Cybersecurity problems in a typical hospital (and probably in all of them)

Harold Thimbleby

Swansea University
Wales

Abstract *A criminal case balancing on the corruption of patient data in a UK hospital resulted in some nurses being acquitted and some given community service and custodial sentences. This paper explains the background, demonstrates the inability of hospital IT systems to provide reliable evidence, and highlights broader problems with IT culture affecting manufacturers, hospitals, police, legal advisors — and ultimately misleading clinicians and compromising delivery of care.*

The NHS (and healthcare more generally) urgently needs to improve its IT awareness, management and policies. The police and the legal system need a more mature approach to IT. Manufacturers need to provide dependable systems that are fit for purpose for complex hospital environments. Regulators should ensure that systems meet better standards of quality and dependability.

This paper includes recommendations; the most fundamental being that hospitals acknowledge that IT is unreliable and they should procure and manage equipment with this in mind. In particular, mature and effective data protection and cybersecurity policies must be in place and used proactively. When problems occur, evidence derived from IT (whether systems or devices) must not be used in legal or disciplinary investigations without extreme care and independent proof of provenance.

1 Introduction

This paper summarizes my insights from being an expert witness in a criminal case involving alleged fabrication of patient data by nurses.

The outcome and details of the court case are in the public domain, but the aim of this paper is not to tell a story about a hospital or its nurses, but rather to tell a more worrying story: *this could happen anywhere — and probably is happening everywhere*.

The court case collapsed because prosecution evidence was derived from flawed data, flawed IT and flawed management of IT. Nurses were blamed, but the underlying causes must be understood as basic cybersecurity issues that should have been taken seriously as such when they happened. The allegedly incriminating data and

the later corruption of data (the final understanding of which led to the trial collapsing) should have been detected as and when they happened.

One wonders how many other cases inappropriately pursue clinicians caught up in fallout from IT chaos, with nobody recognizing or wanting to admit or check that IT can cause such problems. It is worrying that the case here very nearly did not end at all happily, and would not have ended as well as it did without a lot of work correcting widespread misunderstandings of IT — and, harder!, correcting *widespread unconscious and unintentional misunderstandings* of IT.

Although we are not criticizing hospitals or their staff, in view of understandable sensitivities this paper does not provide any citations to the court case or to related evidence. We have avoided using identifiable names in this paper, though we have not changed technical details or standard procedures. However, numbers have been rounded and ward names changed, etc. Although the blood glucometers and databases are made by a company we will call TechCo, we do not think this company is egregious: we think their products are of typical quality and design for the industry. The story is therefore representative of the industry and its regulation, not about any one company; similarly, the story is not about one hospital nor about any nurses in it; it is about *all* hospitals and their staff, and what can unwittingly happen.

At a higher level the story is about the widespread misunderstanding of IT in healthcare, and in particular, about mismanagement of IT by hospitals and by the police. Note that TechCo's systems are used across the NHS and worldwide.

The one sentence take home is that there must be effective, mature procedures and understanding in place to detect and manage cybersecurity problems before they trigger catastrophes. The story here fortunately involved no patient harm or malicious hacking, but that was only by luck. Healthcare cyberattacks are "growing exponentially" (Davidson, 2016) — with 113 million US electronic health records breached in 2015. Pure luck protecting staff and patients cannot hold out for long.

2 A public perspective

Concerns about the quality of patient care in a hospital ward led to a police investigation. For the criminal investigation, the police focused on the treatment of vulnerable adults, which may be criminal even though there is no patient harm. Indeed in this case there was no patient harm caused by poor care.

There was considerable political and public interest in this case, particularly since the powerful and high-profile criticism of Robert Francis's 2013 *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry* ("The Francis Report"); nobody wanted another Mid Staffs.

The news reported stories of many — over 50 — nurses being investigated and soon of an imminent court trial. (I did not follow the internal investigations.) The nurses were alleged to have fabricated blood glucose readings (that is, not having actually taken any readings from patients) and then written them up in paper patient notes. For vulnerable adult patients this would be criminal. The implication was that

the nurses were lazy and dishonest and had put patients at risk. Publically, it was known that three nurses pleaded guilty, but two pleaded not guilty and their case proceeded to a jury trial.

TechCo's blood glucometers are used in the hospital, and they automatically upload glucose readings to a central patient record system. The police established that the central record system had no records of many tests the nurses had written on patient paper notes. Therefore the police concluded that the nurses had written down fictitious readings and not bothered to do their job properly.

The police were thorough in their investigations and considered various ways the nurses may have made accidental errors. The police compared paper records with a computer database, involving around 150,000 test records — a great deal of combined manual and computer work!

