Infusion pump standards guide

chi+med

making medical devices safer

EPSRC Programme Grant EP/G059063/1

Disclaimer: This guidance, is provided by UCL and its partners ‘as is’, without any representation or endorsement made and without warranty of any kind whether express or implied, including but not limited to the implied warranties of satisfactory quality, fitness for a particular purpose, non-infringement, compatibility, security and accuracy. UCL and its partners accept no liability in connection with the use of such materials by a third party for any purpose.

Executive Summary

Standards provide for the growth of markets. Across a market as a whole, standardisation offers efficiencies in terms of maintenance, compatibility and elimination of wasteful duplication or unproductive labour [1]. From a consumer perspective, standards can be the basis for certification schemes which communicate an attribution of quality and safety that would otherwise remain hidden (e.g. use of the CE mark). In the UK, standardisation efforts are estimated to contribute £2.5bn per year to the economy and have been shown to facilitate trade [2]. Standards provide one or more of the following: compatibility, minimum quality and a reduction in variety. Taking the production of medical devices as an illustrative example, compatibility allows for the growth of networks of complementary products (e.g. interconnectivity between bags of fluid and a means of delivery), quality reduces risk (e.g. providing a guarantee of sterility) and reduction in variety helps realise economies of scale (e.g. use of a generic equipment type). A recent US report identified “setting standards and guidelines for safety and efficacy” as contributing to one of the five pillars of medical technology innovation, the global medical device industry equating to $350bn USD a year [3]. This document examines the “standards landscape” by taking a series of commonly applied medical device standards and describing the practicalities of application.

For previous versions please contact Chris Vincent +44 (0)20 7679 0694

1. Introduction

Standards provide compatibility, quality and a reduction in variety. There are multiple approaches to standardisation, for example, specification of a product, implementation of management systems (process) or establishment of common values or principles. The guide will outline the use of standards during medical device development and provide examples of the application of standards associated with medical device design (infusion pumps).

1.1 Mechanisms of standardisation

In the European Union, the "placing onto market" of medical devices is governed by a number of European Council directives that are implemented though national law. The directives specify essential regulatory requirements, corresponding to the quality, safety and performance of medical devices. Modular, open, voluntary and harmonised standards support compliance with the regulatory requirements. Within the EU, this exemplifies a class of "new approach" directive. Contrasting "old approach" directives contain a large amount of technical detail, which adds to the challenge associated with approval and revision. For new approach directives, bodies such as CEN and CENELEC prepare consensus standards to support compliance. National Standards Bodies (NSB) are involved in the generation of consensus standards and private, independent, certification authorities or "Notified Bodies" assess conformity. This means that new approach directives need only contain essential requirements. For medical devices, the adoption of harmonised consensus standards provides benefit as a single European standard replaces numerous national standards. Harmonised standards therefore cut the cost of compliance, provide a single point of access to the market and support free trade. As the adoption of standards is voluntary, organisations are free to innovate, although in many cases, incorporation of tried and tested solutions is appropriate. In these cases, product standards provide a basis for quality, consistency, comparability and testing. For consumers, standards communicate an attribution of quality and safety that would otherwise remain hidden (e.g. use of the CE mark).

In affixing a CE mark, a manufacturer demonstrates compliance with the appropriate parts of the Medical Device Directive(s). The regulatory framework provides a risk based classification process, which impacts on the essential requirements that apply and the conformity process used to demonstrate compliance. For example, for comparatively low risk devices, manufacturers will self certify, for higher risk devices there may be a need to generate a dossier of evidence or "technical file" that is independently audited by a Notified Body. Necessary documentation may include descriptions of the product or process used to manufacture the product, results of risk analysis, development, testing or inspection activities. The instructions for use also provide an opportunity for the manufacturer to state their intended concept of use, user and use environment.

One way to show compliance with the essential requirements of the medical device directives is through adoption of harmonised standards. These standards contains a “Z Annex” which map clauses contained within the standard to the essential requirements of a given directive. Within the EU framework, the adoption of harmonised standards is not mandatory. However many manufacturers feel that they need to utilise appropriate standards when demonstrating conformity. EU member states have a designated “Competent Authority” which is responsible for ensuring compliance with the EU directives. In the UK, the Competent Authority is the Secretary of State for Health, who acts through the Medicines and Healthcare products Regulatory Agency (MHRA). Competent Authorities can designate Notified Bodies to take an active role in conformity assessment. The designation occurs via a process and criteria outlined in the Medical Device Directives. The EC has also mandated CEN and CENELEC to prepare standards to support, through generation of harmonised consensus standards. The fact that a standard is harmonised is indicated the prefix EN in the label. Recent, proposed changes to the Medical Device Directives have been summarised by the MHRA: http://tinyurl.com/cjt4v59

1.2 Why standardisation is not always easy, the story of Luer

There are approximately 300 standards that are current, UK-specific and applicable to general medical devices (harmonised) (for infusion pumps see Figure 1). Examples include the use of standard scalpels (BS EN 27740:1992), surgical gloves (the BS EN 455 series) and standardised connector types such as the Luer conical fitting (BS EN 20594-1:1994) (see also the 80369 series). The final case is interesting, because the Luer fitting was originally developed in a proprietary setting by Karl Schneider, for Wülfing Luer, in 1896. It was then made available to the wider industry to promote interoperability. Without standardisation, health services fail to work together in an effective way. For example, in 1988, following the Ramstein air-show disaster, incompatibilities between the connector types used on IV catheters impeded the emergency response (Brown, 2012). The Luer fitting has since become a global standard. Conversely, although the Luer connector has proved successful in allowing interconnection between multiple equipment types, it has also been implicated in several wrong-route administration errors. These are where mistaken connection of the wrong device or substance results in delivery to an unintended part of the body. A study commissioned by CEN showed that when the potential for misconnection was considered across multiple medical connector types (including the Luer), 27% could be fatal (PD CR 13825:2000). There is therefore an inherent complexity in product standardisation, with a balance to be achieved between flexibility and control.

