Analysis of Medical Device Recalls Owing to Output Information from Software

Ryota KITAWAKI\(^1,2\)*, Mitsuo UMEZU\(^1\), Kiyotaka IWASAKI\(^1\) and Hiroshi KASANUKI\(^1\)

Abstract

Because software as a medical device (SaMD) does not come into direct contact with patients, it does not pose such direct harm as that caused by conventional hardware medical devices (non-SaMD); however, SaMD does present indirect harm. Using US recall data, the objective of this study was to investigate indirect harm caused by software medical devices including both SaMD and non-SaMD. We collected recall data in which the cause was determined by the US Food and Drug Administration to be software design, and we determined whether the product in each recall was SaMD or not (non-SaMD). We also classified the failure mode of each recall into four groups: physical (Group 1); information-related (Group 2); data-related (Group 3); and other (Group 4). We identified 6,393 recalls for 2009-2014, and 712 software-caused recalls. Of those, the leading failure mode was Group 2 that can lead to indirect harm and accounted for 408 (57%), followed by Group 1 that can lead to direct harm and accounted for 122 (17%). Of 408 Group 2 recalls, 387 (95%) were recall class II and 6 (1%) were recall class I. Of the 6, 4 were for non-SaMD and 2 were for SaMD. The findings revealed by this study indicate the importance of the further understanding of safety of information presentation functionality and the necessity of further discussion of safety measures.
Introduction

Increasingly more medical devices have been using software in recent years\(^1\). Conventionally, software has been used in embedded form in medical devices to control hardware. Because the hardware comes into direct contact with the human body, it can pose direct harm to patients and users. In this regard, one of the most severe accidents occurred during the years of 1985 to 1987, which involved massive radiation overdoses caused by the Therac-25—a software-controlled radiation therapy device—and resulted in the deaths or serious injuries of six patients\(^2\).

Another form of software in medicine—software as a medical device (SaMD)—has recently received increased attention. In 2013, the International Medical Device Regulators Forum (IMDRF) defined SaMD as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device\(^3\). In 2014, a major change occurred in Japanese law when standalone software was recognized as a medical device\(^4\).

Because by definition, SaMD does not drive or control hardware, it cannot present a direct harm that could be caused by conventional medical device software (software in a medical device\(^3\)) ; hereafter we refer to this as non-SaMD). Instead, harm may occur as a consequence of the medical decision, action taken/not taken on the basis of information or result(s) provided by the device\(^5\). In contrast to conventional direct harm, this type is termed indirect harm, and SaMD as well as in vitro diagnostics (IVD) may lead to indirect harm\(^3\). As long as non-SaMD outputs information, it can also lead to indirect harm. Thus far, safety issues related to indirect harm in medical devices have not been studied. We analyzed safety issues related to indirect harm caused by medical device software including SaMD and non-SaMD using medical device recalls in the United States.

Methods

1. Recall data source

There are publicly available databases for medical device recalls maintained by the US Food and Drug Administration (FDA)\(^6\)-\(^7\). From these databases, we retrieved recall data from January 1, 2009 to December 31, 2014 to create a local database. Among others, the fields retrieved included the following: Recall Number, Trade Name/Product, Recall Class, Date Posted, Recalling Manufacturer, Reason for Recall, Recall Event ID, Product Classification, Product Code, Product, 510(k) Number, Recalling Firm/Manufacturer, Manufacturer Reason for Recall, FDA Determined Cause, and Action. We identified a single recall with Recall Event ID\(^8\). The FDA has determined and published the cause of each recall. We defined a software-caused recall as a recall whose FDA Determined Cause was "Design: Software Design." Our target of analysis was software-caused recalls.

2. Identifying SaMD

We determined whether the product in each recall was SaMD or non-SaMD by reviewing such fields as Trade Name/Product, Reason for...
Recall, Product Classification, Product Code, Product, and Action. Where necessary, we also reviewed the 510 (k) summary and the company’s website. We carefully classified each recall according to the SaMD definition created by the IMDRF, the accompanying seven notes, the four examples of software that were SaMD, and the six examples of software that were not SaMD.

3. Development of set of failure modes

After reviewing the Reason for Recall and Action fields, we classified each recall into four failure modes.

First, we categorized the following as Group 1: those stated as problems in physical control of the hardware, such as hardware movement, pressure, flow of air or liquid, and radiation; and those declared as non-software issues, such as overheating.

Examples of indirect harm include misdiagnosis, delayed diagnosis, delayed treatment, inappropriate treatment, and lack of treatment; they may be the result of inaccurate results, false-positive and false-negative results, and other causes. According to European Commission guidelines, “Software qualified as medical devices may also lead to indirect HARM (incorrect information generated by software)”5). We classified recalls related to incorrect information generated by the software into Group 2.