In addition to identifying alleged fabrication (that is, paper records with no corresponding computer records) the police also found evidence of many cases of poor operating practice. For example, a nurse is supposed to enter the patient's ID, but sometimes a nurse will scan their own staff card instead. This is easier and enables the glucometer to work, so the nurse can quickly obtain a test reading. From the computer records it is clear there are many cases of this practice. (The next section, below, describes in more detail what nurses are supposed to do.)

The accused nurses had followed such bad practice repeatedly, and under the UK Criminal Justice Act this was considered evidence of "bad character." Put briefly this means that if proven guilty your sentence may be harsher: not only are you guilty of the crime, but you are a bad person. The bad character concept makes sense when a crook is arrested for one, perhaps relatively petty, crime but nevertheless is known to be a hardened criminal.

The prosecution argued that there were no problems with the equipment. There are national databases in the UK and USA for reporting problems, and no related problems had ever been reported with TechCo's systems. Therefore this was a nursing problem, not a system problem, they argued.

The case thus went to trial … but weeks later the trial collapsed. The two nurses were released.

TV crews were there and filmed a patient victim group protesting outside the court. The media presented the collapse of the trial as a failure; as if the nurses were still guilty because the trial only collapsed on legal technicalities.

The other nurses who had previously pleaded guilty did not change their plea and were later sentenced, some to community service and some to prison.

3 What do nurses do on the ward?

To help patients manage blood glucose levels (particularly if the patients have limited capacity to look after themselves) it is important to take and record blood glucose test readings. Using the TechCo blood glucometer, an outline of the operating procedure is follows:

1. Find a glucometer;
2. The nurse then identifies themselves to the device (by scanning the barcode on their staff card or by typing their ID);
3. The patient ID is scanned from their barcode or typed;
4. Scan a glucose test strip, clean the patient's finger; the patient is pricked and a drop of blood placed on the test strip;
5. The test strip is inserted in the glucometer;
6. The glucometer displays the blood glucose level (or possibly an error);
7. The nurse may then take immediate action to address any clinical issues;
8. The nurse then "contemporaneously" writes down on the paper patient notes the time and reading;
9. One further step, that has no immediate clinical significance, is that the glucometer must be placed in a dock, and then its data will be automatically uploaded to TechCo's central systems.

The TechCo glucometer itself can record over 2,000 readings before it needs to be docked, and it will warn if its memory is full. This memory feature means that TechCo's glucometer can be used for batching tests: a nurse can test a whole ward, respond to patients' needs, then later write up the results using the glucometer review screen to recollect individual test details.

Once successfully uploaded to TechCo's systems, the test readings will later appear in the main patient records available on ward PCs — however, this might take days or longer (see below).

3.1 Patient ID workarounds

Barcode workarounds are a well-known problem (Koppel, et al, 2008), and indeed sometimes it is hard to scan or read the patient ID, so one workaround is to type 000 etc on the glucometer keyboard, or more easily just scan the staff ID barcode again just to get a number the glucometer accepts as a valid "patient ID." Using the staff ID instead of the patient ID is called "double tapping" and (perhaps) coincidentally the number of nurses in the database who double tapped at least twice in the period for which I have data is approximately equal to the number of nurses originally investigated.

The glucometer accepts both of these workarounds (arbitrary IDs and double tapping) and will still give a correct blood glucose reading. Behind the scenes, however, the hospital had configured its systems to reject this data, which then requires manual intervention (which may never happen, or may introduce further errors) to fix it. It should be noted that configuring the system in this way makes sense: the blood glucose reading in the database cannot be reliably associated with any particular patient in the database until the issue is manually resolved.

I would argue that double tapping and other workarounds, since they are trivial to detect, should be sorted out immediately rather than ignored — TechCo's systems can be configured to detect these problems immediately they occur. Since double tapping was so prevalent, I believe nurses were not aware it was problematic nor that it was being monitored. Since double-tapped data at this hospital got lost, this fact would further confirm to nurses the irrelevance of docking the glucometer.

4 An expert witness perspective

I was invited by the defence team to be an expert witness. In fact, an expert witness in the UK works for the court, but in this case the defence thought the truth would help their case.

My first comment was that if so many nurses are all alleged to have made the same mistakes, it is more likely there may be a common explanation, such as an IT failure — which would affect everybody. The legal view, however, was that if many nurses are making the same mistakes "they are all in it together" and the bad character concept applies.

My first task was to analyze the prosecution evidence (presented as a CD of Excel spreadsheets and, later, data logged on blood glucometers and XML files) to see if the police had made any mistakes claiming that the test data was not there.

It was easy to show that the data the police claimed was not there was indeed not there. Nor was closely related data, as might be expected if glucometer clocks were not synchronized with nurse watches, or if other minor transcription errors had been made, and so on.

If data was not present, it implied, so the prosecution claimed, that the nurses had fabricated doing actual tests — for if they had actually done the tests, the data would be present in the spreadsheets they argued. That is possible, but I thought it far more likely that IT problems or even a "technician with a grudge" would be a simpler explanation — indeed, normal operation of TechCo's system *requires* administrators to make changes to data, for instance to sort out double tapping.