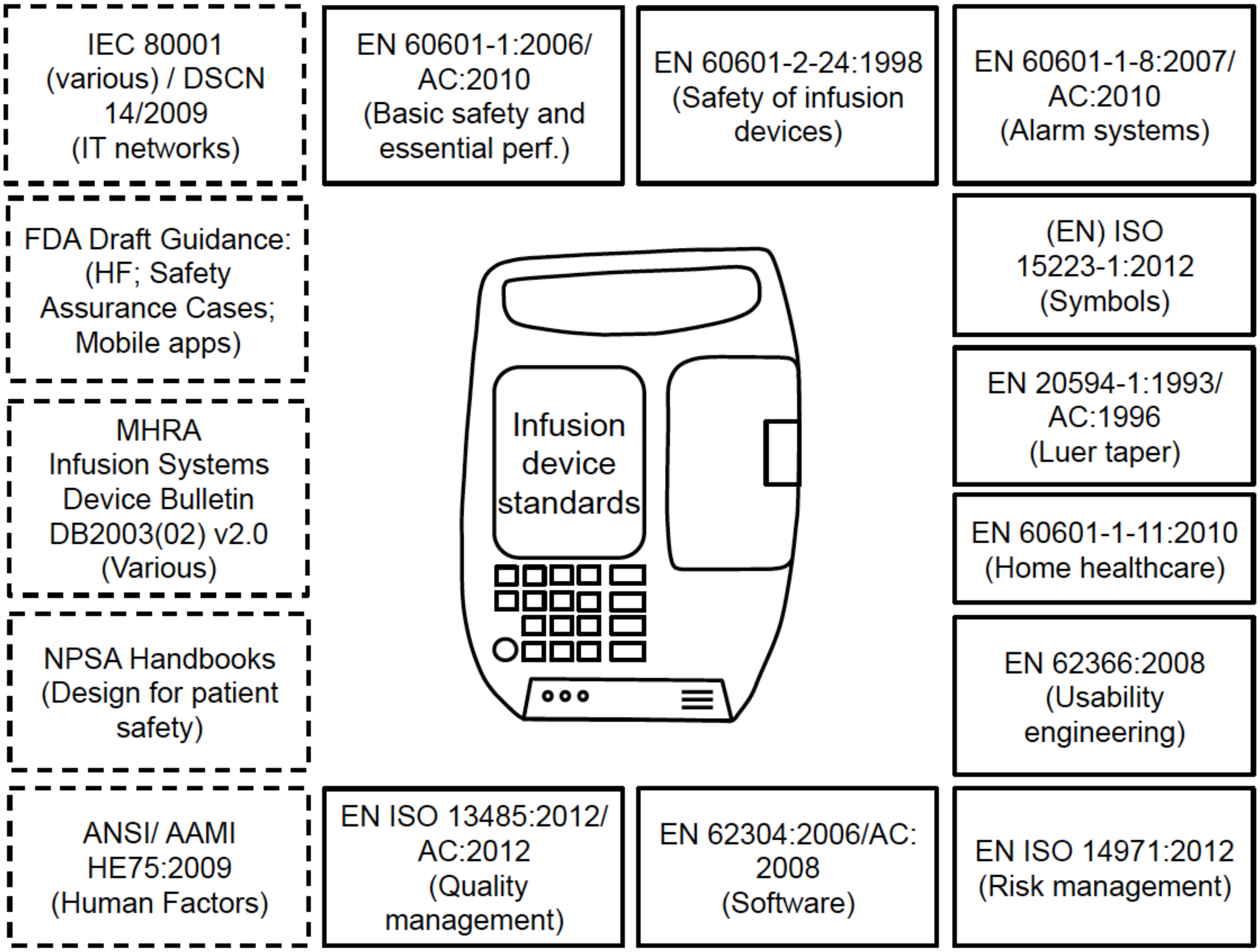
IEC 80001 (various) / DSCN 14/2009 (IT networks)
EN 60601-1:2006/ AC:2010 (Basic safety and essential perf.)
EN 60601-2-24:1998 (Safety of infusion devices)
EN 60601-1-8:2007/ AC:2010 (Alarm systems)
FDA Draft Guidance: (HF; Safety Assurance Cases; Mobile apps)
Infusion device standards
(EN) ISO 15223-1:2012 (Symbols)
EN 20594-1:1993/ AC:1996 (Luer taper)
MHRA Infusion Systems Device Bulletin DB2003(02) v2.0 (Various)
EN 60601-1-11:2010 (Home healthcare)
NPSA Handbooks (Design for patient safety)
EN 62366:2008 (Usability engineering)
ANSI/ AAMI HE75:2009 (Human Factors)
EN ISO 13485:2012/ AC:2012 (Quality management)
EN 62304:2006/AC: 2008 (Software)
EN ISO 14971:2012 (Risk management)

Figure 1: A selection of infusion device standards and guidance (product, process and principles). Dashed boxes are not European (harmonised) standards.

1.3 Advantages of standards

Standards provide for wider social, economic and political objectives. They reduce potential for damage to people, property or the environment. These factors can be hard for a manufacturer to factor into the development process. Standards provide the means to incorporate factors that occur post purchase, prior to an event happening. For example, when packaging is disposed of, the manufacturer is detached from the event. A standard that specifies a property of biodegradability allows the manufacturer to take on responsibility for the cost of the disposal, prior to the event happening. This is helpful, as it allows the sharing of responsibility between public and private entities.

Standards increase in worth, the greater the number of people that adopt them. For example, there is an innate benefit for the group of VHS users as a whole, every time an individual buys a VHS video player and starts consuming VHS standard product. Form the perspective of a consumer, standards can also make the invisible visible (e.g. crash test ratings for a car) and reduce the cost in locating an appropriate product. From the manufactures perspective, standards offer a single route into the market that incorporates the desires of the consumer, without having to conduct market research.

1.4 Disadvantages of standards

For medical devices, inherent in the "new approach" is that the adoption of consensus standards is voluntary. This means that whilst compliance with the directives has a legal mandate, standards only have a quasi-regulatory role. This means that total harmonisation may not achieved (e.g. everyone using the same standard). There is also an argument that standardisation may be to the decrement of SMEs, or suppliers outside of the harmonisation zone, which may not have the means to achieve compliance or the ability to influence the formation of standards. One of the biggest challenges with standards (as with old approach legislation) is keeping them up to date with changing technology, as they can take a significant period of time to draft (CEN reports a three year timeframe). For medical devices, examples of recurrent challenges, frequently cited by manufacturers include:

1) Understanding what is and isn't a medical device (Annex 5.5).
2) Understanding what does and does not fall into the remit of legitimate concerns regarding potential for use error.

2. Examples of product standards that are applicable to infusion pump design

2.1 Symbology (60878:2003)

Common symbology enables a product to be marketed in multiple geographic regions and avoid the need for multiple languages. HE75 [4] specifies that 85% or users should be able to identify the symbol meaning when tested with the intended clinical users in the intended clinical setting. When considered at an international level, studies have shown that for a commonly used set of symbology (IEC 60878:2003), reported comprehensibility varies across symbol type and country. For example, for symbology likely to be used in an ICU, the “bell cancel” symbol (IEC number 5576) was reported comprehendible by 100% of German users, but only 65.4% of Chinese users. The “do not reuse” symbol (IEC number 1051) was reported comprehendible by 32.5% of German users and 46.2% of Chinese users.

Few, if any symbols have a universal meaning. The potential for misunderstanding is described by Cassey [5], in a report where Iraqi peasants ate seed that had been preserved using a mercury based compound. This was stored in bags labelled with a skull and cross bones symbol, the meaning of which was misunderstood by the peasants.

Current UK guidance suggests that users may not be familiar with the meaning of symbology, such as the set specified in IEC 60878:2003. Manufacturers should consider providing an accompanying text label or “improve understanding of symbols in other ways, until the meaning of the symbols are universally recognised by users”.

In hospitals unique symbol sets have arisen based on the requirements of a given context and information exchange requirements (e.g. handover) [6]. In these cases, there may be utility in leveraging previously established sets, during system design, however consideration should be given to estimating the likelihood of misinterpretation, as well as interpretation. This should be specific to the population who are likely to be using the device.

