Observed Usage Errors During Meaningful-Use Stage 3 Safety-Enhanced Design Summative Testing

Jennifer Mauney1, Janey Barnes1, David Clarke1, Caleb Furlough1,2, Erika Orrick1, Teri Brooks1, Megan Brown1, Olga Zielinska1,2, and John Grishin1,2

1User-View, Inc.
2North Carolina State University

This report presents data from summative usability tests conducted by User-View on behalf of multiple vendors as part of MU2 and MU3 certification. The objective is to present findings related to performance metrics and use errors associated with each prioritized certification criteria, shine the light on EHR effectiveness, and contribute to ongoing discussions of EHR usability.

INTRODUCTION

The Office of the National Coordinator (ONC) through its 2015 Edition Health IT Certification, also called Meaningful Use 3 (MU3) certification program, calls for a focus on improved usability in conjunction with improved safety and risk management (“2015 Safety-Enhanced Design Test Procedures,” 2015). Known as Safety-Enhanced Design, formative and summative user-centered design (UCD) activities must take place during the design and development of prioritized certification criteria and associated capabilities designated by ONC. A certification program, focused on improved usability in conjunction with improved safety and risk management, was also included in the ONC’s Meaningful Use Stage 2 (MU2) program (“Test Procedure for §170.314(g)(3) Safety-enhanced design,” 2014)

Prioritized certification criteria required as part of MU3 included:

- Computerized Provider Order Entry (CPOE) - medications (§170.314(a)(1))
- Computerized Provider Order Entry (CPOE) - laboratory (§170.314(a)(2))
- Computerized Provider Order Entry (CPOE) – diagnostic imaging (§170.314(a)(3))
- Drug-drug, drug-allergy interaction checks (§170.314(a)(4))
- Demographics (§170.314(a)(5))
- Problem list (§170.314(a)(6))
- Medication list (§170.314(a)(7))
- Medication allergy list (§170.314(a)(8))
- Clinical decision support (CDS) (§170.314(a)(9))
- Electronic medication administration record (eMAR) (§170.314(a)(16))
- Electronic prescribing (§170.314(b)(3))
- Clinical information reconciliation (§170.314(b)(4))
- Implantable devices (§170.314(a)(14))
- Clinical decision support (CDS) (§170.314(a)(9))
- Electronic prescribing (§170.314(b)(3))
- Clinical information reconciliation (§170.314(b)(4))

Prioritized certification criteria required as part of MU2 included:

- Computerized Provider Order Entry (Medications, Laboratory, Diagnostic Imaging) (CPOE) (§170.314(a)(1))
- Drug-drug, drug-allergy interaction checks (§170.314(a)(2))
- Medication list (§170.314(a)(6))
- Medication allergy list (§170.314(a)(7))
- Clinical decision support (CDS) (§170.314(a)(8))
- Electronic medication administration record (eMAR) (§170.314(a)(16))
- Electronic prescribing (§170.314(b)(3))
- Clinical information reconciliation (§170.314(b)(4))
- Implantable devices (§170.314(a)(14))
- Clinical decision support (CDS) (§170.314(a)(9))
- Electronic prescribing (§170.314(b)(3))
- Clinical information reconciliation (§170.314(b)(4))

METHODS

This report presents data from summative usability tests conducted by User-View on behalf of multiple vendors as part of MU2 and MU3 certification.

- Five hundred seventy-six (576) participants took part in the testing of twelve (12) EHRs.
- Participants included physicians, nurses, registration specialists, and configuration specialists.
- Products tested included ambulatory, acute, and module EHRs.
- Test scenarios included multiple subtasks.

Test environments varied across the twelve (12) summative usability tests. All sessions were moderated single participant sessions.

This report aggregates data from measures of effectiveness. A task was counted as a “Pass” if the participant was able to achieve the correct outcome, without assistance. A task was counted as a “Pass with help” if the participant was able to achieve the correct outcome with one navigation assist from the moderator. “Pass” and “Pass with help” were combined to be Task Success.
A task was counted as “Fail” if the participant abandoned the task, did not complete the task goal, did not complete the task goal in the allotted time or needed more than one navigation assist from the moderator. Failed tasks were discussed with participants at the end of each task.

RESULTS

Figure 1 presents a scatter plot of the percent task failures x critical error risk score. Task failure was calculated as 1- task success. For each product, the criterion subtask with the lowest task success rate and the associated number of participants for that subtask was used to calculate that criterion’s percent task success. For example, the Medication List criterion included subtasks associated with record, change, and access a patient’s active medication list as well as medication history. The subtask with the lowest task success rate was selected to represent scenario performance because that number is the most accurate representation of the number of participants that failed the scenario.

Task failures were subjected to an error analysis. Error analysis included identification of the root cause of observed errors, classification of severity based on the consequence of each use error, and identification of the frequency each use error was observed during summative testing. Identification of the root cause of each error was identified by human factors specialists and verified through discussion with a multi-discipline team. The multi-discipline team included human factors specialists and vendor stakeholders. Vendor stakeholders included product specialists and clinical specialists.

