Overview

The most well-known code of ethics for the medical profession is the Hippocratic Oath, based on the ancient writings of the Greek physician Hippocrates. The modern version of this oath is called the "Declaration of Geneva," and contains many of the same principles. Although neither of them contains the exact words "do no harm," the idea of causing no harm to the patient is the overriding focus of both oaths. Perhaps the most ironic tragedy is when a particular medicine or machine created to save life results in the loss of life.

This tragic scenario was exactly the case with the Therac-25, a machine used to treat cancer patients. There were several failures of the Therac-25 system, with at least one directly resulting in the death of the patient. There have been many studies on the Therac-25 system, and the fact that emerges is that the failures were not caused by a single problem, but rather by a series of mistakes made by nearly all of the involved stakeholders. An investigation of the Therac-25 is a fascinating case study in software engineering, both in the history of the system and in the analysis of the problems discovered in the system.

History

The Therac-25 was a machine that treated cancer by accelerating electrons and creating a very high-energy beam of radiation to kill malignant tumors. In medical terminology, these kinds of devices are referred to as "medical linear accelerators" (sometimes shortened to "linacs"). These machines usually have two modes of operation: a mode that treats shallow tumors using the electrons themselves and a mode that treats deeper tumors by using a special shield to convert the electron beam to X-ray photons. Whenever the X-ray mode is used, the shield ensures that the patient is not injured by the extreme energy levels created by that particular mode of operation (Leveson, 1995, pp. 1-2).

Development

There were two main ancestors to the Therac-25 device: the Therac-6 and the Therac-20. Both devices were produced by a business cooperation between the medical subdivision of Atomic Energy of Canada Limited (AECL) and a French company called CGR. The Therac-6 was an older device and was capable of delivering up to six million electron-volts (MeV) of energy to cancer tumors. The Therac-20 was an updated device that could deliver 20 MeV. Both devices included similar software to control the device, but this software was included merely for operator convenience and the machine was perfectly operable without the software (Leveson, 1995, pp. 2).

The Therac-25 was produced by AECL after their agreement with CGR ended in

1981. It was somewhat based on the Therac-20, but included a new concept referred to as "double-pass acceleration," which resulted in the ability to create the same amount of energy in less space. Thus, the newer machine was more compact. It was also designed to be entirely controlled by software, and did not include the same hardware interlocks that the older machines did (Leveson, 1995, pp. 2-3).This reliance on software to control the machine would turn out to be its major problem.

The biggest concern in the construction and implementation of a medical linear accelerator is to administer enough radiation to kill the tumor without damaging the patient's healthy skin cells. One of the most basic objectives is to prevent the machine from firing in X-ray mode without the shield in place. All of the accelerators before the Therac 20 had mainly relied on physical, hardware interlocks to prevent the machine from doing this. This means that the machine was physically built in such a way that it could not fire in X-ray mode if the shield was not in place. However, starting with the Therac 25, the hardware mechanism for preventing this possibility was no longer included. Instead, the machine relied on the software to prevent the machine from firing under the wrong conditions. Unfortunately, there was at least one unanticipated error in the software that resulted in several horrible malfunctions.

Accidents

There were six accidents as a result of the software error mentioned above. They occurred during the years 1985-1987 across a variety of locations in the United States and Canada. All resulted in significant health problems for the patients and at least two malfunctions directly caused the patient's death.

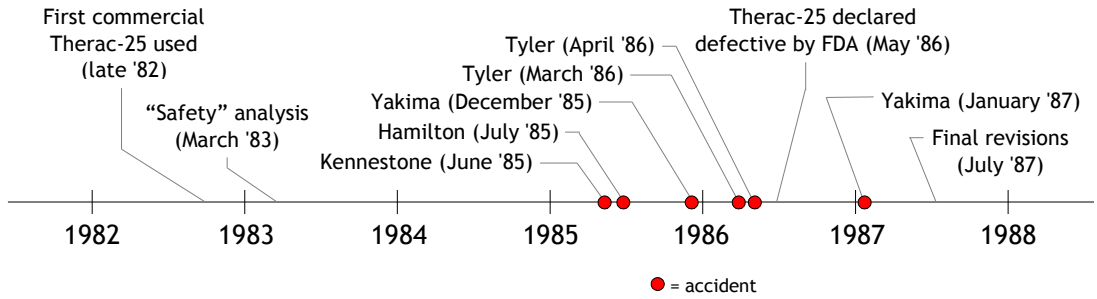

Figure 2. Therac-25 Timeline

The first accident (see Figure 2 for a timeline of all accidents) occurred at the Kennestone Regional Oncology Center in Marietta, Georgia on June 3, 1985. Many of the details of this incident were unclear since it was never reported or investigated, but the patient was a 61-year-old woman who was undergoing radiation treatment for a malignant tumor in lymph nodes near her left breast. A

normal treatment for this kind of tumor involves exposure to around 200 “rads” (the unit used to measure radiation). Instead, she received a much higher treatment, later estimated to be in the 15,000- to 20,000-rad range. This accident was not reported to the FDA until after the Tyler incidents discussed below (Leveson, 1995, pp. 9-11).

