Human Factors Standards and the Hard Human Factor Problems: Observations on Medical Usability Standards

Lorenzo Strigini[a] and Marwa Gadala[b]
Centre for Software Reliability, City, University of London, Northampton Square, London, U.K.

Abstract: With increasing variety and sophistication of computer-based medical devices, and more diverse users and use environments, usability is essential, especially to ensure safety. Usability standards and guidelines play an important role. We reviewed several, focusing on the IEC 62366 and 60601 sets. It is plausible that these standards have reduced risks for patients, but we raise concerns regarding: (1) complex design trade-offs that are not addressed, (2) a focus on user interface design (e.g., making alarms audible) to the detriment of other human factors (e.g., ensuring users actually act upon alarms they hear), and (3) some definitions and scope restrictions that may create "blind spots". We highlight potential related risks, e.g. that clear directives on "easier to understand" risks, though useful, may preclude mitigating other, more "difficult" ones; but ask to what extent these negative effects can be avoided by standard writers, given objective constraints. Our critique is motivated by current research and incident reports, and considers standards from other domains and countries. It is meant to highlight problems, relevant to designers, standards committees, and human factors researchers, and to trigger discussion about the potential and limits of standards.

1 INTRODUCTION

Glucose meters, infusion pumps, and radiation therapy systems are a few of the many computer-based medical devices becoming increasingly essential in medical practice. These devices are evolving from simple, one-function designs to sophisticated, multi-function abilities; their range of users is expanding to less skilled users (including patients); and smaller, more portable devices are introducing a variety of new use environments.

In the Aggregated Quality Assurance for Systems (AQUAS) project, which addresses engineering challenges arising from the inter-dependence between system safety, security and performance, one use case concerns extensions to a blood pressure and neuromuscular transmission monitoring device to provide closed-loop control of these physiological parameters via an infusion pump. We were tasked to explore some human factor issues in the switch from human to automated control. Of specific interest were "*exceptions*" – situations involving extra user interventions: for instance, scenarios in which a device fails to perform as specified, or reverts to a fallback mode of operation in response to detecting failures, and/or an alarm requires intervention by a clinician, and/or a clinician overrides a device (rightly: e.g. due to a failure, or wrongly: e.g., a user that mistrusts a device overrides some of its correct decisions). This focus was linked to our own previous research, about effects of warnings and so-called "automation bias" (Alberdi, 2009; Povyakalo, 2013); many of our examples will be in these areas, but some observations are of more general relevance.

Our example in the abstract, that for an alarm to be effective the designer has a responsibility not only to ensure that it is audible, but that the overall device design does not make it likely to be ignored, is but one of a class of problems arising from the complex interactions between the user, the device and the use environment. These issues are known, to extents that vary from references in the popular press to e.g. "cry-wolf" events (users failing to intervene when they should, because a high rate of false alarms "trained" them to ignore alarms), to scientific literature about "automation bias" (certain human errors becoming more likely, or new errors being created, by use of automated devices), "situation awareness",

[a] https://orcid.org/0000-0002-4246-2866

[b] https://orcid.org/0000-0002-9148-6522

"complacency" (user inaction when they should intervene, attributed to users delegating to the automated alarms their responsibility to monitor for dangers), etc. A review of medical devices, considering a subset of such phenomena, suggested they affected as much as 6-11% of user decisions (Goddard, 2014). These problems are due not only to user interface design, but also to factors such as accuracy of algorithms, user adaptation to the device, etc. They are also not limited to alarm-emitting devices but to a range of decision-support devices providing prompts, warnings, advice, etc.; e.g., in interpreting ECGs (Tsai, 2003) and screening mammograms (Povyakalo, 2013).

Our work in AQUAS started with studying how "exceptions" and security/safety/performance interactions are covered in a set of human factors standards that govern medical devices in the European market: IEC 62366-1:2015 (Application of Usability Engineering to Medical Devices) and guidance on its application: IEC 62366-2:2016; IEC 60601-1-6:2010 (General Requirements for Basic Safety and Essential Performance - Collateral Standard: Usability), and IEC 60601-1-8:2007 (Collateral Standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems). For brevity, we refer to these as "*European medical usability standards*" or *EMUSs*.