2.2 Alarms (60601-1-8)

For medical devices marketed in the EU, there is a voluntary consensus standard 60601-1-8:2007 which describes requirements, guidance and tests for alarm systems used in medical electrical equipment. The document contains a partial specification of aspects regarding the duration and frequency of tones, colour, duty cycle and brightness of accompanying indicator lights. It also details requirements for test. The 2005 version of the standard went much further in providing recommendations regarding the use of melodic alarms to discriminate between alarms emanating from various sources. Feedback regarding the implementation of the 2005 version of the standard has suggested that it can be hard to establish a suitable set of melodic alarms, which the standard does not detail. It also suggests that without an accompanying component of user testing, there is no way to establish that users would be able to discriminate a chosen set [7-10]. An important lesson learned during the revisions of 60601-1-8 was that if a standard allows variants (which is often important and / or unavoidable) then a method should be provided to determine that the alternatives are safe and effective.

HE75 [4] includes the following points relating to this topic:

- There is a need to provide the means for the user to verify that the alarm system works.
- There may be the need to consider contextual factors relating to the state of the device (for example, is it connected to the patient? – if not, then an alarm may not be appropriate).
- False alarms, alarm fatigue and cry wolf syndrome all relate to potential concerns regarding over alarming devices. These need to be considered in the design.
- Different models of alarm system are required dependent upon the concept of operation. For example, is the device user always going to be near the device, does a single user have responsibility for a single piece of equipment, should the alarm we disabled when the user is aware?
- Does the alarm system form part of a wider distributed system?
- Is there an allowable latency, priority or urgency associated with an alarm condition?
- Would it be appropriate for the alarm to latch?
- Under what conditions should an alarm be triggered?
- Under what conditions should an alarm terminate?
- Are default limits appropriate, are these under the control of the user?
- What happens to limits when a device is reset?
- Should the limits be variable or under the control of the user?
- Should it be possible to disable an alarm?
- Is the alarm likely to interfere with other visual or auditory feedback?
- Is the user likely to understand why the device is alarming?
- There is a need to consider users with cognitive, perceptual or physical impairments.
- In some situations speech based alarms may be appropriate.

Ultimately, medical device developers face difficult decisions when deciding how and when devices should alarm. There is a balance to be established between alarm necessity and frequency. This problem is exasperated given the number and range of devices that may be alarming simultaneously.

3. Examples of process standards applicable to infusion pump design

3.1 Usability engineering (IEC 62366) (With particular regard to use error)

IEC 62366 outlines the application of usability engineering to medical devices and provides for the safety of those who may interact with the device. These interactions are broadly defined and include (but are not limited to) transport, storage, installation, operation, maintenance, repair and disposal. The standard is focused on managing and reducing the hazards associated with the user interface, so the traditional approach of identifying hazards, estimating and evaluating risks, controlling those risks and monitoring the effectiveness of the control applies equally well to aspects of the device user interface as it does to almost any other area. The usability engineering standard links to a well-known risk management standard (ISO 14971) and is explicit in encouraging the linkage between design features and the mitigation of risk.

It is important to note that there is no systematic or formal test method that can predict in advance the likelihood of people making errors with a particular design, although research does offer a tantalising hint that this may be possible in some situations. For example there are reasonably developed models that can predict the chance of numeric entry error or human performance limitations when it come to behaviours such as multitasking. However as far as the standard is concerned, tried and tested methodologies such cognitive task analysis or workload assessment are recommended for the formative stages. The standard specifies the usability engineering process, as defined by multiple phases – namely user research, conceptual design, requirement and criteria development, user interface design, implementation, verification and validation. The standard specifies an iterative process and outlines the fact that activities may occur in parallel and may be revisited on multiple occasions. The standard maps tools and techniques such as contextual inquiry, task analysis and heuristic analysis to the various process components and contains quite a complex decision diagram that informs how to proceed given the results of various phases or interaction with risk management techniques.

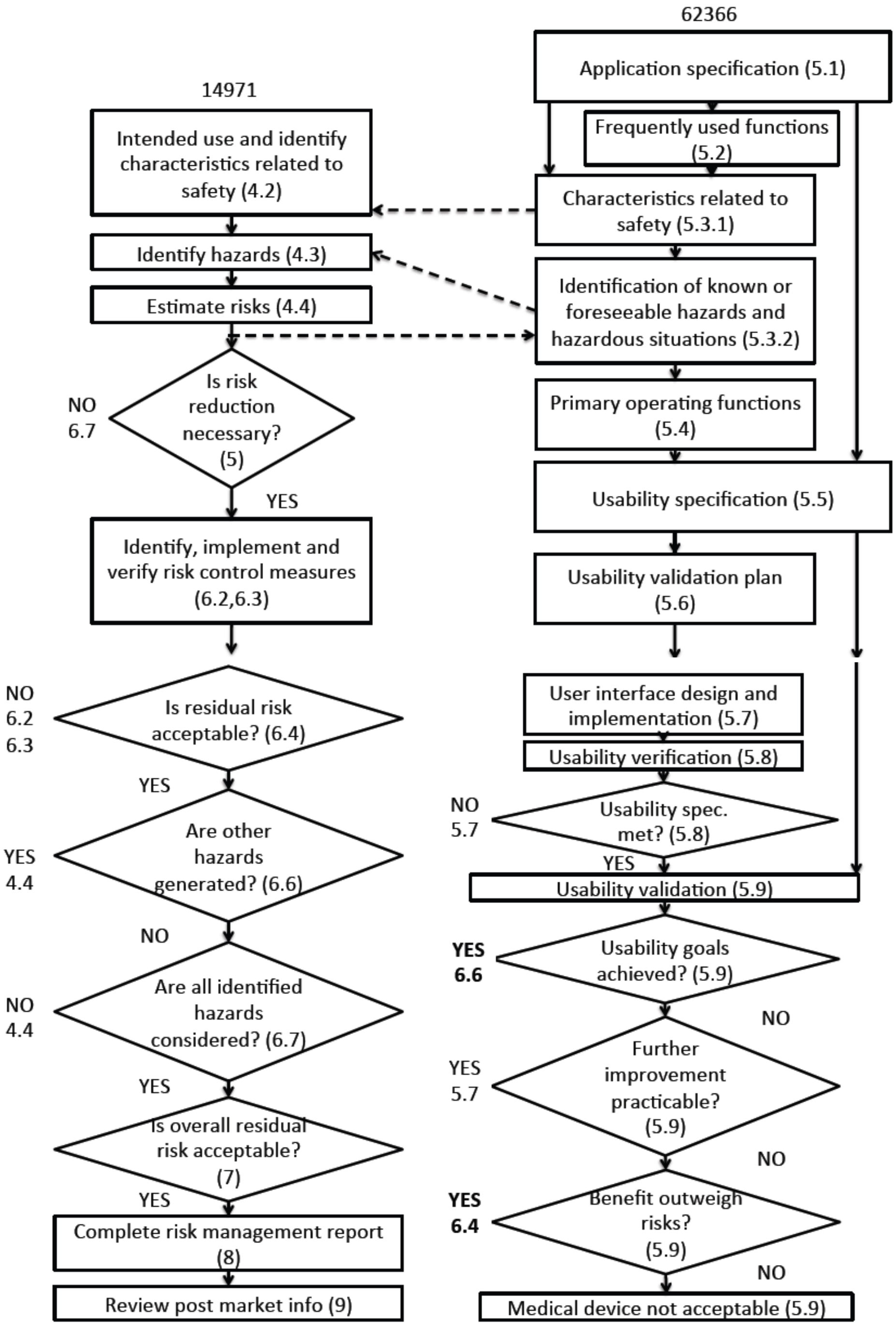
14971
Intended use and identify characteristics related to safety (4.2)
Identify hazards (4.3)
Estimate risks (4.4)
Is risk reduction necessary? (5)
NO 6.7
YES
Identify, implement and verify risk control measures (6.2,6.3)
Is residual risk acceptable? (6.4)
NO 6.2 6.3
YES
Are other hazards generated? (6.6)
YES 4.4
NO
Are all identified hazards considered? (6.7)
NO 4.4
YES
Is overall residual risk acceptable? (7)
YES
Complete risk management report (8)
Review post market info (9)
62366
Application specification (5.1)
Frequently used functions (5.2)
Characteristics related to safety (5.3.1)
Identification of known or foreseeable hazards and hazardous situations (5.3.2)
Primary operating functions (5.4)
Usability specification (5.5)
Usability validation plan (5.6)
User interface design and implementation (5.7)
Usability verification (5.8)
Usability spec. met? (5.8)
NO 5.7
YES
Usability validation (5.9)
Usability goals achieved? (5.9)
YES 6.6
NO
Further improvement practicable? (5.9)
YES 5.7
NO
Benefit outweigh risks? (5.9)
YES 6.4
NO
Medical device not acceptable (5.9)