We defined Group 3 recalls as those reported as follows: disappearance of data (data loss, overwritten, removed, or corrupted); problems in communication (errors in transmission, forwarding, receiving, downloading, or uploading); problems in data access (errors in loading or opening); problems in data saving; problems in data copying; and problems in data backup.

We classified problems other than those in Group 1 (physical), Group 2 (information), or Group 3 (data) as Group 4 (other). This group included problems in behavior other than specification and runtime errors.

4. FDA’s recall class

FDA assigns each recall each of the following three classes for the relative degree of health hazard:10)

1) Class I is a situation in which there is a reasonable probability that the use of, or exposure to, a violative product will cause serious adverse health consequences or death.

2) Class II is a situation in which use of, or exposure to, a violative product may cause temporary or medically reversible adverse health consequences or where the probability of serious adverse health consequences is remote.

3) Class III is a situation in which use of, or exposure to, a violative product is not likely to cause adverse health consequences.

By using the two types of classifications together, our failure mode classification and the FDA’s recall class classification, we aimed to understand the relative level of health hazard of indirect harm.

Results


In all, 6,393 medical device recalls occurred for 2009–2014. Fig. 1 shows the total number of recalls per year and Design:Software Design category. We determined that there were 770 recalls (12% of the total) in this category.

Of 770 software-caused recalls, 712 recalls were classifiable in terms of SaMD identification and failure modes. The following results and discussion are based on those 712 recalls.
2. SaMD/Non-SaMD

Of the 712 recalls in the study period, 261 (37%) were the result of SaMD. 22 types of Product classification/Medical specialty/Product were found and the three most frequently recalled types of SaMD were as follows: 94 (36%) for "System, Image Processing, Radiological" / Radiology / LLZ; 55 (21%) for "System, Planning, Radiation Therapy Treatment" / Radiology / MUJ; and 39 (15%) for "Calculator/Data Processing Module, for Clinical Use" / Clinical chemistry / JQP.

Of the 712 recalls, 451 (63%) were the result of non-SaMD. 128 types of Product classification/Medical specialty/Product were found and the three most frequently recalled types of the non-SaMD were as follows: 49 (11%) for "Accelerator, Linear, Medical" / Radiology / IYE; 39 (9%) for "System, X-Ray, Tomography, Computed" / Radiology / JAK; and 32 (7%) for "Monitor, Physiological, Patient (With Arrhythmia Detection Or Alarms)" / Cardiovascular / MHX.

3. Failure modes

Fig. 2 presents the failure modes of all software-caused recalls, non-SaMD recalls, and SaMD recalls. Group 2 was the leading failure mode and accounted for 57% (408) of the recalls; it was followed by Group 1 (17%, 122), Group 4 (15%, 108), and Group 3 (10%, 74).

(1) Group 1 (Physical)

Of 451 non-SaMD recalls, 122 (27%) were in Group 1. Examples of Group 1 recalls were as follows: faults with the physical behavior of ventilators, defibrillators and infusion pumps; unintended laser emission; burning; overheating; interference caused by moving parts; and errors in radiation therapy.

(2) Group 2 (Information)

Group 2 was the leading failure mode for SaMD (69%, 181 of 261), non-SaMD (50%, 227 of 451), and all software-caused recalls (57%, 408 of 712). Examples of Group 2 recalls were as follows: alarm stopped or delayed and no or false alarm; incorrect message or information on the
monitor: incorrect or inaccurate analysis results: incorrect dose, therapy, or treatment plan: mix-up of patients or mix-up of right and left: inconsistent presentation of information in multiple displays: image quality issues: and incorrect or inaccurate measurements.

(3) Group 3 (Data)

In Group 3, there were 44 (17%) out of 261 SaMD recalls, 30 (7%) out of 451 non-SaMD recalls and 74 (10%) out of 712 in total. Examples of Group 3 recalls were as follows: data loss (data disappeared, were overwritten, removed, or corrupted): errors in data communication (mistakes in sending or receiving data): errors in download or upload: errors in accessing, loading, or opening: errors in data saving: errors in copying data: and errors in data backup.

(4) Group 4 (Other)

In Group 4, there were 36 (14%) out of 261 SaMD recalls, 72 (16%) out of 451 non-SaMD recalls and 108 (15%) out of 712 in total. Examples of Group 4 recalls were as follows: device behaved differently to specifications: runtime errors (system crash, freeze, or hang-up).