Usability standards have important roles. They expose designers to concerns and design principles produced by the specialist human factors (HF) community, particularly important for safety. It seems plausible that these standards have reduced risk for patients. However, our reading raised concerns, broader than our initial focus implied, about whether EMUSs may fail to deliver some benefits. The following three sections detail concerns about: complex design trade-offs arising within usability and between usability and other system attributes; a focus on user interface design risking relative neglect of other causes of use errors; and some definitions of terms and of the scope of the EMUSs that may create "blind spots". Last, we discuss challenges in dealing with these concerns.

2 THE ISSUE OF TRADE-OFFS

The writers of EMUSs appear concerned with the need to 'sell' usability to designers, by explaining its many advantages (e.g. Section 5.2 of IEC 62366-2). This is in line with frequent concerns in human factors circles that technically trained designers may

regard usability as not needing special attention. But this presentation may end up depicting usability as a unidimensional quality, with compliance to prescriptions being a win-win option, without concern for possible needs for trade-offs. For example, an important concern about exceptions is the rate of false positive alarms, which can lead to inappropriate user action. IEC 60601-1-8 states "algorithms that determine alarm conditions should be carefully optimized to provide, on balance, an overall benefit to patient care" and "should be designed to minimize the number of false negative and false positive alarm conditions." This latter statement seems to mask: the necessary trade-offs between rates of false positive and false negative errors; experience that designers can make alarms more sensitive, only to find user decisions become less sensitive; and evidence that the best sensitivity/specificity combination may vary between users, suggesting that user-adjustable thresholds may be advantageous, within constraints. We note that elsewhere (Section 15.2.5 of the American HE75 standard) false alarms are addressed with a more balanced approach, which discusses such trade-offs.

It seems dangerous that standards may be read as reducing usability engineering to a set of design precautions that will improve design from all viewpoints. For instance, about different aspects of usability itself, sometimes improving usability for certain operations may only be achieved by reducing it for others; e.g., putting obstacles in the way of access to functions that change the settings of a device may be required, for safety, so that more frequently used features can be used quickly without accidentally changing the settings.

The EMUSs do highlight some design trade-offs in the interaction between usability and other system attributes, such as safety or performance. A good example (Section 5.1 of IEC 62366-2) is how design for high performance in user tasks might make a device safer, as it speeds up urgent therapy, but also introduce hazards, if critical confirmation steps are omitted. On the other hand, slow task performance could "lead a well-meaning user to pass over steps in a procedure to increase speed of the procedure. This can result in a higher probability of use error linked to a potentially unacceptable risk" (IEC 62366-2).

Emphasizing such relationships is important because usability standards may be intended "to provide a single easy-to-use source of human factors design criteria" (Ahlstrom, 2008), separate from other standards on, for example, safety or security; however, considering each attribute in isolation can lead, among other issues, to unidentified hazards.

The role of usability towards device safety, mentioned for example in Section 5.1 of IEC 62366-2, seems especially important to stress since "the majority of medical device incident reports can primarily be attributed to use error" (van der Peijl, 2012) but some designers may consider usability as a secondary, almost cosmetic attribute: e.g. "keeping users happy with a user interface", much less critical than e.g. "ensuring a pump delivers the correct dose". However, as an example, displaying dose limits on a user interface not only "reduce[s] the burden on users' memory and increase[s] their confidence when programming the pump", but can also prevent a harmful dose (IEC 62366-2).

EMUSs do not yet address the relationship between usability and security, which presents a good example of necessary trade-offs. Security has attracted attention because the trend towards greater integration and connectivity between medical devices and networks brings benefits, but also security challenges about patient safety and confidentiality. A recent report (Francis, 2017) documents medical devices being targeted by cybercriminals, and that these attacks are rising. For example, blood gas analyzers and radiology equipment were found to offer backdoors into hospital networks allowing attackers to send patient records to unknown locations abroad.

HF issues with a need for trade-offs arise in user authentication. Authentication may be needed to prevent malicious use of a medical device (FDA, 2018). But requiring user authentication may be a nuisance for users, especially if required often; may delay necessary work, and, in emergencies, inhibit a user's ability to respond in a timely manner, thus posing a safety hazard. User authentication in a medical device is a good example of a many-way trade-off that cannot be solved by focusing on a single system attribute; a designer must consider the interaction between safety, security, performance, and usability.