Figure 2: Overview of 62366 Process

Figure 2 shows an indicative flow through the process. A manufacturer would need to start by outlining the description of the intended application. This would include: A definition of the condition to be addressed by the medical device, the patient population, user profile(s), conditions of use and operating principles. A manufacturer would go on to identify frequently used functions. Examples of “frequently used functions” would include turning on or off the device or for infusion pumps, loading a giving set. Many of these functions will map directly onto the section regarding characteristics that relate to safety and act as an input for the risk analysis process. When defining characteristics that relate to safety, there is a list of questions in the back of the standard that can be used as a prompt, for example, is the medical device used in an environment where distractions are commonplace? The manufacturer would compile a list of hazards and hazardous situations associated with the device. A hazard is defined as a potential source of harm, which includes physical injury or damage to the health of people, property or the environment. Primary operating functions, which include functions relating to safety and frequently used functions would then provide input into the usability specification. The section would list scenarios and provide testable acceptance criteria or usability goals. It could also include indications of menu flows, screen layouts, dialogues or control panels. The usability validation plan facilitates collection of objective evidence that the product meets the intended use. This may be qualitative or quantitative. It is structured using the primary operating functions and references scenarios outlined in previous sections (including worst case scenarios).

During implementation, checks need to be made to ensure the device is being designed as intended. Verification ensures that the product meets design requirements; validation ensures that the product meets user needs (in context). Verification may involve comparing design sketches or interface layouts with the current product to ensure they have been implemented correctly. In terms of validation, manufacturers may choose to test a product in a hospital setting (although there are limitations to doing this), or use simulation techniques. Usability goals may be incorporated as acceptance criteria. During validation, if acceptance criteria are not met then the manufacturer has the option to demonstrate that the benefits outweigh the risk using a risk analysis approach, however a case needs to be made for further improvements not being practicable. There are plenty of techniques that can be used to support risk analysis, for example Fault Tree Analysis (FTA), Failure Mode and Effect Analysis (FMEA), Hazard and Operability Studies (HAZOPs). The annexes of 14971 outline these techniques, however don’t mention some of the techniques that have been applied to specifically cater for user interface design – for example THEA – Technique for Human Error Analysis which includes consideration of human information processing. This will also allow developers to take into account taxonomies of human error such as those that define slips, lapses and mistakes.

A recent case study has outlined the application of 62366 during the development of ventilation systems [11].

62366 is currently due for revision, with a supplementary annex (Annex K) that will address process for legacy devices and changes to legacy devices. It will also address the assessment of usability for equipment of unknown provenance. The standard is due to be rewritten and split into 62366-1 (a shorter normative standard) and 62366-2 (informative / supplementary information). There is an expectation that it will be combined with the draft FDA guidance and published in 2014.

3.1 Human factors engineering (HE75)

HE75 is a horizontal standard, which means that it provides fundamental advice, which may be applied across multiple product types or development contexts. HE75 sets out a series of design principles that can be used optimise design, in conjunction with process-based standards such as 62366. Originally, the different styles of advice were presented together (for example HE48:1993). As part of the FDA recognition of HE75, they signpost additional guidance. This guidance makes it clear how an organisation can satisfy the conduct and reporting of HFE / UE activities. It is called: “Medical Device Use Safety: Incorporating Human Factors into Risk Management”. Although the advice is in line with the content of HE75, it is not identical. The FDA guidance is outlined in the following section and is a reasonably concise overview of the way in which manufacturers can conduct HFE/UE. HE75 provides a more comprehensive range of data, methods and principles that support the development of medical equipment. In terms of the interactive properties of devices, there are some key points from HE75 that are worth considering:

- Unlike approaches that are purely focused on managing and reducing risk (e.g. 14971 and to some extent 62366), HE75 encourages consideration of overall user experience.

- “Human factors are not only about safe and effective task performance, but also about user satisfaction. Designers should try to make medical devices pleasing to use.” [4]

- If product specific advice is not available then there is the option to incorporate evidence from a wider range of sources: “Readers seeking basic software–user interface design guidance are referred to the resource listing at the end of this section (e.g., ANSI/HFES 200:2008), as well as the Association for Computing Machinery’s (ACM’s) Special Interest Group on Computer-Human Interaction website (www.sigchi.org)[1], which provides updated references to design standards and guides.” [4]

[1] It is also worth referring to the US Human Factors and Ergonomics Society content, for example freely available material from the annual healthcare symposium: http://tinyurl.com/lrydohl

- The advice in HE75 Annex A regarding the justification of sample size is not identical to the advice presented in the FDA guidance document [12]. The advice in the Annex of the FDA guidance appears to be more sympathetic to the fact that setting an appropriate sample size may be influenced by real world practicalities and needs to be considered on a case by case basis.

- HE75 also contains UE/HFE process information (figure 9.2). This is similar but not identical to equivalent process specified in 62366 (figure A.1). The content in HE75 is clearer about the initial research stages and places a greater emphasis on contextual enquiry. The content in 62366 is clearer about links to the risk analysis process, but does not indicate what inputs inform the application specification.

