

**Health IT Implementation, Usability, and Safety (IUS) Workgroup**

**Certification NPRM Comment Template (Group #2)**

| **§ 170.315(g)(3) Safety-enhanced design Page #191** |
| --- |
| **Included in 2015 Edition Base EHR Definition?**No |
| **2015 Edition EHR Certification Criterion**(3) Safety-enhanced design. (i) User-centered design processes must be applied to each capability technology includes that is specified in the following certification criteria: § 170.315(a)(1) through (10) and (18), (20), (22), (23), and (b)(2) through (4).(ii) The following information must be submitted on the user-centered design processed used:(A) Name, description and citation (URL and/or publication citation) for an industry or federal government standard; or(B) Name the process(es), provide an outline of the process(es), a short description of the process(es), and an explanation of the reason(s) why use of any of the existing user-centered design standards was impractical.(iii) The following information/sections from NISTIR 7742 must be submitted for each capability to which user-centered design processes were applied:(A) Name and version of the product; date and location of the test; test environment; description of the intended users; and total number of participants;(B) Description of participants, including: sex; age; education; occupation/role; professional experience; computer experience; and product experience;(C) Description of the user tasks that were tested and association of each task to corresponding certification criteria;(D) List of the specific metrics captured during the testing, including; task success (%); task failures (%); task standard deviations (%); task performance time; and user satisfaction rating (based on a scale with 1 as very difficult and 5 as very easy);(E) Test results for each task using metrics listed above in paragraphs (g)(3)(ii)(A) through (D) of this section;(F) Results and data analysis narrative, including: major test finding; effectiveness; efficiency; satisfaction; and areas for improvement. (iv) Submit test scenarios used in summative usability testing. |
| **Preamble FR Citation:**  | **Specific questions in preamble?** *Yes* |

Group 2 has not reached consensus on some areas of 170.315(g)(3). At a high level there are areas in 170.315(g)(3) where Group 2 members agree with the specified criteria for reporting Safety-Enhanced Design activities, there are areas where some feel the elements represent overly burdensome specificity without corresponding value, and there are some areas where the groups agrees clarification is needed.

Regarding Identify industry recognized/federal user centered design (UCD) standard:

- Some Group 2 members expressed that following a UCD process teams will be applying formative, iterative UCD activities to inform the design + conducting a summative test(s) to validate safe and effective use of production user interface. Both formative and summative activities are required as part of a UCD process.

- Other Group 2 members expressed that teams might submit artifacts from formative activities as evidence for applying a UCD process. - With focus on safety – potentially submit artifact that includes the audit trail from identifying a potential usage error through validation that mitigation decreased/removed the potential for usage error.

What artifacts provide evidence of UCD process? Should the formative artifacts be public or should the artifacts be used by the ONC-ACB review only?

Providing an artifact from a formative usability activity is not going to “make the product better”. Discuss with Group – What should teams submit? What does the industry learn for having access to these artifacts? What will stakeholders do with the knowledge from these artifacts?

Regarding how will ONC-ACBs use to evaluate if teams follow the identified UCD process e.g., for MU2 some teams as reported on the CHPL site identified following a recognized UCD process but the details in the associated summative usability test report indicated violations of known UCD practices e.g., not including representative end users in summative test, summative test sample size.

-There are a couple of issues here. First, the reports on the CHPL site are not user friendly (understandable)

- Second, the reports on the CHPL site are not always complete or high quality, as noted above

- Third, on page 196 of the document, the first full paragraph notes that “health IT developers can perform many iterations of the usability testing” and this should at least be strengthened by changing it to “health IT developers should be performing many iterations.”

- Fourth, MU2 ONC-ACB review for certification was very focused on report format and inclusion of data fields not related to safe and effective use. The document says that there are seven new certification criteria that strengthen safety-enhanced design. The inclusion of additional criteria with a safety focus is a positive effort in the right direction. However, there also needs to be an emphasis on formative evaluation. This may be covered when a Quality Management System is used, but should be specifically mentioned in this section as well.

