don’t have time to do this</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fourth Heart Sound should also be an option</td>
<td>1</td>
</tr>
<tr>
<td>3.</td>
<td>Assessment of Tests results</td>
<td>Negative or positive is the way we usually report instead of Normal and Abnormal options in the tool</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shouldn’t there be a not done option, because I never do BNP, at least should have been with BNP</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not very intuitive if I want to go to the tabs for the next test, perhaps they should be color coded</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Information about each test should be on a page before</td>
<td>1</td>
</tr>
</tbody>
</table>
**Data Coding**

We prepared *a priori* list of **categories of potential usability problems** based on Neilson usability heuristics.

Constant comparative approach was employed for the coding of the data in order to group data, which was then compared against the usability categories.

<table>
<thead>
<tr>
<th>No.</th>
<th>Usability Problem Categories</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Time and effort of data entry</td>
<td>Requires a lot of time and effort for data entry</td>
</tr>
<tr>
<td>2.</td>
<td>Interface consistency</td>
<td>Users should not have to wonder whether different words, situations, or actions mean the same thing</td>
</tr>
<tr>
<td>3.</td>
<td>Inadequate and less useful information content</td>
<td>The system provides inadequate and less useful information for the users to be able to complete their tasks properly</td>
</tr>
<tr>
<td>4.</td>
<td>Information content less comprehensible</td>
<td>The system does not use the users' language, i.e., words, phrases and concepts that are familiar to the user</td>
</tr>
<tr>
<td>5.</td>
<td>Cognitive overload</td>
<td>The user have to remember information from one part of the dialogue to another</td>
</tr>
<tr>
<td>6.</td>
<td>Navigation through the protocols and finding desired information</td>
<td>Information does not appear in logical order and it is not be easy to find the desired information</td>
</tr>
<tr>
<td>7.</td>
<td>Visibility of system status</td>
<td>The system does not always keep users informed about what is going on, through appropriate feedback within reasonable time</td>
</tr>
<tr>
<td>8.</td>
<td>Inadequate user control</td>
<td>User is unable able to exit the unwanted state without having to go through an extended dialogue</td>
</tr>
<tr>
<td>9.</td>
<td>Error prevention</td>
<td>Error-prone situations are not abolished and users are not presented with a confirmation option before they commit to an action</td>
</tr>
<tr>
<td>10.</td>
<td>Error recognition and recovery</td>
<td>Error messages are not expressed in plain language or does not suggest a solution</td>
</tr>
</tbody>
</table>
Classification & grouping of data into 10 usability categories

In order to ensure reliability of these categories, we took into account the frequency and consistency with which the family physicians stipulate them in the protocols.

The appearance of an issue in a category during coding consistently and frequently means that this issue/category is more credible and should be taken into account when improving the design of the COMET tool.

<table>
<thead>
<tr>
<th>No.</th>
<th>Usability Problem Category</th>
<th>No of Usability Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Inadequate and less useful information content</td>
<td>10 usability issues</td>
</tr>
<tr>
<td>2.</td>
<td>Navigation through protocols and finding desired information</td>
<td>6 usability issues</td>
</tr>
<tr>
<td>3.</td>
<td>Time and effort of data entry</td>
<td>4 usability issues</td>
</tr>
<tr>
<td>4.</td>
<td>Information content less comprehensible</td>
<td>3 usability issues</td>
</tr>
<tr>
<td>5.</td>
<td>Cognitive overload</td>
<td>1 usability issue</td>
</tr>
<tr>
<td>6.</td>
<td>Visibility and system status</td>
<td>1 usability issue</td>
</tr>
<tr>
<td>7.</td>
<td>Error prevention</td>
<td>1 usability issue</td>
</tr>
</tbody>
</table>
**Most Frequently Occurring Usability Categories**

**10 usability problems**
Inadequate and less useful information content

- High blood pressure should be an option in the drop down menu for the blood pressure (7 times)
- Bradycardia should be an option in the drop down list for the pulse rate (6 times)
- There should be a resource or a link for the patient for free water restriction or low sodium diet (2 times)

**6 usability problems**
Navigation through protocols and finding desired information

- Sodium and potassium tests should be presented sequentially. We usually order them at the same time. Realistically you order all the electrolytes at the same time and recheck them at the same time if needed (8 times),
- It is not very intuitive to select contraindication to beta-blocker or ARB tab to access the list of contraindications. The contraindication list should be intuitively displayed and then you should select them (7 times)
- I have to select the sign of overload present button to display the list of overload signs. This is not very intuitive, shouldn’t be it other way round (6 times).

**4 usability problems**
Time and effort of data entry

- Entering the date of birth takes too long, especially that most heart failure patient are quite old (7 times)
- I see many patients every day, entering the demographic information takes too long (4 times)
- It should be hooked with my EMR, so that the demographic information enters the tool directly and I don’t have to do that again (3 times).
Quantitative Analysis of Survey Data & Results

Internal Consistency and Reliability of the Responses

We used Cronbach’s alpha to evaluate consistency of the responses within streams.