Even for the purpose of security alone, trade-offs arise in that stringent security policies can be self-defeating if they reduce usability: they encourage users to circumvent them. For example, users required to memorise many complex passwords often respond by sharing passwords, posting them on paper notes, etc. (Zhang-Kennedy, 2016). These reactions have prompted the U.S. National Institute of Standards and Technology to reverse parts of their previous advice on password policies (Grassi, 2017).

In conclusion, we suggest that usability be presented as a multi-dimensional attribute, requiring a balanced understanding of the various trade-offs between effects of a design decision on multiple aspects, both within usability and between usability and other attributes including security.

3 CAUSES OF USE ERRORS AND RISK OF TUNNEL VISION

An important, recurring term in the EMUSs is *use error*, defined as "user action or lack of user action while using the medical device that leads to a different result than that intended by the manufacturer or expected by the user" (IEC 62366-1). The term *use error* is chosen over *user error* or *human error* to educate designers to accept responsibility for usability rather than blaming users: "although human beings are imperfect, it is inappropriate to blame the user when problems occur" (IEC 62366-2).

Removing any "blame the user" attitudes seems indeed important for good design. But current explanations seem to shift the blame from users to user interface (UI) designers: this too may cause blind spots in designers' vision. UI design is mentioned frequently both as cause of use errors – "much more commonly, use errors are the direct result of poor user interface design" (IEC 62366-1), "user interface design shortcomings can lead to use errors" (IEC 62366-2) – and as the solution: "usability engineering is a principle means to reduce [...] risk and improve patient care by reducing the potential for harmful use error through enlightened user interface design" (IEC 62366-2). Usability itself is defined as a "characteristic of the *user interface*" (IEC 62366-1). We are concerned that focusing readers' attention on user interfaces may come at the cost of neglecting other, often harder to mitigate and/or more dangerous, problems in human factors, which we discuss next.

3.1 Beyond User Interface Design

Research and incident reports indicate that thorough identification of use errors must consider: (1) the users (e.g. experience, functional state, biases towards automation), (2) the use environment (e.g. policies, time constraints, noise), (3) the device design (e.g., user interface, device reliability, level of automation) and (4) the complex interactions between these three components (e.g. automation bias, user adaptation, complacency). User interface design is but one player in a web of potential causes of use error. The EMUSs *do* mention some non-user interface related causes of use errors (e.g., Section 6.5.1 and Annex H in IEC 62366-2), but our concern

is that the focus on user interface as both a cause of, and remedy for, use errors, and the relative ease of prescriptive solutions about it (e.g., display colours, pitch of auditory alarms) may divert designers' attention from warnings about non-user interface related causes, generally not accompanied by prescriptions. We discuss examples related to exceptions, close to our own expertise, without any claim to exhausting the range of problems of interest.

As an illustration of the complex interaction of various factors to trigger a use error, consider a scenario where a patient parameter, monitored by a medical device, reaches a dangerously low level that warrants immediate user action. To start with, user action is likely influenced by whether the device algorithm is designed to detect this danger with high

enough probability. If the device does detect this danger, the alarm signal needs to be effectively communicated. However, to prevent a hazardous situation, it also matters whether, in practice, the alarm leads to correct user action, with high enough probability. This may depend on environmental factors such as whether the user is busy dealing with other, simultaneous tasks. It is also influenced by user-related factors such as users' mental models of how the device works (IEC 60601-1-10). In turn, mental models are based on users' knowledge and thus depend not only on training but also on users' previous experience of interaction with the device.

There may well be a need for standards to emphasise the role of sound user interface design, to ensure that designers take certain precautions. But

Table 1: Various causes of specific use errors beyond shortcomings in user interface design. All quoted text is from Section 16.3 in IEC 62366-2.