Regarding recruiting representative test participants for each category of intended end users:

- Some Group 2 members expressed that as part of a UCD process, teams must identify the intended user of a feature/prioritized criteria and recruit that user role to participate in formative and summative activities. As an example 170.315(a)(2) Drug-Drug, Drug-Allergy Interaction Checking configuration to adjust the severity level of alerts is an activity that is typically not a part of the workflow conducted by clinical users; but instead is carried out by configuration specialists. As another example, the criteria being added for MU3, § 170.315(a)(5) Demographics, is an activity that is likely carried out by admission/check in user roles. Applying UCD activities to these types of criteria requires recruiting non-clinical users. Some Group 2 members expressed including non-clinical users was acceptable and needed. The regulation would be contradictory if it states, apply UCD process but only for clinical users.

Other Group 2 members expressed that NISTIR 7804 guidance for recruiting representative test participants for each category of anticipated clinical end users does not account for elements of the Safety-Enhanced Design criterion whose cohort of intended users are not clinical. For example, criteria covered by the rule includes 170.315(a)(2) Drug-Drug, Drug-Allergy Interaction Checking configuration to adjust the severity level and § 170.315(a)(5) Demographics which is not a part of the workflow covered by clinical users. These members recommend specification of the application of safety-enhanced design to criteria that are only tasks performed by clinical users.

Regarding the description of the participants including:

- (3)(iii)(B) The description of the participants including: sex, age, education, occupation/role, professional experience, computer experience, and product experience is taken from the NISTIR 7742 sample content. The descriptors such as sex, age, and education when collected from clinical users would not represent information that could provide evidence of correct use and application of User Centered Design procedures. IN addition, providing participant data at this level of detail and in the table format on page 11 of NIST7742 may allow for identification of specific participants. Recommend providing summary descriptor information that demonstrates the participants are representative of clinical users such as occupation/role and professional experience.

Regarding user tasks included in the test and corresponding to the certification criteria:

- Should NIST provide standard scenarios for usability test tasks so that summative test reports can be compared to each other? Should risk-based analysis of test tasks be included as part of the test task description?

Regarding specific metrics captured during testing:

- (3)(iii)(D) Task Standard Deviations (%) does not make sense in the context provided in the rule as a metric by which to measure User Centered Design. The reference standard of NISTIR 772 uses Task Deviations as a metric for effectiveness, which is not a reported as a percentage. Request clarification on whether this portion of the rule is defining the statistical procedure to use for measuring effectiveness (which would not make sense in the context of Successes or Failures) or the Effectiveness metric (which is not standard deviations, but simply task deviations).

- Task Performance Time – needs clarification – is ONC requiring Mean Task Performance Time (and SD of Task Performance Time).

- The rule guidance to use a User Satisfaction Rating with a scale of 1 to 5 as a task-based measure from the NISTIR 7742 content is suggested as an example and is not representative of an industry standard questionnaire for analyzing software usability. The primary scientifically recognized single task-based satisfaction rating is the Single Ease-of-use Question which uses a 7-point scale. Recommend supporting the ability to use literature recognized satisfaction measures such as the Single Ease-of-use Question, System Usability Scale, or Software Usability Measurement Inventory, which all use different measurement scales and not specify which satisfaction measure is used or the specification of task-based satisfaction measures versus a post-session satisfaction measures.

Regarding Results and data analysis narrative:

- (3)(F) asks for the major findings related to Effectiveness, Efficiency, Satisfaction, and Areas of Improvement. The focus is on areas identified to be at risk for error, but no requirement for error analysis as part of the report? Should a specific results section be required focused on error analysis?

Regarding areas for comment are missing in this template:

- In lieu of simply providing guidance on the number of, and user cohort for, test participants, we request comment on whether we should establish a minimum number(s) and user cohort(s) for test participants for the purposes of testing and certification to the 2015 Edition under the ONC Health IT Certification Program.