The second accident occurred at the Ontario Cancer Foundation in Hamilton, Ontario on July 26, 1985. This time, the patient was a 40-year-old woman who was undergoing treatment for carcinoma of the cervix. She received a dosage of around 13,000 to 17,000 rads. Soon after, she died from cancer unrelated to the accident, but the specialist who performed the autopsy reported that if she had not died, hip replacement surgery would have been necessary because of the radiation overdose. AECL was notified of this accident and they issued a temporary recall after a short investigation. They determined the cause to be a microswitch failure and after only minor testing issued a new version of the software that they claimed was “an improvement over the old system by at least five orders of magnitude (Leveson, 1995, p. 11-14).”

Despite the modifications made after the second incident, there was obviously still a problem. A third accident occurred at Yakima Valley Memorial Hospital in Yakima, Washington sometime during December 1985. Unfortunately, the skin redness that developed after the incident was not considered abnormal at the time and the patient, a woman, continued receiving treatments until January 6, 1986. Later investigations of the burn patterns on her skin led to the conclusion that she had received a radiation overdose. Fortunately, the overdose was not as major as with other patients, and the only consequences of this incident were minor disabilities and scarring (Leveson, 1995, pp. 15-16).

The fourth accident occurred at the East Texas Cancer Center in Tyler, Texas on March 21, 1986. Thanks to the investigative work of the hospital physicist Fritz Hager, much more is known about this incident than any of the previous ones. The patient in this case was a male who was undergoing treatment for a tumor on his back. The intended treatment was 180 rads. Instead, he was subjected to an X-ray beam of somewhere between 16,500 and 25,000 rads. The patient experienced severe pain in his neck and shoulder, later developing paralysis on most of the left side of his body and multiple bladder, diaphragm and lung problems. Five months later, he died from complications involving these injuries. Hager reported the problem to AECL, but they were unable to replicate the error and determine the cause (Leveson, 1995, pp. 16-19).

The fifth accident also occurred at the East Texas Cancer Center on April 11, 1986, only three weeks after the previous accident. After the overdose, the patient

experienced disorientation and eventually entered a coma with neurological damage and a fever of 104 degrees Fahrenheit. Three weeks later, he too died from complications involving the accident. After this second accident, Hager immediately removed the machine from use and began a extended investigation on his own with the assistance of the staff member who was operating the machine at the time of the malfunction. It took considerable time and effort, but they were eventually able to isolate the problem, which involved the operator making a very specific kind of mistake during input and then correcting it in a very short time period. This incident and the previous one were reported to the Texas health department, which then eventually relayed the information to the FDA (Leveson, 1995, pp. 19-33).

The sixth accident occurred at Yakima Valley Memorial Hospital in Yakima, Washington on January 17, 1987. Scheduled to receive a minor treatment of 86 rads, the patient instead received an estimated 8,000 or 10,000 rads. The patient had terminal cancer but because of the overdose, he died in April, even sooner than expected (Leveson, 1995, pp. 33-35).

Post-Accident History

On May 2, 1986, the Therac-25 was declared to be defective by the FDA, who demanded that AECL form a corrective action plan (CAP) and notify all Therac-25 users. By June 13, AECL had created the first CAP, which contained six items:

(1) fix the software to eliminate the specific behavior that caused the Tyler problem,
(2) modify the software high-voltage protection circuits to be stricter,
(3) make Malfunctions 1 through 64 result in a suspend rather than simply a pause,
(4) add a new circuit to shut down the beam independently of the software if high voltage is detected,
(5) modify the software to limit the editing keys to the up-cursor, backspace, and return, and
(6) modify the manuals to reflect the changes (Leveson, 1993, p. 3).

After the first CAP, there were many other revisions and new proposals as AECL and the FDA communicated about exactly what needed to be done to "fix" the Therac-25. The final results came on July 21, 1987, when the AECL released the final CAP with 23 individual proposed changes to the Therac-25 system (Rose, 1994, p. 5). Some of the major changes include:

(1) a new version of the software to correct the Tyler and Yakima problems,
(2) a method of allowing the operators to independently monitor the turntable

position,
(3) a foot-switch that the operator had to hold down during operation to prevent the machine from firing without the operator's consent,
(4) a hardware turntable interlock like that in the Therac-20, and
(5) a modification of the control keyboard to eliminate ambiguous keys (Leveson, 1993, p. 4).