Classification of the severity of each use error was determined through discussion with a multi-discipline team. The primary factor for determining severity was the clinical consequence of the use error. Use errors resulting in subtask failures that might impact patient safety were categorized as critical and assigned a criticality weight of 3. Use errors resulting in subtask failures that are known industry risk issues and/or related to aspects of the user interface that are configured per customer site were categorized as artifacts of testing and assigned a criticality weight of 2. Use errors impacting efficiency were categorized as noncritical and assigned a criticality weight of 1. This report does not include presentation or discussion of use errors impacting efficiency.

As seen in the scatterplot (Figure 1), criteria in the top right quadrant of the graph (Implantable Devices, eMAR, Demographics, and Drug/Drug Configuration) are ranked with relatively higher risk scores and higher percent fails compared to the other criteria. Criteria in the bottom left quadrant of the graph (Medication List, Medication Allergy List, Problem List, CPOE – Laboratory, and CPOE – Diagnostic Imaging) are ranked with relatively lower risk scores and lower percent fails compared to the other criteria. Drug/Drug Configuration and ePrescribe are ranked with the highest risk score relative to the other criteria but ranked in the higher to midrange in percent fails. Drug/Drug Clinical is ranked with relatively low percent fails but with a relatively high risk score.

![Figure 1. Combined MU2 and MU3 Percent Task Fails by Critical Error Risk Scores.](image-url)
DISCUSSION

Task Performance and Error Analysis
Opportunities for improvement and enhancement are available by applying Neilsen’s (1995) usability heuristic guidelines, and Lowry, Quinn, Ramaiah, Schumacher, Patterson, North, Zhang, Gibbons, & Abbott (2012) and Lowry, Ramaiah, Taylor, Patterson, Prettyman, Simmons, Brick, Latkany, & Gibbons (2015) specific guidance to EHR design.

Drug/Drug Configuration and Clinical Decision Support-Configuration
- Opportunities to apply heuristic principles enhancement (Nielsen, 1995) for improvement:
  - Error prevention
  - Help users recognize, diagnose and recover from errors
  - Help and documentation
  - Provide error messages expressed in plain language that precisely indicate the problem and constructively suggest a solution
  - Provide accessible, searchable help and documentation with concrete steps and explanations

Electronic Prescribing
- Highlights the need for mature UCD processes for more complex and higher risk tasks. Test participants had difficulty with the more complex prescribing tasks.
- Opportunities to apply heuristic principles enhancement for improvement:
  - Accurate data display (Lowry et al, 2012)
  - Assure visibility of system status (Nielsen, 1995)
  - Error prevention (Nielsen, 1995)
  - Help users recognize, diagnose, and recover from errors when reliable systems fail (Nielsen, 1995)
  - Speak the user’s language, follow real-world conventions and make information appear in a natural and logical order. (Nielsen, 1995)

Implantable Devices
- Highlights the importance of consistent design of similar functionality across different parts of the application and the need for mature UCD processes.
- Opportunities to apply heuristic principles enhancement for improvement:
  - Consistency and standards (Nielsen, 1995)
  - Help users recognize, diagnose, and recover from errors (Nielsen, 1995)
  - Pleasurable and respectful interaction with the user (Lowry et al, 2012)

Electronic Medication Administration Reconciliation (eMAR) (MU2 criterion only)
- Multiple environments (e.g., computer workstation, medication storage areas, bedside) and multiple technologies (e.g., EHRs, bar code scanners) contribute to the complexity.
- Findings underscore the importance of shared partnership and responsibility between vendors and their customers as the EHR becomes a technology in the medication administration workflow. Apply User Centered Design activities to the design of the medication administration workflow.

Demographics
- Highlights the importance of implementing simple, usable controls and the need for mature UCD processes.
- Opportunities to apply heuristic principles enhancement (Nielsen, 1995) for improvement:
  - Consistency and standards
  - Speak the user’s language, follow real-world conventions and make information appear in a natural and logical order.

Clinical Information Reconciliation and Incorporation
- More test participants were able to complete the reconciliation task since MU2, and there were still users who were able to complete the task scenario mechanically, but did not have a clear understanding of the interaction model of the feature.
- Opportunities to apply heuristic principles enhancement (Nielsen, 1995) for improvement:
  - Match between system and the real world
  - Speak the user’s language, follow real-world conventions and make information appear in a natural and logical order.