- HE75 is explicit in stating a link to 62366, particularly with regard to defining, user population, user profiles, intended use (e.g. task analysis), scenarios of use, and descriptions of the likely use environment. HE75 outlines how to do this in Section 5 and Section 9, which is similar to the requirements and illustrative examples in both the FDA guidance and 62366.

- There has been a historic confusion over nomenclature which applies to multiple terms contained within the standard. For example in some cases, user has referred to the owner of a device, where as operator has referred to the person who actually uses it. More recently (since HE74 and the 3rd edition of 60601), the definition of user has been broadened to include both of these aspects, however terms like verification, validation, summative and formative continue to cause confusion.

3.2 FDA draft guidance: Applying Human Factors and Usability Engineering to Optimise Medical Device Design: http://tinyurl.com/qdl6mkw

The FDA draft guidance is not a standard, but is due to be combined with 62366. In the US, the regulatory basis for HFE is contained within Quality System Regulation, 21 CFR Part 820 Section 30, Design Controls. The document sets out the content of a HFE/UE report and describes some of the techniques that can be used to support. The scope of the document suggests that HFE/UE activities are necessary if:

- A device is being modified due to problems associated with use (for example as a result of corrective or preventative action – CAPA).

- Analysis reveals that there is a "moderate to high" risk of use error. Section D.3 in 14971 may be used to help to inform this judgment, for example, if use error is unlikely or the severity associated with use error negligible then HFE/UE activities may not be necessary.

If HFE/UE activities are necessary then as with 62366, an overarching process of risk/hazard analysis applies. This is consistent with the broad process specified in 14971, although the draft FDA guidance places a greater emphasis on methods to analyse and evaluate the use of a device, as well as being more specific about the types of activity that might inform these considerations. These overlap with similar suggestions illustrated in HE75 (Figure 9.2) and 62366 (Figure A.1). Presentations to industry have containing the following points:

- For infusion pumps, extra validation testing may be required involving simulated use and clinical evaluations. There may be a need for an assurance case.
    - http://tinyurl.com/c3kpx78
- Testing the speed of task completion may not be appropriate, compared with success, failure, confusion or use error.
- Training needs to be representative, and so may require a period of decay.
- There is a need to test with US citizens and/or to not test with the employees of the organization(s) involved in marketing the product.
    - http://tinyurl.com/bxoo7w9
    - http://tinyurl.com/aup5pwc
- Think aloud is not an acceptable method for collecting data for HF validation testing.
    - http://tinyurl.com/aup5pwc
- For validation testing, the test environment needs to be representative.
- Tasks used for validation testing should include those that are essential and those that are safety-critical. Anything to do with an alarm, or warning / caution in the instructions is safety critical.
    - http://tinyurl.com/aup5pwc
- There is a need to document use error, irrespective of whether or not a participant succeeds in a task. The description of the use error needs to be detailed.
    - http://tinyurl.com/bxoo7w9
- If use error is documented, there may be a need to modify the design.
- There is a need to consider worst case users.
- If a device is to be used by multiple user groups there is a need to test using multiple user groups.

Since the FDA guidance was produced, consultation has resulted in 600 comments.

3.3 Risk Analysis: 14971

14971 provides a framework for the management of risk, when considering across the perspective of a number of stakeholders. 14971 was prepared by a technical committee focusing on quality management and maps onto the EU medical device directives, as specified in the three Z Annexes. The 2012 version is specific in detailing which parts of the standard map to the various essential requirements held within the directives.

A generic approach to risk analysis is as follows:

- Identify characteristics that relate to safety
- Identify hazards and hazardous situations
- Estimate risk for hazardous situation
- Evaluate the need for risk reduction
- Implement risk control measures
- Evaluate residual risk
- Produce a risk management report
- Review post market data

14971 outlines use of a number of techniques that can contribute to this process including:

Prospective Hazard Analysis (PHA)

These methods have been described in detail in the Prospective Hazard Analysis Toolkit, available from the Cambridge Engineering Design Centre:

http://tinyurl.com/pz7765o

Fault Tree Analysis (FTA)

Tree like diagram used to graphically represent system failures and causes. These are often used in nuclear power / chemical processing industries (see BS EN 61025:2007 or NUREG CR 2300). The process is as follows:

- Define failure event;
- Determine cause;
- Take first cause - can you decompose, are there any more causal events?
- Repeat.
- It is also worth considering the use of Event Tree Analysis ETA (BS EN 62502:2011).

Failure Mode and Effects Analysis (FMEA) (BS EN 60812:2006)

Variants include Healthcare Failure Mode and Effects Analysis, Failure Mode and Effects and Criticality Analysis (HFMEA / FMECA). FMEA is a frequently used inductive technique for the identification of problems that might occur within a system of interest. An example healthcare application is detailed in [13].

For FMEA and associated variants, a brief summary of the process as follows:

- Create a plan: purpose; scope; relationship to the project; who is involved; schedule.
- Create a description of system structure (redundancy, connections, IO, modes of operation).
- Define the system boundary.
- Decide the highest and lowest level of analysis (e.g. stop at resistor level).
- Collect operational information (rules, regs, procedures).
- Specify assumptions / environment.
- List failure modes. Failure modes could include information like failure during operation; failure to operate at a given time; failure to cease operation; premature operation.
- For each failure mode, identify most likely cause(s).
- Detail the effect of the failure:
- Consider detection methods.
- Assign a severity classification (Catastrophic, critical, marginal, insignificant).
- Estimate frequency (1 in 1’000’000).
- Come up with a control and then re-evaluate.

HAZard and OPerability Studies (HAZOP) (IEC 61882:2001)

Hazard and Operability Studies (used in the chemical process industries) use (for example) Piping and Instrumentation Diagrams to help structure the risk assessment process. The process is as follows:

- Assemble HAZOP team.
- Provide structure of diagram to support the analysis.
- Select Guidewords (Leakage, Less Than More Than, Mis-Ordered…)
- For each task step or component: Take the step, transfer or component; apply first guideword; discuss effect of guideword; note any credible errors; describe errors; describe consequences; determine causes; determine recovery; determine error remedy.

Hazard Analysis and Critical Control Points HACCP

Also used in process industries. The FDA provides detailed guidance on this topic (for other domains, e.g. outside of the remit of medical devices).

http://tinyurl.com/pamahhn

Practical Considerations During Risk Analysis

During the application of any risk analysis technique there are several practical considerations, for example:

External Boundary Setting

- What questions do you ask prior to setting the boundary?
- How are spatial and / or conditional boundaries set?
- How do you determine if components or conditions have an affect on one another?

Resolution of analysis

- How do you determine an appropriate level of resolution?
- How are components defined?
- Are lists of components available from historical systems or design drawings / prototypes?
- When can you group or coalesce data for efficiency in handling and evaluation?

Content and type of the diagram or structure used to conduct the analysis

- What are the readymade descriptions that are available (e.g. pathways)?
- What techniques are used and why?
- How are the different types of diagram categorised?
- What taxonomies or frameworks are in existence to support diagraming
- How can you describe the social aspects of systems (oppose to the purely technical aspects)
- How many diagrams are necessary and what type of diagram?
- How are iterations managed?
- How are hierarchies managed?