Use Error	User Interface (UI) Design Shortcomings	Other Possible Causes Not Related to UI Design	Potential Mitigations Addressing the Other Possible Causes
"Users fail to detect a dangerous increase in heart rate because alarm limit is set too high and users do not look at medical device display because they are over-reliant on the alarm system"	"User-adjusted high and low alarm limits on a heart-rate monitor are not continuously displayed" *(implicit solution: continuously display alarm limits)*	User chose inappropriate alarm limits either due to inexperience or in an effort to reduce the device alarm rate which they find distracting	- Consider how the alarm threshold (sensitivity/specificity combination) is set – not just choosing a more/less sensitive threshold, but also considering default settings, degrees of freedom by users, and customization according to certain attributes such as user ability.
"User ignored a warning label telling the user to disconnect the patient tube before turning the medical device off"	"The medical device did not require the user to confirm patient disconnection before powering-off" *(implicit solution: add a verification step to confirm patient disconnection before powering off is allowed)*	User, at the end of a long medical procedure, is fatigued and overlooks the importance of this step. Or other devices, to which the user is accustomed, dictate that equipment must be turned off before disconnecting from the patient.	- Add a verification step to confirm patient disconnection before powering off is allowed. - Redesign the device so that the order of these operations does not matter.
"User disregarded a warning symbol and allowed a portable medical device to run out of battery power"	"The warning symbol was not sufficiently attention-getting" *(implicit solution: make the warning symbol more visible/audible to attract the user's attention)*	Lack of reaction to an alarm due to factors such as "cry wolf". In other words, it may not be that a user did not see/hear the warning, but that their experience with the device has led them to ignore it.	- Ensure that the time between when an alarm is emitted and when the actual danger occurs is so chosen as not to cause users to ignore alarms/delay action, yet gives them enough time to react. - During user training, raise awareness against behaviours such as "cry wolf". - Consider potential unwanted interactions between different alarms, and how to group/prioritise alarms to reduce them.

designers who focus solely on the role of user interface are likely to overlook other causes of hazards, and thus fail to address them properly.

Table 1 helps illustrate the danger of such tunnel vision. The first two columns describe specific use errors and user interface design shortcomings that may cause them. They are taken from IEC 62366-2, with our comments added in italics. We add the third column to illustrate other plausible, non-user interface causes of those errors, and the fourth column for possible remedies against these latter causes. Table 1 is not meant to deny the role of effective user interface design, but to help shift the focus, using concrete examples, emphasizing that:

- some use errors can result from non-UI causes,
- although some of these other causes can be remedied by the same design mitigations that address interface design shortcomings (e.g., Row 2), some require different remedies (e.g., Rows 1 and 3). In fact, in Row 3, making the warning symbol more attention-getting not only does not address the "cry wolf" phenomena, but may even exacerbate it.
- mitigation strategies may extend beyond changes to the user interface and instead address the user or environment (e.g., Rows 1 and 3).

3.2 Potential Mitigation Strategies

It is useful for standards to mention difficult human factors issues, but equally important is discussing how they may be mitigated. Such mitigations may address: (1) user characteristics (e.g., via effective training), (2) device properties (e.g., manipulating the level of automation, adjusting alarms to focus on hazards that are difficult for unaided users to detect – i.e., increasing diversity between the device and the user), and (3) environmental factors (e.g., more effective policies on device use such as detailing how best to integrate a device into a user's workflow).

Importantly, when considering mitigation strategies, designers will need to consider the effect of *human adaptation* to automation– a critical issue that seems to be left implicit in these standards. The presence of automation makes people adapt their working procedures and cognitive processes (consciously or not) in ways that may, at times, negatively affect their performance. For instance, a very reliable alarm system may cause users to adapt to completely rely on it to detect dangers, which could result in users failing to react to dangers not alerted by the device – even dangers that they would have tackled properly without the alarm system (Povyakalo, 2013). We note that many adaptations,

even when unintentional, can be defended as "rational" in that they improve some aspect of performance, e.g. time or resources. Yet, they may also increase the risk from use error, even compared to the unaided user, at least for some category of situations (the device, while possibly reducing overall risk, could *transfer* risks between kinds of situations, and possibly kinds of patients) (Povyakalo, 2013).

Testing that only incorporates a single, often first-time, use is unlikely to reveal dangerous effects such as complacency, overreliance, automation bias, etc.; instead, these may only become apparent in post-production testing and evaluation of device logs. We suggest that it would be advantageous to encourage such focused post-production evaluation/testing.

To illustrate the significance of human adaptation on user decisions, consider a clinician's mental model of a computer-aided detection (CAD) device for cancer. Understanding mental models is important in addressing use errors (e.g., dealing with false prompts based on a user's mental model may cause a user to miss a true prompt in an area habitually known to have false prompts (Alberdi, 2014). The user may start with a sceptical view of the CAD device's capabilities, but after interacting with the device, find that it highlights difficult to find masses. This interaction shapes the user's understanding of the device's capabilities and is also likely to increase the user's trust in the device. However, even a single error may then reverse this trust (Parasuraman, 2010), which can be difficult to regain (Wiegmann 2002).