In 1988, AECL changed the name of its medical division to Theratronics International Ltd. and tried unsuccessfully to sell it to a private corporation. Theratronics suffered under an FDA ban on its medical devices until 1994 as a result of the Therac-25 incidents and it no longer produces linear accelerators. (Rose, 10 Nov. 2004)

Problem Analysis

As mentioned before, there were many areas of weakness in the Therac-25 system. These problems can be conveniently divided into three main categories: software defects, systematic inadequacies and communication deficiencies.

Software Defects

The term "software defect" is the technical term for what some people would refer to as a "bug." The problem with the word "bug" is that it implies that the problem happened by itself. In software engineering, however, it is taught that software problems do not cause themselves, but rather are "injected" into the software by the developers, usually unintentionally. A "defect" is a flaw in the software instructions (code) that ultimately causes a failure to occur. The Therac-25 code had several defects: (1) a possibility for mode overlaps, (2) a reliance on slow input and (3) ambiguous error messages.

The ideal situation is that the operator would know the exact state of the machine at all times and that the machine carries out every instruction exactly as intended by the operator. Unfortunately, this is not always the case. There was a small defect hidden deep within the code for the Therac-25 that, with a particular set and method of inputs, allowed for the machine to perform in the X-ray mode while displaying information on the screen corresponding to the electron mode (Rose, 1994, p. 4). The unfortunate result of this problem is that the machine could operate in a mode contrary to the operator's expectations or knowledge.

This problem is compounded by a reliance on slow input. The mode overlaps only occurred when the operator made a certain kind of mistake and then performed the corrective actions very rapidly. This only happened after the operator had operated the machine many times before, so many times that he or she was performing the

actions purely from memory without thinking about it. The software was not prepared to deal with this possibility.

This reliance on slow input would not be such a problem if the software had contained more descriptive error messages. During the accidents, the operators received a "Malfunction 54" message (Leveson, 1993, p. 2). This is extremely ambiguous and does not indicate the type or severity of the error, nor does it suggest possible solutions to the problem. The operators had gotten used to many such "errors" during normal execution and the normal procedure when an error occurred was simply to re-try the treatment. This kind of "error handling" system is unacceptable.

Why were these defects still present in the final version of the Therac-25 that was released to hospitals? Why were they not detected and removed during the design or testing phases of software development? The unfortunate, surprising and inexplicable fact is that the Therac-25 software was developed largely by a single person who did all of the design, programming and testing. This reliance on a single individual to perform every phase of development is bad software engineering practice and led to the injection of the major software defects discussed above.

System Problems

After examining the software defects, some people might decide that the cause of the failure had been found and therefore that the analysis process was done. However, this would be an oversimplification of the situation. The software defects mentioned above were indeed problems, but the failures still could have been prevented if it had not been for several systematic problems. "System problems" are problems that concern the entire Therac-25 system, including the hardware, the software, the patient, the operator and the interactions between them all. There were two main system problems: (1) an excessive reliance on software and (2) a lack of sufficient testing.

The overriding system problem in the Therac-25 project was that the developers (as well as the manufacturers and hospitals) placed too much confidence in the system's software. First of all, their fault tree analysis in March 1983 of the "safety" of the Therac-25 system did not include the software at all (Leveson, 1993, p.1). Secondly, they omitted a hardware interlock to physically prevent the machine from being misused. As mentioned earlier, this apparatus was present in earlier versions of the machine but removed for the Therac-25 in favor of a purely software-controlled version. These two factors combined with their unwillingness to admit to a software problem until after several major incidents betray an extreme over-reliance on the software subsystem. This kind of blind confidence is unjustified and naïve.

To ensure the safety of a system such as the Therac-25, the software must be just as rigorously tested as the hardware. Unfortunately, as mentioned above, the system was not tested as well as it could reasonably have been. It should have been tested with real operators to the extent that the operators became experts and were able to produce the quick keystrokes that eventually led to the accidents. Another problem is that the system was assumed to be fixed after the first malfunction and revision without extensive testing. Testing the product after revision is imperative to ensure that the problem was indeed fixed and that new defects were not injected during the revision process.

These system problems are well evident on hindsight. However, sometimes it is very difficult to distinguish them during the process of design and implementation. In particular, the process of testing is a very difficult issue. How much testing is enough? How does a company balance testing with budget, schedule and project scope? The only realistic solution is to set reliability goals and attempt to achieve them. If the reliability goal of the Therac-25 system was zero probability of failure on demand (POFOD), it is obvious that the system failed to meet its goals.