Other

- How best to get people motivated?
- How can you prioritise effort given limited resource?
- How to avoid tedium?
- How much time goes into diagraming and how much time goes into analysis?
- What trade-offs and how to select sweet spot?
- How are diagrams linked with the wider risk assessment process?
- How are parts of the diagram selected for analysis?

3.4 Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management

The document was issued as guidance in 2000, and overlaps with the FDA draft guidance, 62366 and 14971. The document outlines the use of empirical and analytical approaches in combined HFE and Risk Management process. Approaches include:

Analytical: FMEAs, FTAs, HAZOPs, task analysis, heuristic analysis and expert review.

Empirical: User testing, walk-throughs, assessment of perceived workload.

The FDA recognition statement associated with HE75, lists the document as additional guidance.

3.5 Safety Cases / Assurance Cases

The safety case or assurance case provides evidence that supports the claim that the system is acceptably safe to operate in a given environment for a given amount of time. This approach to regulation became necessary when a more prescriptive approach resulted in several high profile accidents for installations that were theoretically “safe”. For example the Cullen report [14], concluded that safety assurance activities in the offshore oil industry were:

- Too superficial;
- Too restrictive or poorly scoped;
- Too generic;
- Overly mechanistic;
- Demonstrated insufficient appreciation of human factors;
- Were carried out by managers who lack key competences;

- Were applied by managers who lack understanding;
- Failed to consider interactions between people, components and systems.

Assurance cases support demonstration of regulatory requirements across multiple industries, for example the nuclear, aviation and defence sectors in the UK. An assurance / safety case therefore lays out an argument; and supporting evidence; to show that safety claims are valid. There are multiple ways in which compliance can be demonstrated, for example, through standards, testing or formal proofs. A common approach is to base an argument upon hazard analysis. The FDA is requesting that manufacturers of infusion pumps produce assurance cases. It is not unusual for assurance cases to be specified using a graphical notation, although it is also possible to present them using a textual or table notation. For graphical presentation, there are two commonly applied variants, Goal Structured Notation and Claims Arguments and Evidence.

When compiling an argument relating to safety or usability, questions emerge regarding the nature of evidence that should be included. Evidence can come from many sources and include (for example), adherence to standards, results of testing, the use of formal methods or formal proofs, use of tried and tested solutions, the fact that equipment is proven in use or the fact that design corrections have been made to mitigate a known problem.

The question emerges, what is the evidence that should be included. For example, under the heading of relevance; Direct evidence will often be collected from the system in question. Indirect evidence will be the fact that those using a system will be relatively skilled. Coverage relates to the extent to which evidence shows that a requirement is met across the domain in question. There may also be a demonstration of the fact that multiple lines of evidence taken as a whole provide coverage. Trust is the perceived ability to rely on the character, ability, strength or truth of someone or something. Conceptual independence would relate to the fact that different techniques had been used to collect evidence, mechanistic independence would relate to the fact that perhaps different teams had applied the same technique. Corroboration of evidence comes from the legal profession, where in some cases, in order for evidence to count it needs to be backed up by another source.

An assurance case is therefore a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given environment. They are encouraged because they: Move away from the check-list approach; They are appropriate for systems are diverse, distributed and dynamic; They provide top-level argument to ease the review process. They can provide for automatic checks; They expose incorrect assumptions / flawed logic; They can also be used to show to impact of changes in evidence. For an example of related safety approaches used in another domains see: http://tinyurl.com/kgxyf9v

4. Frequently Asked Questions

4.1 Mobile Apps

The FDA has released the following documents outlining the approach to regulating mobile medical applications:

http://tinyurl.com/orejqll

4.2 Comparison of EU and US Systems

DRAFT FDA Guidance	HE 75	62366
	Research (Customer Requirements, Draft Usability Objectives)	
	Contextual Inquiry (Task Flows, User Profiles, Use Environment)	
Identification of Known Problems (Comparative Review, CAPA), ID, Eval and Understand Use Related Hazards (Contextual Inquiry, Interviews, Focus Groups, Function and Task Analysis, Heuristic Analysis, Expert Review) Mitigation and Control of Hazards	Use Error Risk Analysis (FMEA / Hazard Lists)	Characteristics that Relate to Safety and ID of Known / Foreseeable Hazards and Hazardous Situations (User Research, Contextual Inquiry, Conceptual Model, Comparative Analysis, Task Analysis, Cognitive Task Analysis, Workload Assessment, Interviews)
		Primary Operating Functions (Functional Analysis)
	Use Scenario (Product Requirements, Description and Usability Objectives)	Usability Specification (Detailed Specification, Use Scenario, Usability Goals)

	Usability Objectives (Specifications, Product Requirements, Validated Usability Objectives)	Acceptance Criteria, Production Unit Validation, Design Specifications
Formative Eval. (Cognitive Walk-Through, Simulated Use Testing)	Iterative Design (Simulations, Prototyping)	Prototyping, Participatory Design, Style Guide
Design Verification Testing Human Factors Validation Testing (Simulated Use, Task and Use Scenarios, Testing with Participants, Clinical Validation Testing).	Usability Evaluation / Testing (Formative Usability Testing Protocols and Reports, Summative Usability Testing Protocols and Reports, Expert Reviews, Cognitive Walkthroughs, Verification and Validation Reports)	Expert Review, Heuristic Analysis, Design Audits, Cognitive Walkthroughs, Usability Testing
	Post-Implementation Analysis, Customer Complaints, Customer Surveys, Medical Device Reports, CAPA, Product Actions	

It may be worth noting the following:

- Depending on the scheme, user research (e.g. contextual enquiry) could inform the understanding of user and usage (termed the application in 62366) or form the input the risk / hazard analysis (e.g. characteristics that relate to safety). Perhaps this depends on what an individual wants to get out of user research in terms of informing conceptual development, identifying potential concerns (with or without reference to the concept in question).

- Across schemes, techniques or material with the same name (for example use scenarios), may be applied at different stages in the process and applied in different ways.

- Similar information (e.g. CAPA) may be applied prospectively or retrospectively.

- Although the scheme from HE75 is the only one that details post implementation analysis, the process is implicit across EU and US regulatory environments.

4.3 Numbers of participants and usability engineering

In Appendix B of the Food and Drug Administration (FDA) Human Factors Draft Guidance there is consideration of the number of participants required for validation testing. The ANSI/AAMI HE75:2009 definition of validation testing states that users are sampled to ensure:

“user requirements in the form of usability objectives are met”

the draft guidance references two studies detailing how theoretical models can be used to predict a trade-off between the number of participants involved in user testing and the number of usability problems discovered [15, 16]. As validation testing is often applied using a pass/fail criteria, the studies referenced may be more applicable to formative testing where there is a need to demonstrate that a sufficient coverage of usability problems have been “discovered” during prototyping phases.