A sound UCD process requires a sufficient number of test participants so as a team is confident in making/stating their conclusions.

It is our recommendation that organizations follow industry accepted guidelines on compliance with a number of test participants for formative testing as well as summative testing. Several elements are crucial in determining sample size: the type of test, the representative users for that test/type of work, type of task, context, and risk factors (patient safety versus cosmetic).

Summative testing with fewer than 15 participants should not be accepted going forward. Summative testing is a benchmark test so more users are required; 15-20 are suggested per user group for which the application is being designed as per NIST IR 7804. If organizations use smaller samples, justification with rationale is needed since this would be outside industry standards and best practice recommendations. This is also consistent with language in the 2011 FDA guidance on validation testing (see Appendix B). " If users with distinctly different characteristics [e.g., use responsibilities, age ranges, skills sets or experience levels] will use the device, validation testing activities should include 15 from each major user group."1

Resources on sample size include Falkner2 and Virzi.3

1. FDA. Draft Guidance for Industry and Food and Drug Administration Staff - Applying Human Factors and Usability Engineering to Optimize Medical Device Design. Washington DC2011.

2. Falkner L. Beyond the five user assumption: Benefits of increaed smple sizes. Behavior Research, Methods and Computers 2003:379-83.

3. Virzi. Refining the test phase of usability evaluation: How many subjects is enough Human Factors 1992;34:457-69.

An area covered during work group meetings that is not specifically addressed in the NPRM is the need for formative testing with a focus on implementation due to the safety issues that arises from customization.

| **Request for Comment on Summative Testing Page # 196** |
| --- |
| **2015 Edition EHR Certification Criterion**We understand that some health IT developers are concerned that the summative testing report may not adequately reflect the design research that has been performed throughout a product’s lifecycle. We request public comment regarding options that we might consider in addition to – or as alternatives to – summative testing. For example, if formative testing reflects a thorough process that has tested and improved the usability of a product, could a standardized report of the formative testing be submitted for one or more of the 17 certification criteria for which summative testing is now required? What would be the requirements for this formative testing report, and how would purchasers evaluate these reports?  |
|  |

When a team attests to applying a nationally recognized user centered design process, they are attesting to carrying out formative activities and attesting to conducting a summative usability (validation that the product has met identified usability goals). To replace the summative usability test with only formative testing is prescribing a process that is not a nationally recognized user centered design process.

Formative activities are required as a part of UCD attestation. Formative evaluation needs to be mentioned and required in some way. Formative activities will likely be covered when a quality management system is implemented. Our work group also heard about risk management processes that, if implemented, would embrace formative testing. Both formative and summative are needed.

Summative testing with summative metrics is the only standard for the usability validation and all of our efforts to ensure effective validation for the safe and usable design will be in vain if this will be voted out.

**INTERNATIONAL STANDARD ISO/IEC 25062 A.22**

A usability test of a completed product to determine how well it meets its usability objectives

The report format assumes sound practice [1, 2] has been followed in the design and execution of the test.

Test procedures which produce measures that summarize usability should be used, i.e. the test is summative in nature.

Some usability evaluation methods, such as formative tests, are intended to identify problems rather than produce measures; the format is not structured to support the results of such testing methods.

**ANSI/AAMI HE 75:2009 Human Factors Engineering-Design of Medical Devices**

9.3.1.1 Formative usability testing

Formative usability testing is performed early, using simulations and early working prototypes; it is intended to explore whether usability objectives are attainable, but does not have strict acceptance criteria.

3.89 summative usability testing: Usability testing performed in the late stages of design.

NOTE—Summative usability tests include verification and validation, and it is a recommended best practice to have formal acceptance criteria (e.g., usability objectives for human performance and satisfaction ratings).