The EMUSs state that, "Ideally, an operator's mental model can be easily created through interaction with the [device] or it can be acquired through explanation from training or the accompanying documents" (IEC 60601-1-10). But non-ideal situations may exist in which a correct mental model is hard to create and maintain. E.g., users may easily learn about deterministic functions of a device from trial use of it, but be unable to conceptualise how likely some very infrequent error modes are. The dynamic nature of mental models also matters: they may change over time and depend on factors such as number and type of error committed by the device; it may be useful to alert designers to possible discrepancies between users' mental models and the true abilities of devices. E.g., a study found that users' explanations of how a CAD device behaved were based on false notions of its capability to detect breast asymmetries (Hartswood, 1997).

Such complexities may prompt designers to choose a "simple" solution: to give users more information regarding a device's capabilities and algorithms. The difficult question is: exactly how

much information? The standards suggest, in a similar manner to win-win examples presented in Section 2, that "reporting the false positive and false negative alarm condition accuracy in a standardized format allows operators and responsible organizations to understand the performance of equipment" (IEC 60601-1-8). But, by the same token, informing users of these rates can paradoxically lead to probability matching (users agreeing with the device at a rate equal to the device's reliability), which can result in decreased overall performance (Wiegmann, 2002). Furthermore, good explanations of device behaviour can make inaccurate device advice more convincing and thus increase the chance of automation bias. Providing too much information can also lead to unnecessary complexity and jeopardize users' acceptance of device advice (Alberdi, 2014). In conclusion, dealing with users' mental models by giving users more information is one example of the difficult trade-off decisions inherent in the application of most mitigation strategies for difficult human factors issues.

In concluding Section 3, we note that some of these difficult HF issues are highlighted in the aviation domain. The HF standard approved for use by the Federal Aviation Administration states that "complacency is a major concern with automation" (HF-STD-001B). This standard also has relevant references alerting designers to design decisions that may promote "complacency and may cause users to monitor automation with less vigilance"; although focus is mostly on training users "to recognize inappropriate uses of an automated device including automation bias", instead of improving by adapting the device.

We note that in a recently proposed amendment to IEC 60601-1-8 (not yet to be regarded as a standard, but released for public feedback until January 2019), new terms, such as "alarm fatigue", "alarm flood", and "nuisance alarm signal", have been added to address some of these difficult concepts. We welcome these recent additions, but reason that to help designers appreciate the true danger of these issues, the definitions need to be accompanied with examples, explanations and potential mitigations.

4 DEFINITIONS, SCOPE AND RISK OF BLIND SPOTS

Standards try to define precisely concepts they use and the scope of each rule. But precise definitions may do harm if they are inappropriate or inconsistent.

We found examples of definitions that, while they may cause no confusion for an experienced designer in a safety-aware company, are otherwise liable to cause similar dangers to those discussed earlier: missed or mis-prioritized hazards.

4.1 Conceptual Gaps from Definitions

Alarm condition is defined as: "state of the alarm system when it has determined that a potential or actual hazardous situation exists for which operator awareness or response is required" (IEC 60601-1-8). As noted directly after the definition, this suggests that an alarm condition can be invalid (a false positive). However, another note states that an alarm condition may also be missed (a false negative). But the definition implies that if the alarm system has not detected the hazardous situation then the situation is *not* an alarm condition. There is a logical inconsistency. One could think that false negatives can at least be attributed to an *alarm signal*; but this is defined as "type of signal generated by the alarm system to indicate the presence (or occurrence) of an alarm condition", thus excluding false negatives, when an alarm signal is absent despite there being a hazard. This could reduce attention to problems like mode confusion due to lack of a clear alert that a device entered fallback mode (IEC 60601-1-10); an error type known to cause accidents.

To be sure, *alarm system* is defined as "parts of [..] a medical electrical system that detect alarm conditions and, as appropriate, generate alarm signals": the intended meaning must be that alarm conditions exist in a device's environment, rather than inside it as in the definition of *alarm condition.* We note that this inconsistency remains in the current draft amendment to IEC 60601-1-8.