For example:

“Formative usability tests require only five to eight subjects per homogenous user group. Many HFE experts recommend this sample size because only a few subjects are needed to uncover major usability issues.” [16] [4]

The draft guidance also details an empirical study which argues the benefits of a comparatively larger (than five) sample size. The study establishes limits in the proportion of problems detected for varying sample sizes, with 90% of issues found by fifteen subjects (worse case scenario). This number may also apply to later stage validation testing which:

“requires larger sample sizes so that statistical tests can be performed” [4]

and

“for simple devices, each formative test has five subjects and the final summative test has 15” [4]

For validation testing, a statistical approach would incorporate the number of participants in the assessment of significance and so would overcome concerns about sample size (the important fact being that significance has been reached). Consideration needs to also be given to the probability of falsely accepting or rejecting the null hypothesis and statistical power.

Regardless of the type of usability test conducted (formative v summative), there remain considerations and limitations relating to any theoretical analysis of sufficiency in testing. The

FDA draft guidance is explicit in detailing these limitations. For example, the assumption of independence between usability problems would mean that discovering one problem does not influence the chance of finding another problem. This is usually not the case. In addition, case studies have shown that usability problems may not be discovered during user testing and the reduction of residual risk may not relate to the number of participants involved in user testing. Some problems will have a higher level of severity than others and despite early reports to the contrary may not be discovered prior to less severe problems [16].

1) Usability problems may not be discovered during user testing

For similar methods applied in a different domain (Air Traffic Control), (formal) modelling approaches were found to be beneficial in discovering Human Machine Interaction (HMI) problems. For example, in the accidental message deletion scenario detailed by Simon Buckingham Shum and colleagues [17], the HMI problem was contingent on near simultaneous receipt of multiple messages. In this case system modelling was able to identify the issue, which could have easily been missed during user tests. In any case, the number of participants included in user tests would not have been as critical as the range of conditions in which the system was placed during testing.

2) Reduction of residual risk may not relate to the number of participants involved in user testing

In many cases, additional analysis may be required to understand the potential severity of the problems encountered:

“if, for example only novice users were tested, a large number of usability problems may have been revealed, but the test would not show which are the most severe” [18]

Research has shown that novices and experts vary in their estimates of severity and has highlighted the potential for less severe problems to mask the presence of comparatively severe problems. The guidance is explicit in this respect:

“individual likelihoods of encountering a problem with a user interface vary considerably, depending on the user’s personal capabilities, knowledge and experience levels, nature of interaction with the device, frequency of task performance, attributes of the use environment and use conditions.” [12]

Recommending a sample size for the purpose of user testing is possibly misleading when the objectives of the practitioner is to reduce the risk associated with use. This is because there is no guarantee that a set number of participants will provide the type of information that allows the risk to be mitigated to an acceptable level (although satisfaction of usability objectives and the involvement of user testing is likely to be beneficial). During formative evaluation small numbers of insightful participant could outweigh the benefit of large numbers of poorly informed or ineffective participants, for example if they are overly focussed on a given feature or unaware of operational constraints (careful design of the user test may help mitigate this). Additional research is required to clarify these issues.

4.4 Potential for bias

Concerns have arisen regarding the potential for bias to occur during risk analysis / usability engineering activities. The following sources of bias might apply:

Conformation Bias

"Confirmation bias is a tendency for people to favour information that confirms their preconceptions or hypotheses regardless of whether the information is true. People will focus on and interpret evidence in a way that confirms the goal they have set for themselves." [19]

Hindsight Bias

"Investigations that are anchored to outcome knowledge run the risk of not capturing the complexities and uncertainties facing sharp end personnel and why their actions made sense at the time. Important lessons go unlearned if the exercise is simply to back track someone else's decision landmarks." [20]

Limited Scope – Out of Sight Out of Mind

"Fischoff, Slavin, and Lichtenstein conducted an experiment in which information was left out of fault trees. Both novices and experts failed to use the omitted information in their arguments, even though the experts could be expected to be aware of this information…. …being provided with an incomplete problem representation (argument) can actually lead to worse performance than having no representation at all." [19, 21]

There may also be biases associated with representativeness, availability, anchoring and adjustment, overconfidence, illusions of control and conformation, affective forecasting, causality errors, fixation, framing effects, memory errors, miserly information processing, perception errors, probabilistic reasoning errors, inertia, resistance to self-criticism and unrealistic optimism.

Fischoff talked about debiasing techniques and outlined general methods to:

1) warn those involved about biases in general
2) identify particular biases in play
3) provide feedback and explain the implications
4) extend training.

There were some more detailed strategies outlined in a 1982 chapter which broke down the potential for bias to arise and detailed strategies. These included (for example), asking how much people were prepared to gamble on their decision / assessment.

That said, it may be that the best thing would be for those involved to identify potential biases and then come up with interventions themselves. People use development scenarios to do this. For example, by giving people a scenario outlining a development context, define threats, associated biases and then agree on a mitigation.

Other possible interventions include:

Monitoring the potential for bias and then avoiding the use of those particularly prone to it.

Give reviewers a criteria or decision-making instrument to help identify concerns.

Put in firewalls / controls to stop people from being exposed to biasing information (exposure control) - goes against research calling for a need to bolster communication within development organisations [22].

Get people to admit to a bias and then correct for it (apparently people do not do this very well).

Include sceptics or individuals briefed to propose the seemingly absurd, in order to mitigate potential for fixation or overconfidence.

Include naïve individuals.

Include independent individuals.

Clarify the objective of the review task up front, and make clear definitions / terminology / nomenclature.

Test people using fictitious / materials containing a set of known flaws.

5. Annex

5.5 Definition of a medical device and software as a medical device (EC)

Definitions (1) as amended by 2007/47/EC (extract from MHRA slides)

'medical device' means any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of:

— diagnosis, prevention, monitoring, treatment or alleviation of disease,
— diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap,
— investigation, replacement or modification of the anatomy or of a physiological process,
— control of conception,

and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means.

Definitions (2) (extract from MHRA slides)

It follows that software that does not have such a medical purpose, but is placed on the market for a more general purpose will therefore not be a Medical Device even if a user decides to use it for a medical purpose.

Other examples of software that would not meet the definition of a Medical Device include:

- Software only intended for archiving/ retrieving patient records/images without intending to change or interpret them.
- Electronic prescription software that only replaces conventional paper based prescriptions and sends them out to a pharmacy.
- Patient administration software that only deals with appointments, admissions, referrals and billing/invoicing.

Definitions (3) (extract from MHRA slides)

Software that qualifies under the definition of a Medical Device if placed on the market for such a purpose would include:

- Software that carries out further calculations or interpretations of captured patient data for a therapeutic purpose, e.g. radiation treatment planning, medication dosage calculations.

- Software that carries out further calculations, enhancements or interpretations of captured data for a diagnostic purpose, e.g. tele-health and remote diagnostics, mass screening and risk assessment tools, helpline/telephone services algorithms.

5.6 Abnormal Use and Reasonably Foreseeable Misuse according to GHTF/SG2/N54R8:2006

Abnormal use:

Act or omission of an act by the operator or user of a medical device as a result of conduct that is beyond and reasonable means of risk control by the manufacturer.