5.6 Identification of use-related hazards

5.6.1 Overview

Extent of testing effort: Usability testing can involve varying degrees of structure, complexity, and realism.

Tests can consist of informal, “quick and dirty” preliminary testing (sometimes called formative evaluations) or more comprehensive

We recommend continued use of Summative Test reports.

| **Retesting and Certification Page #197** |
| --- |
| We believe that ONC-ACB determinations related to the ongoing applicability of the SED certification criterion to certified health IT for the purposes of inherited certified status (§ 170.550(h)), adaptations and other updates would be based on the extent of changes to user interface aspects of one or more capabilities to which UCD had previously been applied. We believe that ONC-ACBs should be notified when applicable changes to user-interface aspects occur. Therefore, we include these types of changes in our proposal to address adaptations and updates under the ONC-ACB Principles of Proper Conduct (§ 170.523). Please see sectionIV.D.6 of this preamble for further discussion of this proposal. |
|  |

This is already addressed through the existing processes. Release notes include a summary of the software changes; including user interface changes. No change is required to the current process to consider user interface changes.

| **§ 170.315(g)(4) Quality management system Page #197** |
| --- |
| **Included in 2015 Edition Base EHR Definition?**No |
| **MU Objective** N/A |
| **2015 Edition EHR Certification Criterion****(**4) Quality management system. (i) For each capability that a technology includes and for which that capability's certification is sought, the use of a Quality Management System (QMS) in the development, testing, implementation, and maintenance of that capability must be identified that is:(A) Compliant with a QMS established by the Federal government or a standards developing organization; or(B) Mapped to one or more QMS established by the Federal government or standards developing organization(s).(ii) If a single QMS was used for applicable capabilities, it would only need to be identified once.(iii) If different QMS were applied to specific capabilities, each QMS applied would need to be identified.  |
| **Preamble FR Citation:**  | **Specific questions in preamble?** *No* |
|  |

There great potential in risk management and quality improvement processes. During presentations to the workgroup, it seemed that the AAMI idea of convening diverse groups to develop a risk management process for EHRs was appealing, so while there's potential, there's a lot of work to be done.

Recommend supporting the proposed language.

| **§ 170.315(g)(5) Accessibility technology compatibility Page #199** |
| --- |
| **Included in 2015 Edition Base EHR Definition?**No |
| **MU Objective** N/A |
| **2015 Edition EHR Certification Criterion**(5) Accessibility technology compatibility. For each capability technology includes that is specified in the certification criteria at § 170.315(a), (b), and (e), the capability must be compatible with at least one accessibility technology that includes text-to-speech functionality. |
| **Preamble FR Citation:**  | **Specific questions in preamble?** *Yes* |
|  |

Some Group 2 members waiting to hear from their Accessibility Experts and stakeholders. Other Group 2 members support the proposed language.

| **§ 170.315(g)(8) Accessibility – centered design Pg #210, 406** |
| --- |
| **Included in 2015 Edition Base EHR Definition?**No |
| **MU Objective** N/A |
| **2015 Edition EHR Certification Criterion**(8) Accessibility-centered design. For each capability that a Health IT Module includes and for which that capability's certification is sought, the use of a health IT accessibility-centered design standard or law in the development, testing, implementation and maintenance of that capability must be identified.(i) If a single accessibility-centered design standard or law was used for applicable capabilities, it would only need to be identified once.(ii) If different accessibility-centered design standards and laws were applied to specific capabilities, each accessibility-centered design standard or law applied would need to be identified. This would include the application of an accessibility-centered design standard or law to some capabilities and none to others.(iii) If no accessibility-centered design standard or law was applied to all applicable capabilities such a response is acceptable to satisfy this certification criterion. |
| **Preamble FR Citation:**  | **Specific questions in preamble?** *Yes* |
|  |

Some Group 2 members waiting to hear from their Accessibility Experts and stakeholders. Other Group 2 members support the proposed language.