4. Failure modes and recall class

Fig. 3 shows the recalls according to recall class. Fig. 4 shows the recalls according to failure mode in each recall class (Fig. 4a) the same graph as Fig. 4a for non-SaMD (Fig. 4b) and the same for SaMD (Fig. 4c).

As indicated in Fig. 3 recall class II was the leading recall class for software-caused recalls, and it accounted for 94% (667/712) of the recalls: recall class I amounted to 3% (21). With regard to non-SaMD recalls, as shown in Fig. 3 recall class II accounted for 92% (417/451) of the recalls and recall class I accounted for 4% (19). As shown in Fig. 4b of 19, Group 1 accounted for 68% (13/19) and Group 2 accounted for 21% (4/19). With respect to SaMD, as shown in Fig. 3 recall class II accounted for 96% (250/261) of the recalls and recall class I accounted for 1% (2). As shown in Fig. 4c the 2 were in Group 2.

Table 1 shows all the recalls in Group 2.

Discussion

1. Characteristics of software-caused recalls

(1) Principal findings

Group 1 represents physical problems and can in principle appear in non-SaMD and means recalls that can lead to direct harm. Because Group 2 represents issues in information presented to the user, if a Group 2 recall actually led to harm, then it must be a recall of indirect harm, which is why we focused on Group 2 recalls in this study. Conversely, we cannot tell if a Group 3 or Group 4 recall led to either direct or indirect harm from given recall information. (For a Group 3 or Group 4 recall, if it actually
led to harm, it must be only indirect harm for SaMD but it could be both direct and indirect harm for non-SaMD. The affected output in Groups 3 and 4 may be for the user or for a subsequent process of the system, in non-SaMD cases. If the output was for the user, then the harm must be indirect, while if the output was for a subsequent process of the system, then the harm may lead to direct harm.)

As shown in Fig. 2 the leading failure mode was Group 2 that can lead to indirect harm and accounted for 408 (57%). This was followed by Group 1 that can lead to direct harm and accounted for 123 (17%). As shown in Fig. 3 of the 408 Group 2 recalls, 387 (95%) were recall class II and 6 (1%) were recall class I. Of 6, 4 were in non-SaMD and 2 were in SaMD.

The findings revealed by this study indicate the importance of further understanding of indirect harm from information from medical devices. The most Group 2 recalls were assigned as recall class II, meaning the middle level of health hazard, and there were even recalls with recall class I meaning the most critical level of health hazard. Stakeholders including manufacturers, regulators and users need to understand the nature of indirect harm to know what kind of information and what kind of situation is the most critical and to know how to prevent indirect harm. Also, as shown in Table 1 Class I recalls of Group 2 were highly diverse in medical specialty and device type/form (non-SaMD, IVD, SaMD). This indicates the importance of risk management for information presentation functionality in a wide range of medical devices regardless of medical specialty and device type/form.
(2) Class I recalls in Group 2

The two SaMD class I recalls in Group 2 were Z—1060—2014 (recall of a device categorized as “Gas-Machine, Anesthesia” [BSZ] in Anesthesiology) and Z—1566—2014 (recall of “System/Device, Pharmacy Compounding” [NEP] in General Hospital). During the study period, there were only those two class I recalls of SaMD. Other than those two, just one recall in 2003 was reported [13]. Two other recalls were Z—1985—2012 and Z—2136—2012, which were both recalls of “System, Image Processing, Radiological” (LLZ) in Radiology: two additional recalls were Z—1074—2013 and Z—1245—2013, which were both “System, Test, Blood Glucose, over the Counter” (NBW) in Toxicology. The latter two products were IVD medical devices.

A scenario leading to harm from inaccurate or incorrect information from medical devices was discussed and illustrated in Annex H of ISO 14971:2007 [12], where a schematic view was illustrated with this explanation: “Incorrect or delayed results can lead to inappropriate or delayed medical decisions and actions that result in harm to patients.” This scenario is also applicable to stand-alone software [12]. Toward global convergence on how this kind of risk should be recognized, the IMDRF published a guideline as a possible framework for risk categorization [9].