Communication Problems

Even after discussing the software defects and the system problems, there is still more to investigate. After the Therac-25 failures occurred, there were several major communication problems that prevented the failures from being reported. This lack of reporting delayed solutions from being formulated and through inaction may have caused more accidents to occur. There were communication problems on the part of nearly every group involved: the machine itself, the operators, the hospital staff, the manufacturers and the Food and Drug Administration (FDA).

The communication problems actually started with the machine and the operators. As mentioned earlier, the Therac-25 gave very unhelpful error messages. In this sense, the machine was unable to communicate to the operator the details of the malfunction. The machine gave minor (no risk to the patient) error messages regularly, and this put the operators in a very awkward position. Should they carefully investigate every error or simply ignore them and continue the treatment? By default, they assumed the latter response after the errors became routine. More useful and detailed communication between the machine itself and its operators may have prevented the malfunctions.

However, there were many more communication problems on higher levels, especially the hospital level. Hospitals have three general routes to reporting medical device failures: (1) report to the device manufacturer, (2) report to a third-party

reporting agency or (3) report directly to the FDA. Unfortunately, there is little evidence that any reporting was done regularly, and most of what was done was only passed on by word-of-mouth. The United States General Accounting Office conducted several reports on medical device malfunction reporting and received some startling results. One report indicated that only 51% of device malfunctions were reported to any organization outside the hospital and only a mere 1% actually ended up in the files of the Food and Drug Administration. Perhaps even more concerning was the finding that the more serious accidents were actually less likely to be reported than the trivial accidents (Chelimsky, 1987, p. 5-7).

In investigating the communication problems involving the Therac-25, one major theme surfaces time after time: no one wanted to talk about the problem. The operators didn't know what had happened and didn't want to be blamed for the accident, the hospitals didn't want an expensive lawsuit, the manufacturers were concerned about reputation and legal problems, and the FDA had no way of efficiently gathering information. After the Therac-25 incidents and other problems with accident reporting, changes were made to the process of device malfunction reporting to make it easier for anyone to report and deal with problems.

Conclusion

After examining all of the problems associated with the Therac-25, ranging from software problems to communication problems, the question must be asked: which problem was the worst? Unfortunately, there is no answer to this question. All of the problems contributed to the scale of the tragedies involved. The goal is to recognize the problems and the relationships between them in order that they might be prevented in the future. Mistakes were made, but pointing fingers is useless. To learn from the past and to put the newly-found knowledge into use in the future would be the best possible outcome of the Therac-25 incidents. The future for medical linear accelerators is bright, and the lessons learned from the Therac-25 could soon push the use of these machines to entirely new levels of effectiveness, reliability and safety for cancer patients across the nation and the world.

Works Cited

Bowsher, C. (1989, November 6) *Medical Devices: The Public Health at Risk*. United States General Accounting Office. GAO/T-PEMD-90-2. Retrieved November 10, 2004, from http://www.gao.gov/

Center for the Study of Ethics in the Professions, Illinois Institute of Technology. *Declaration of Geneva*. Retrieved November 10, 2004, from http://www.iit.edu/departments/csep/PublicWWW/codes/coe/World_Medical_Association_Declaration_of_Geneva_1994.html

Chelimsky, E. (1987, May 4) *Medical Devices: Early Warning of Problems is Hampered by Severe Underreporting*. United States General Accounting Office. GAO/T-PEMD-87-4. Retrieved November 10, 2004, from http://www.gao.gov/

ComputingCases.org. *Therac-25 Case Materials*. Retrieved November 10, 2004, from http://www.computingcases.org/case_materials/therac/therac_case_intro.html

Leveson, N., and Turner, C. (1993, July). An Investigation of the Therac-25 Accidents. *IEEE Computer*. Volume 26, Number 7. Retrieved November 10, 1004, from http://courses.cs.vt.edu/~cs3604/lib/Therac_25/

Leveson, N. (1995) Medical Devices: The Therac-25. Appendix taken from *Safeware: System Safety and Computers*. Addison-Wesley, 1995. Retrieved November 10, 2004, from http://sunnyday.mit.edu/papers/therac.pdf

McDaniel, J. (2002, May). Improving system quality through software evaluation. *Computers in Biology and Medicine*. Volume 32, Issue 3. Retrieved November 10, 2004, from ScienceDirect.

Medical Device Recalls - Examination of Selected Cases. (1989, October) United States General Accounting Office. GAO/PEMD-90-6. Retrieved November 10, 2004, from http://www.gao.gov/

Rose, B. (1994, June 6). Fatal dose. *Saturday Night*. Volume 109, Issue 5. Retrieved November 10, 2004, from WilsonWeb Journal Directory.

Taylor, R. John Hopkins University. *The Therac 25 – A Case Study in Safety Failure*. Retrieved November 10, 2004, from http://www.cs.jhu.edu/~cis/cista/445/Lectures/Therac.pdf