4.2 Scope: What Is "Abnormal"?

EMUSs are written to assess and mitigate risks caused by normal use, and to help identify but not asses or mitigate risks associated with abnormal use (IEC 60601-1-6); "abnormal use" is defined as "conscious, intentional act or intentional omission [...] that is counter to or violates normal use and is also beyond any further reasonable means of user interface-related risk control by the manufacturer" followed by a note that "an intended but erroneous action that is not abnormal use is considered a type of use error". The standards suggest that abnormal use can be distinguished from normal use through a post-test interview which establishes whether "the user understood appropriate use and made a *conscious*

decision to act (or not act) in opposition" (IEC 62366-1). But this criterion may exclude scenarios that we (and perhaps the authors of the standard, depending on how one reads the complex definition) think should be covered by risk mitigation rules.

For example, consider a device that allows users to adjust an alarm threshold for some patient parameters. A user that finds the device's alarms dangerously distracting could consciously set wider alarm thresholds than ideal for a given patient, to reduce the distraction from spurious alarms (IEC 60601-1-8). Such a *conscious* (perhaps safety-motivated) decision is likely a result of design choices: perhaps too high a false alarm rate (to achieve high sensitivity), or alarms displayed in a distracting manner. The user's conscious, inappropriate choice of threshold may cause a hazardous situation where a patient whose parameters reach a dangerous level goes unnoticed. To complicate the scenario, such user behaviour is likely to change over time depending on experience with the device and factors such as trust in the device. We think that controlling such risks should be considered in the usability engineering process. Despite the "*conscious* decision to act [...] in opposition", this behaviour is not necessarily "beyond any additional means of risk control by the manufacturers"; careful consideration of the device's alarm rate is one way manufacturers can address this risk.

5 DISCUSSION

In our review of some medical usability standards for difficult human factor issues (including issues such as automation bias, complacency, human adaptation, triggered/unmotivated user interventions, etc.), we identified a broader set of concerns than our initial focus implied, regarding: the complex design trade-offs inherent in usability decisions, a focus on UI design to the possible detriment of difficult HF issues, and finally definitions and scope. We highlighted potential risks but wish here to discuss the possible challenges in addressing them.

Easy to understand and articulate use-related hazards are not necessarily the greatest risks, and dealing with them should not preclude mitigating other, more obscure use-related hazards (HE75). Many of our observations above are in the form "this 'hard' topic is not fully addressed" followed by "in fairness, these standards refer to the problem in various passages, but lack focus or do not give a coherent warning or approach". The "obvious" remedy, "give as much concrete advice about these

issues as about the simpler topics" may however be difficult because:

- Writing and following prescriptions about known solutions to well-understood problems is easier than prescribing a valid approach to complex problems; and there is a lack of consensus between researchers about how to address many of these difficult HF issues.
- Providing practical solutions is not trivial (such as the dilemma whether to provide users with more information regarding device capabilities).
- Testing for these issues is difficult; it needs to incorporate the effect of time on user behaviour, often requiring post-production analysis, which may be infeasible and/or expensive.
- Standards need to be simple - many of these standards are already over 100 pages, contributing to the "usability paradox of usability standards" (Ahlstrom, 2008) - but this is difficult to achieve without neglecting key concepts or masking the true complexity of issues, as we exemplified in our discussions.

We nonetheless offer some ideas of possible improvements for discussion. Regarding how "hard" issues may be de-emphasised by being only raised in terms of somewhat vague warnings, a possible improvement could be to have sections individually dedicated to them and to proposed solutions, so as to add emphasis and make it easier for designers to follow a coherent approach to these problems.

We especially noted some concerns about effects of time and human adaptation. Possible improvements could be:

- adding to existing lists of questions that designers should ask themselves others like: "Does the device design encourage unnecessary interventions that may reduce the overall benefit of the device and/or increase the probability of hazards?"; "Does the device help users in situations where help is indeed useful/most needed (i.e., is there adequate diversity between the device and the user)?".
- highlighting the need for post-production analysis that focuses on identifying risks introduced by evolving user behaviour and adaptation to devices.

The easiest problem to solve seems that of inconsistent definitions or vague restricting exemptions, although the latter may also be related to contentious issues of limits to the responsibility and liability of manufacturers.