That over 20% of the database had an error flag set my raised suspicions; this was becoming a much more complex story than the prosecution painted. Another worry was that a comment field on each test said "Wrong patient" for just 2 entries and *nothing at all* for the remaining hundreds of thousands of entries — suggesting to me that nobody was really paying much attention to the management of the database; indeed, the "reviewed" flag was false for almost all data entries.

I noted that staff names occurred with many implausibly close variant spellings (e.g., differing only in capitalization or spacing, or variants like Jon and John but with the same surname and ID). Many identical staff names occurred with different numeric IDs. All this, and more, suggested the database was not well-managed and might not be reliable for the purposes the prosecution wanted it for. Moreover, the poor staff data suggests that the TechCo features for only permitting authorized staff

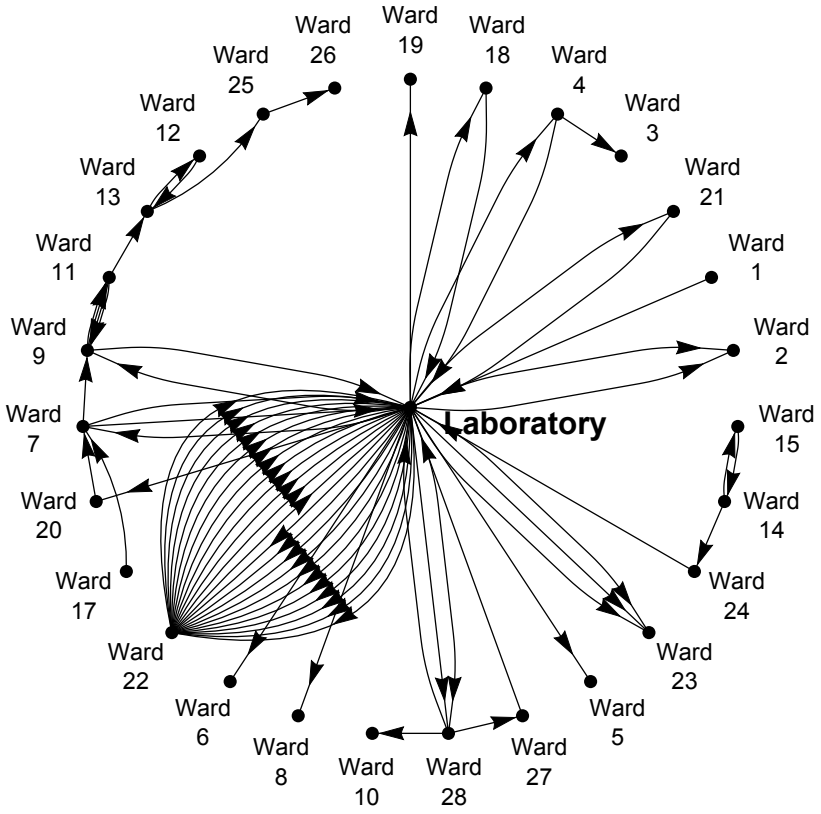

Figure 1. Recorded movement of glucometer dockings around the hospital over a period of one year. Note the centrality of the laboratory as a hub of movement, and that Ward 22 seems to have a lot of activity — 25 movements. Since wards presumably try to maintain a constant stock of glucometers, there must be other movement that is not being recorded. (The diagram layout is arbitrary and unrelated to the real numbering and locations of wards, and to further help preserve anonymity, some of the "wards" aren't strictly wards at all.)

to use the glucometer might easily have been compromised (how can it reliably lock out unauthorized staff when the staff data is so poor?).

Unfortunately for the police, proving absent data is absent *for a specific reason* is hard when the provenance and quality of the data is in question. Moreover, Excel spreadsheets have no way to audit: it is impossible to tell whether rows or columns have been deleted, been edited, or even had never existed.

The police claimed they had used forensic methods to make the copies of the database. In fact, some being CSV (an Excel data format) files proved there had been manual intervention: the hospital database was SQL so a manual process had converted it. Some of the Excel worksheets had differences further strongly suggesting an unreliable process had been used to create or edit them.