Thus far, this topic has been addressed as follows: by Miller and Gardner [13] in 1997, a guideline by the Swedish Medical Products Agency [14] in 2009, a 2011 FDA workshop report [15] by Meier, a report by the mHealth Regulatory Coalition and CDS coalition [16], and by Yang and Thompson [17] in 2015. Those reports addressed HIT, SaMD, CDS, and software information systems. One of the earliest discussions of this issue appeared in the draft policy of the FDA [18] in 1989. The FDA stated in the draft that medical software devices (unclassified medical software devices that are not components, parts, or acess-

---

**Table 1** Recalls in recall class I for Group 2

<table>
<thead>
<tr>
<th>Recall Number</th>
<th>Summary of reason for recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z—1060—2014</td>
<td>There was an occurrence where the patient case data did not match the patient data when the case was recalled in the Anesthesia Care Record (ACR) in that it included data from another case.</td>
</tr>
<tr>
<td>Z—1566—2014</td>
<td>PRODUCT may calculate quantities of electrolytes that are double the expected values during the creation of TPN orders.</td>
</tr>
<tr>
<td>Z—1985—2012</td>
<td>The Surgical Planning is displayed inaccurately during Navigation. This may erroneously be interpreted as a tracking related inaccuracy that may lead to patient injury during surgery. (Excerpt from the Urgent Safety Advisory Notice mentioned in Action column)</td>
</tr>
<tr>
<td>Z—2136—2012</td>
<td>PRODUCT could result in an incorrect position of the navigated instrument(s) versus the displayed reference image.</td>
</tr>
</tbody>
</table>

Note: Names of products and manufacturers not indicated.
ories to classified devices) would not be subject to active regulatory oversight if they “are intended to involve competent human intervention before any impact on human health occurs (e.g., where clinical judgment and experience can be used to check and interpret a system’s output).” This draft policy was withdrawn in 2005\(^{19}\), because “the draft policy did not adequately address all of the issues related to the regulation of all medical devices containing software”, as a result of exponential growth in the use of software products as medical devices and an ongoing trend of increasingly diverse and complex types of products\(^1\).

The FDA explicitly excluded CDS from its scope in its mobile medical application guidance of 2015\(^1\), the FDA’s current position regarding the risk of information from CDS is unclear\(^20\). However, the six class I recalls in Group 2 may partly reflect the FDA’s view. We believe that the context was similar in the case of all six recalls, the user tended to perform an action immediately based on the output information, and the action or inaction was significant to the patient. In those situations, there was no room for “competent human intervention”\(^{18}\), there were few “opportunities for practitioners to ignore or override Clinical Software System functions”\(^{13}\), the “device overruled human responsibility”\(^{14}\), there was insufficient “time to reflect”\(^{17}\), there may have been “substantial dependence”\(^{17}\) of the user on the medical device, and the “significance of the information provided by the SaMD to the healthcare decision”\(^9\) was at the highest level because the output information was used to treat a patient\(^9\). The FDA may have adopted a holistic perspective in classifying those recalls as recall class I.

Little is known about indirect harm and the events that can lead to indirect harm. The above six recalls are good examples to help us determine what kind of information and which situations are the most critical in terms of indirect harm. It will be necessary to collect additional cases similar to the ones here and analyze them individually in greater detail. This will help reinforce the current analysis of indirect harm for future prevention efforts. To carry out such investigations, more detailed description in publicized recall data and a mechanism that links recall data and adverse events would be helpful.

2. Comparison with preceding studies

(1) Relevance to safety issues in HIT

The objective of this study was to analyze indirect harm caused by medical devices using medical device recalls. Some studies have analyzed medical device recalls\(^{21–22}\), but to our knowledge none have focused on indirect harm. On the other hand, we think that the analysis of indirect harm from medical devices bears relevance to the analysis of safety issues in Health Information Technology (HIT).

The term HIT is a broad concept that encompasses an array of technologies to store, share, and analyze health information\(^23\). In terms of functionality, HIT can be broadly categorized into three categories: administrative functionality, health management functionality, and medical device functionality\(^24\). HIT offers many advantages\(^25\); however, if it is not properly designed, implemented, or maintained, it can pose risks to patients\(^26\). Accordingly, safety issues in HIT have been studied with respect to socio-technical aspects\(^27–37\). Safety issues in HIT have been investigated mainly in products categorized as HIT with health management functionality; examples include electronic health
HIT has no physical entity that comes into direct contact with patients, and it does not cause harm directly. Instead, harm may occur through improper action taken as a result of HIT in the clinical workflow. The sequence of events and factors leading to harm tend to be complicated owing to a combination of clinical workflow, use and user issues, and the interconnectedness of HIT. We think SaMD shares the above safety characteristics.

Studies about safety issues in HIT, which has complex socio-technical aspects, differ from one another in terms of data sources (as discussed in “Data sources” below) and approaches (as discussed in “Classification approach” below). However, we think our method of failure mode classification has relevance to the classifications of preceding studies (as discussed in “Classification” below).