As frequent in standards, the scientific bibliography is rather old, and not necessarily

because limited to authoritative or seminal papers. This highlights the problem of separation between the standard writing process and large sectors of the research community that could provide scrutiny of the scientific basis of prescriptions, if appropriate reward mechanisms could be organised.

Certainly important to understand is how standards shape designers' decisions, focus their attention, and shift their priorities. Sociological research seems necessary. This paper is one step towards addressing challenging human factors concepts in medical standards.

ACKNOWLEDGEMENTS

REFERENCES

Alberdi, E., Strigini, L., Povyakalo, A. A., & Ayton, P. (2009, September). Why are people's decisions sometimes worse with computer support?. In *International Conference on Computer Safety, Reliability, and Security* (pp. 18-31). Springer, Berlin, Heidelberg.

Alberdi, E. P. A., Strigini, L., & Ayton, P. (2014). CAD: risks and benefits for radiologists' decision. In Samei, E., & Krupinski, E. A. (Eds.), *The handbook of medical image perception and techniques* (pp 326-330). Cambridge University Press.

Ahlstrom, V. (2008, September). The usability paradox of the Human Factors Standard. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 52, No. 24, pp. 1994-1998). Sage CA: Los Angeles, CA: SAGE Publications.

ANSI/AAMI. (2009). *Human Factors Engineering – Design of Medical Devices.* (HE75)

FAA. (2016). U.S. Department of Transportation Federal Aviation Administration. *Human Factors Design Standard.* (HF-STD-001B).

FDA. (2018). *Content of Premarket Submissions for Management of Cybersecurity in Medical Devices.*

Francis, R. (2017). *Hospital devices left vulnerable, leave patients at risk*. Retrieved on June 30, 2019 from https://www.csoonline.com/article/ 3167911/hospital-devices-left-vulnerable-leave-patients-vulnerable.html.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2014). Automation bias: empirical results assessing influencing factors. *International journal of medical informatics*, *83*(5), 368-375.

Grassi, P. A., Perlner, R. A., Newton, E. M., Regenscheid, A. R., Burr, W. E., Richer, J. P., ... & Theofanos, M. F. (2017). *Digital Identity Guidelines: Authentication and Lifecycle Management [including updates as of 12-01-2017]* (No. Special Publication (NIST SP)-800-63B).

Hartswood, M., Procter, R., Williams, L., Prescott, R., & Dixon, P. (1997). Drawing the line between perception and interpretation in computer-aided mammography. In *Proceedings of the First International Conference on Allocation of Functions* (pp. 275-291).

IEC. (2007). The International Electrotechnical Commission. *General Requirements for Basic Safety and Essential Performance - Collateral Standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems.* (IEC 60601-1-8).

IEC. (2008). The International Electrotechnical Commission. *General Requirements for Basic Safety and Essential Performance - Collateral Standard: Requirements for the development of physiologic closed-loop controllers.* (IEC 60601-1-10).

IEC. (2010). The International Electrotechnical Commission. *General Requirements for Basic Safety and Essential Performance - Collateral Standard: Usability.* (IEC 60601-1-6).

IEC. (2015). The International Electrotechnical Commission. *Application of Usability Engineering to Medical Devices.* (IEC 62366-1).

IEC. (2016). The International Electrotechnical Commission. *Guidance on the Application of Usability Engineering to Medical Devices.* (IEC 62366-2).

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human factors*, *52*(3), 381-410.

Povyakalo, A. A., Alberdi, E., Strigini, L., & Ayton, P. (2013). How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography. *Medical Decision Making*, *33*(1), 98-107.

Tsai, T. L., Fridsma, D. B., & Gatti, G. (2003). Computer decision support as a source of interpretation error: the case of electrocardiograms. *Journal of the American Medical Informatics Association*, *10*(5), 478-483.

van der Peijl, J., Klein, J., Grass, C., & Freudenthal, A. (2012). Design for risk control: the role of usability engineering in the management of use-related risks. *Journal of biomedical informatics*, *45*(4), 795-812.

Wiegmann, D. A. (2002). Agreeing with automated diagnostic aids: A study of users' concurrence strategies. *Human Factors*, *44*(1), 44-50.

Zhang-Kennedy, L., Chiasson, S., & van Oorschot, P. (2016, June). Revisiting password rules: facilitating human management of passwords. In *2016 APWG symposium on electronic crime research (eCrime)* (pp. 1-10). IEEE.