Computer Provider Order Entry (CPOE) - Medications
- Highlight industry known issues with ordering similarly named medications and lack of entering detailed information as discrete data.
- Opportunities to apply heuristic principles enhancement for improvement:
  - Accurate data display (Lowry et al, 2012)
  - Visibility of system status (Nielsen, 1995)
  - Error prevention (Nielsen, 1995)
  - Help users recognize, diagnose, and recover from errors (Nielsen, 1995)

Drug/Drug and Drug/Allergy Interaction Checks (Drug/Drug Clinical) and Clinical Decision Support (CDS-Clinical)
- Both capabilities highlight the industry known issue of alert fatigue.
- Alert fatigue provides an opportunity for the industry to refocus efforts to develop and share industry best practices, guidelines, and templates regarding safety-enhanced design associated with addressing alert fatigue. Sharing across vendor and provider communities should take place.
- CDS solutions should be considered based on the model of the workflow. Some CDS solutions interrupt the workflow; appropriately and inappropriately. Some CDS solutions are available when and if the provider decides to access the information.

Medication Allergy List, Medication List, Problem List, Computer Provider Order Entry (CPOE) – Laboratory, and
Computer Provider Order Entry (CPOE) – Diagnostic Imaging

- Considered mature EHR functionality in the EHRs included in this report.
- Opportunities to apply heuristic principles enhancement for improvement:
  - Accurate data display (Lowry et al., 2012)
  - Assure visibility of system status (Nielsen, 1995)
  - Help users recognize, diagnose, and recover from errors (Nielsen, 1995)

Learnings from MU2 to MU3

A number of learnings were gained after performing multiple MU2 summative usability tests and multiple MU3 summative usability tests.

- In general, areas that showed greater failure rates and higher risk scores continued this trend from MU2 to MU3. Exceptions:
  - Electronic Prescription had more complex tasks than MU2 testing.
  - CPOE-Medications was separated from CPOE-Laboratory and CPOE-Diagnostic Imaging from MU2.
  - A better application of a User Centered Design (UCD) process should be incorporated to assure improvements in the areas of error prevention and user experience.
  - Artifacts of testing must be reported and they can affect the understanding of some use errors.
- One criterion from MU2, Electronic Medication Administration Record (eMAR) was not included in MU3. This criterion MUST be included for testing based on calculations of failure rate and critical error risk score from MU2 data.
- Continued education is needed across all stakeholders’ groups (e.g., regulatory bodies, EHR buyers, EHR end users, and the vendor community).

The Barriers to Meaningful Dialog and Productive Use

A number of barriers to having a meaningful dialog and productive use of the summative test reports were identified.

- Summative usability test reports are available on the ONC website, [https://chpl.healthit.gov/#/search](https://chpl.healthit.gov/#/search).
- Finding the reports is quite difficult. Once one arrives at the ONC CHPL site ([https://chpl.healthit.gov/#/search](https://chpl.healthit.gov/#/search)), to see all the vendors who have submitted a MU3 usability report, multiple filters need to be applied to the site. To see a specific MU3 summative usability report, one needs to perform a multi-step navigation process in which the usability of the website can hinder completing the navigation correctly.
- At the time of this writing, very few MU3 reports are published on the website. Of the reports that are published, many reports are not available or not accessible due to broken or incorrect links.
  - A handful of ICSA Labs certified product reports appear to be available except two: one link is an incorrect link and one link is broken.
- The few Drummond Group certified product reports all have broken links except one, which is an incorrect link.
- One InfoGard product report is available.
- ONC-Authorized Testing and Certification Bodies have inconsistency in the testing requirements.
- Different certifying bodies require different levels of test task scenarios to be completed to meet the requirements of an MU3 usability test.
- Feedback from the ONC showed a lack of adherence to best practices.
  - There was a lack of direction, common data, and reporting requirements for hazards/error analysis and risk assessment.
  - Unchanged capabilities/functionality were required to be re-tested and reported in summative test.
  - Discrete data was required for inclusion on the CHPL site, but providing the discrete data without context can be misleading.
  - Lack of common data scoring techniques and required reporting of inappropriate descriptive statistics was seen through the reports.
  - Summative test recommended sample size was too low.
  - As written NISTIR 7804-01 Use Cases are impractical to implement.

Because of the lack of MU3 summative usability reports on the CHPL website and because of inconsistencies in the accessible reports including the lack of required tasks, the levels of detail provided regarding tasks and scenario details, and the levels of reported error analysis, it is not clear that each public report can be used as a contributor to the types of findings reported here.

This report represents a first view into the publicly available EHR summative usability reports required as part of MU3 certification. It also represents a view of the data similarities and the learnings from MU2 to MU3. Potential benefits to multiple stakeholder groups (e.g., regulatory bodies, EHR buyers, EHR end users, and the vendor community) may be realized from further analyses. Now that EHR usability reports are publicly available, reflection and evaluation as to if and how the data can be mined is taking place.

We conclude first and foremost that continued education is needed across all stakeholder groups. Specifically, improvements can be made in the application of user-centered design and continuous processes improvement methods to assure improvements in the areas of risk analysis, error prevention and user experience.

We encourage the human factors community and related communities to keep the discussion on the forefront:

- How can these reports be used in a productive way?
• How can the data in these reports be used to make EHRs safer and more effective in support of providing quality care?
• How can the data in these reports be used to make EHRs more efficient in support of providing quality care?
• How can the data in these reports inform future policy?

REFERENCES