(2) Data sources

With respect to data sources, some investigations have used observations in and interviews with selected hospitals [27-30]; others have employed reports in existing databases [32, 34-37]. Regarding socio-technical aspects, the former have tended to involve more human-related issues than the latter, which have tended to cover more technical issues. We believe that recall data reflect the actions and thinking of the manufacturer and regulator (FDA), mainly with regard to technical aspects of the broader socio-technical area.

(3) Classification approach

For their classification approach, some studies have used grounded theory [38], allowing classifications to emerge from the database [27-30, 34]; others have assigned each event to preconceived categories [32, 35, 37]. We adopted the latter approach to show a picture of indirect harm.

(4) Classification

a) Information and data

We sought to identify recalls that may have caused indirect harm owing to incorrect information generated by software. However, the term “information” is often used without a clear definition and interchangeably with the term “data.” In an attempt to clarify biomedical informatics, Bernstam, Smith and Johnson defined information as “data + meaning”, “syntax + semantics”, or “form + content” [39]. According to the National Information Assurance Glossary, information is defined as “any communication or representation of knowledge, such as facts, data, or opinions in any medium or form including textual, numerical, graphic, cartographic, narrative, and audiovisual”; data is defined as “a subset of information in an electronic format that allows it to be retrieved or transmitted” [40].

Following the above definitions, we defined Group 2 recalls as those stated as problems in the accuracy of output information (incorrect content, inaccurate information). As defined in the National Information Assurance Glossary [40], Group 2 recalls included issues related to alarms, alerts, and audiovisual monitoring. We defined Group 3 recalls as problems reported in electronic form/format that are processed by software. Similar distinctions between information and data have been found in previous studies, as follows.

b) Information-related problems

We found that safety issues in receiving infor-
mation and in medical practice based upon that information have been commonly addressed by studies that differ in data source or approach. However, studies vary according to whether they focus on human-related issues or on technical issues. For example, such categories as “errors in the process of entering and retrieving information in or from the system” identified by Ash, Berg and Coiera27) and “Information Errors: Fragmentation and Systems Integration Failure” recognized by Koppel Metlay, Cohen, Abaluck, Localio, Kimmel et al.28) discussed safety issues related to the use of HIT owing to poor product specifications that did not properly accommodate differences in workflow and users. Conversely, “Incorrect calculation” and “Incorrect content” of Myers, Jones and Sittig34), “Output” of Magrabi, Ong, Runciman and Coiera35), “Unmet display needs” of Meeks, Smith, Taylor, Sittig, Scott, Singh et al.36), and “Errors of omission or transmission” of Shuren41) discussed more technical issues, where the behavior of a product differed from its specifications.

c) Data-related problems

Similarly, the HIT studies above discussed issues related to data handling. Examples include “Transfer” and “General technical” of Magrabi, Ong, Runciman and Coiera35), “Software modifications or upgrades” and “System-system interfaces” of Meeks, Smith, Taylor, Sittig, Scott, Singh et al.36) and “Errors of omission or transmission” of Shuren41).

3. Limitations

This study relied only on medical device recall data in the United States within a limited period. Though recall data are useful, the described reasons for recall often lack full details, and so it is not possible to completely understand the cause of the recall or determine if an adverse event, including indirect harm, actually occurred. To supplement the lack of complete details, we used the recall class assigned by the FDA as a relative risk level. Our aim was to investigate medical device recalls based on indirect harm, which has not previously been studied. Thus, we did not use existing failure modes or classification frameworks for safety issues, and created a new classification. We do not report inter-rater reliability, as our aim was not a rigorous classification of medical device recalls. Subdividing Group 2 into such areas as type of information and context would be necessary to identify critical areas of information presentation.

Conclusion

To our knowledge, this study is the first study that has analyzed indirect harm using medical device recall data. We developed a failure mode classification and used it with the FDA’s recall classification to determine the relative level of health hazard of indirect harm. We identified 6,393 recalls for 2009–2014 and 712 software-caused recalls. Of those, the leading failure mode was Group 2 that can lead to indirect harm and accounted for 408 (57%), followed by Group 1 that can lead to direct harm and accounted for 122 (17%). Of 408 Group 2 recalls, 387 (95%) were recall class II and 6 (1%) were recall class I. Of the 6, 4 were for non-SaMD and 2 were for SaMD. The findings revealed by this study indicate the importance of the further understanding of the safety of information presentation functionality and the necessity of further discussion of safety measures, by diverse stakeholders taking sociotechnical aspects into consideration.
REFERENCES


9) International Medical Device Regulators Forum. Software as a Medical Device : Possible Framework for Risk Categorization and Corresponding Considerations. 2014.


40) Committee on National Security Systems.