Whether the questions in the questionnaire all reliably measure the same latent variable that is inferred from the responses.

An alpha > 0.8 is generally used as evidence that survey items measure an underlying variable and indicate good consistency.

<table>
<thead>
<tr>
<th>Cronbach's Alpha</th>
<th>Learnability</th>
<th>Design &amp; Navigation</th>
<th>Content</th>
<th>Task Related Usability</th>
<th>Overall Satisfact.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9345</td>
<td>0.9244</td>
<td>0.7944</td>
<td>0.8164</td>
<td>0.9599</td>
<td></td>
</tr>
</tbody>
</table>

Overall, the survey responses have sufficient internal consistency and reliability.
Line plots depicting individual responses to the five survey categories

To interpret the consistency of the responses in the survey:

Clock-wise from: Learnability, Screen Layout and Navigation, Content Helpfulness, Task-Related Usability and Overall Satisfaction

the Tasks-Related Usability were the least consistent section of the survey

Moreover, the purple user was the least satisfied overall

We can also conclude that there were no specific statements within any survey category that saw a significant dip or climb in ratings
Mean scores for each category along with Confidence Interval. (1=Strongly Agree, 5 = Strongly Disagree)

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean Score (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learnability</td>
<td>1.425 (1.072, 1.778)</td>
</tr>
<tr>
<td>Design and Navigation</td>
<td>1.456 (1.012, 1.900)</td>
</tr>
<tr>
<td>Content</td>
<td>1.411 (1.066, 1.7575)</td>
</tr>
<tr>
<td>Task Related Usability</td>
<td>1.391 (1.032, 1.75)</td>
</tr>
<tr>
<td>Overall Satisfaction</td>
<td>2.040 (1.293, 2.787)</td>
</tr>
</tbody>
</table>

For all the categories the average score was between strongly agree and moderately agree indicating that the participants were satisfied with the system.

There is a drop in overall satisfaction with the system, however, it still averaged a score of moderately agree.
Histogram showing individual responses in each category

All the categories are skewed positively (P-values < 0.001), indicating that the users responded positively to COMET tool.

‘Task Related Usability’ received highest ‘Strongly Agree’ score as compared to ‘Overall Satisfaction and Recommendation’ that received lowest ‘Strongly Agree’ score.

Responses to the overall satisfaction category were not as positive as the rest of the four individual categories.

<table>
<thead>
<tr>
<th>Survey Category</th>
<th>Strongly Agree</th>
<th>Mod. Agree</th>
<th>Neither Agree or Disagree</th>
<th>Mod. Disagree</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learnability</td>
<td>50</td>
<td>26</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Screen design layout and navigation</td>
<td>60</td>
<td>23</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Content adequacy, understandability and helpfulness</td>
<td>70</td>
<td>22</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Task related usability</td>
<td>78</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Overall satisfaction and recommendation</td>
<td>20</td>
<td>16</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Individual likert scale scores for each of the survey categories.
Comparing survey scores for all 5 categories with the Years of Practice.

Comparing survey scores for all 5 categories with the Age of the family physicians.

There is no evidence that either impacts how doctors responded to the survey.
Comparing Scores to the Current Medical Record Method used in the Practice

Linear regression results predicting survey category scores based on current record storage system.

There is a noticeable linear decrease in scores, which corresponds to an increase in appreciation for users that partially or completely use electronic medical records.

As expected, the lowest scores were from the single respondent who used a paper-based record system.

Highest from the pure electronic group, with the combination group in between.

Distribution of scores for all five categories by the medical record type.
Depicting frequency of Likert scale scores for each statement within individual survey categories

Overall the bar charts show that most participants either strongly or moderately agreed with most of the statements in the 5 categories. None of the participant strongly disagrees with any of the statement in any of the five themes of the survey.

Overall satisfaction & recommendation

Task-related usability

Content

Learnability

Screen design layout & navigation
Limitations

• Small sample size

• Developers of the COMET system carried out its usability evaluation
  – Strength: Developers while testing the system are able to elicit more information from the participants regarding any issues they have while interacting with the system

• Interactions just after 15 min training session
Conclusions & Recommendation

• The questionnaire method revealed that the FP are potential users of the clinical decision support software like COMET

• The Think Aloud approach allowed us to uncover specific usability-related concerns that could not have been apparent from the questionnaires alone

• Adding a more interactive methodology such as Think Aloud Protocol to questionnaires or surveys is very useful to uncover specific usability-related concerns

• These usability issues need to be addressed before deployment of a CDSS and the subsequent outcome evaluation
Acknowledgement

This research has been supported by grant from Green Shield Foundation, Canada

Thank you