Note foreseeable misuse that is warned against in the instructions for use is considered abnormal use if all other reasonable means of risk control have been exhausted (60601-1-6:2004) (from GHTF/SG2/N54R8:2006)

Use Error according to GHTF/SG2/N54R8:2006

Use error: Act, or omission of an act, that has a different result to that intended by the manufacturer or expected by the operator. Use error includes slips, lapses, mistakes and reasonably foreseeable misuse.

Reasonably foreseeable misuse: - use of a product (throughout its life cycle), a process or service in a way not intended by the supplier, but which may result from readily predictable human behaviour.

Examples of abnormal use according to GHTF/SG2/N54R8:2006

- Use of a medical device in installation prior to completing all initial performance checks as specified by the manufacturer.
- Failure to conduct device checks prior to each use as defined by the manufacturer.
- Continued use of a medical device beyond the manufacturer defined planned maintenance interval as a result of operator's or user's failure to arrange for maintenance.
- Contrary to the instructions for use, the device was not sterilised prior to implantation.
- Pacemaker showed no output after use of electro cautery device on the patient despite appropriate warnings.
- Product analysis showed that the device was working in accordance to specifications, further investigation revealed that the operator was inadequately trained due to failure to obtain proper training.
- During placement of a pacemaker lead, an inexperienced physician or other nonqualified individual perforates the heart.
- The labelling for a centrifugal pump clearly indicates that it is intended for use in bypass operations of less than 6 hours in duration. After considering the pump options, a clinician

decides that the pump will be used in paediatric extra-corporeal membrane oxygenation (ECMO) procedures, most of which may last several days. A pump fails due to fatigue cracking and patient bled to death.

- Safety interlock on a medical laser removed by operator or user.
- Filter removed and intentionally not replaced resulting in particulate contamination and subsequent device failure.
- Tanks delivered to a health care facility are supposed to contain oxygen but have nitrogen in them with nitrogen fittings. The maintenance person at the health care facility is instructed to make them fit the oxygen receptacles. Nitrogen is delivered by mistake resulting in several serious injuries.
- Use of an automated analyser regardless of the warnings on the screen that calibration is to be verified.
- Pacemaker patient placed into MRI system with the knowledge of the physician.
- Ventilator alarm is disabled, preventing detection of risk condition.
- Patient's relative intentionally altered infusion pump to deliver a lethal overdose of the infusing drug to the patient.
- Home care worker uses bed rails and mattress to suffocate patient.

Use Error according to GHTF/SG2/N54R8:2006

- Operator presses the wrong button.
- Operator misinterprets the icon and selects the wrong function.
- Operator enters incorrect sequence and fails to initiate infusion.
- Operator fails to detect a dangerous increase in heart rate because the alarm limit is set too high and operator is over-reliant on alarm system.
- Operator cracks catheter connector when tightening.
- Centrifugal pump is made from material that is known to be incompatible with alcohol according to the labelling, marking, and product warnings provided with the pump. Some pumps are found to have cracked due to inadvertent cleaning with alcohol.
- Unintentional use of pipette out of calibration range.
- Analyser placed in direct sunlight causing higher reaction temperature than specified.
- MRI system and suite have large orange warning labels concerning bringing metal near the magnet. Technician brings an oxygen tank into presence of magnet and it moves swiftly across the room into the magnet.

5.7 References

1. DTI, *The empirical economics of standards*. 2005, London: Dept. of Trade and Industry.
2. Swann, P., P. Temple, and M. Shurmer, *Standards and trade performance: The UK experience.* Economic Journal, 1996. **106**(438): p. 1297-1313.
3. PWC, *Medical Technology Innovation Scorecard: The race for global leadership*, 2011: NY.
4. AAMI, *HE75: Human factors engineering - Design of medical devices* 2009, AAMI: Arlington, VA.
5. Cassey, S., *Set Phasers on Stun: And Other True Tales of Design, Technology, and Human Error*. 1998, Santa Barbara, CA: Aegean.
6. Galliers, J., et al., *Safe use of symbols in handover documentation for medical teams.* Behaviour & Information Technology, 2011. **30**(4): p. 499-506.
7. Williams, S. and P.C.W. Beatty, *Measuring the performance of audible alarms for anaesthesia.* Physiological Measurement, 2005. **26**(4): p. 571-581.
8. Sanderson, P.M., A. Wee, and P. Lacherez, *Learnability and discriminability of melodic medical equipment alarms.* Anaesthesia, 2006. **61**(2): p. 142-147.
9. Lacherez, P., E.L. Seah, and P. Sanderson, *Overlapping melodic alarms are almost indiscriminable.* Human Factors, 2007. **49**(4): p. 637-645.
10. Wee, A.N. and P.M. Sanderson, *Are melodic medical equipment alarms easily learned?* Anesthesia and Analgesia, 2008. **106**(2): p. 501-508.
11. van der Peijl, J., et al., *Design for risk control: the role of usability engineering in the management of use-related risks.* Journal of Biomedical Informatics, 2012. **45**(4): p. 795-812.
12. FDA. *Draft Guidance for Industry and Food and Drug Administration Staff - Applying Human Factors and Usability Engineering to Optimize Medical Device Design*. 2011 22/6/11 [cited 2011 11/8/11].
13. Linkin, D.R., et al., *Applicability of Healthcare Failure Mode and Effects Analysis to healthcare epidemiology: Evaluation of the sterilization and use of surgical instruments.* Clinical Infectious Diseases, 2005. **41**(7): p. 1014-1019.
14. Cullen, W.D.L., *The public inquiry into the Piper Alpha disaster*. 1990, London: HMSO.
15. Nielsen, J. and T.K. Landauer, *A Mathematical-Model of the Finding of Usability Problems.* Human Factors in Computing Systems, 1993: p. 206-213.
16. Virzi, R.A., *Refining the Test Phase of Usability Evaluation - How Many Subjects Is Enough.* Human Factors, 1992. **34**(4): p. 457-468.
17. Buckingham Shum, S., et al. *People and Computers XI: Proceedings of HCI'96*. in *HCI'96*. 1996. Springer.
18. Faulkner, L., *Beyond the five-user assumption: Benefits of increased sample sizes in usability testing.* Behavior Research Methods Instruments & Computers, 2003. **35**(3): p. 379-383.
19. Leveson. *White Paper on the Use of Safety Cases in Certification and Regulation*. 2012; Available from: http://sunnyday.mit.edu/SafetyCases.pdf.

20. Henriksen, K. and H. Kaplan, *Hindsight bias, outcome knowledge and adaptive learning.* Quality and Safety in Health Care, 2003. **12**(suppl 2): p. ii46-ii50.

21. Fischhoff, B., P. Slovic, and S. Lichtenstein, *Fault Trees - Sensitivity of Estimated Failure Probabilities to Problem Representation.* Journal of Experimental Psychology-Human Perception and Performance, 1978. **4**(2): p. 330-344.

22. Vincent, C.J., Y. Li, and A. Blandford, *Integration of human factors and ergonomics during medical device design and development: It's all about communication.* Applied Ergonomics, 2014. **45**(3): p. 413-419.