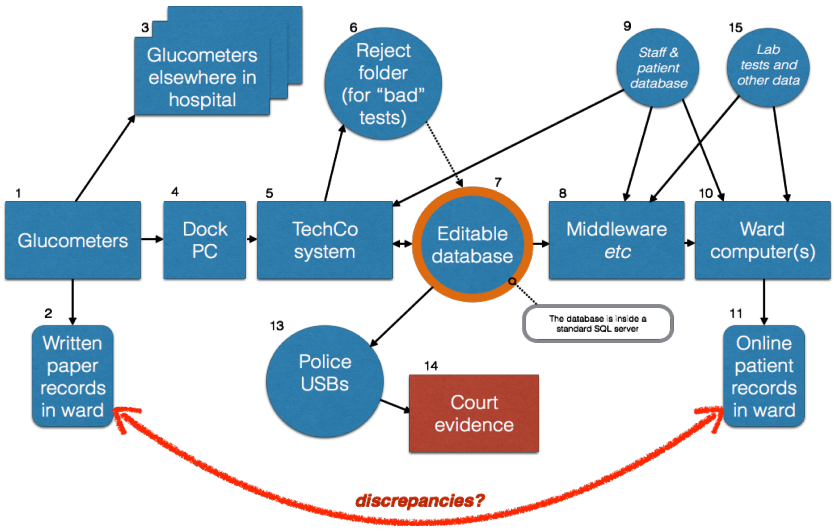

Figure 2. Diagram presented in court (anonymized) originally sized as A4; the smaller reproduction here serves to indicate the complexity of the network, but note that Box 8 contains unknown further software. The basic problem was that there were discrepancies between the paper records (bottom left) and the final computer records (bottom right). Note that not all relevant systems are TechCo's (the "middleware" box may, and probably does, contain further complications). Numbers in the figure were used to cross reference this diagram to other expert evidence.

The police had copied Excel spreadsheets at the hospital to a USB stick and only then digitally signed the data and held it securely. Unfortunately, this "forensic rigor" came too late: the police should have made a signed copy of the original database, not a manually created copy on a USB stick. Anything could have happened to contaminate the Excel data earlier than the signing. Had rows or columns been deleted or edited, Excel provides no way to tell.

The police seized several blood glucometers from the ward in question and presented the data on them as evidence. The police failed to seize at least two other glucometers that had been docked on the ward over the period of the alleged fabrications but which were (presumably) in other wards or being serviced when the police seized the ones they did. (Note that the database only shows where glucometers are docked, not where blood glucose tests are made — to know that, the test data needs to be related to patient/ward data.)

The police sent the seized glucometers to TechCo to confirm they worked correctly and to get their data. Unsurprisingly, TechCo said they worked correctly.

In fact, as figure 1 shows, glucometers move around the hospital. If a glucometer has a fault (e.g., a dead battery) would be returned for servicing and replaced by another. If a nurse needs a glucometer but cannot find one, they may borrow one from another ward. Glucometers may also get lost, perhaps at the back of a cupboard or sent off for repair. In the Excel data, glucometers were used almost hourly during the day on the ward in this story, but some glucometers in the hospital were not used

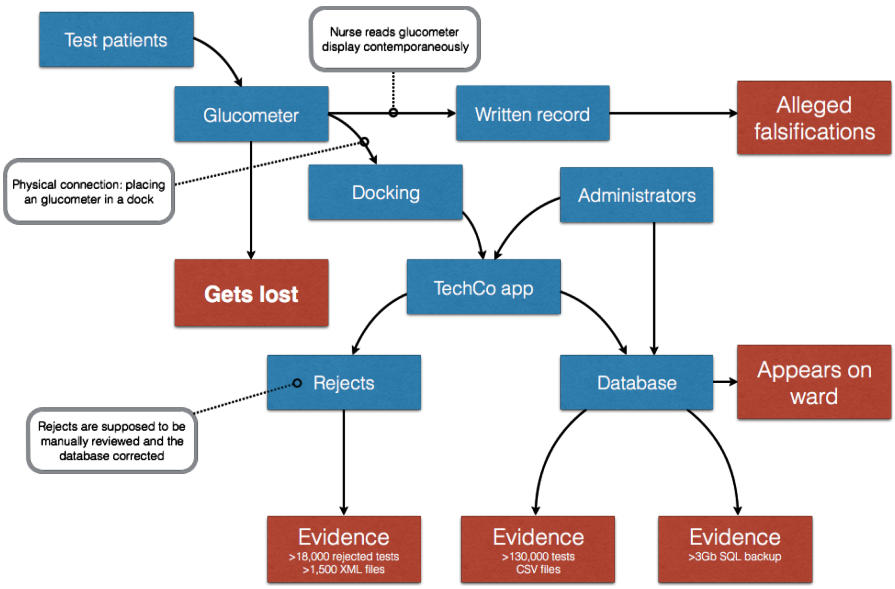

Figure 3. Diagram presented in court (anonymized) to show key sources of evidence, though originally sized as A4 — for the present paper the details are less important than noting the complexity, with no end-to-end checking. The police made digitally signed copies of Evidence 1, 2 and 3, but signing occurred *after* the police had manually copied the data, so it could only be used to show the data had not changed after it had been collected. The digital signatures did not assure that the data was what it was claimed to be. This diagram does not show TechCo's modifications of data, which only became apparent after the diagram submitted to the court.

at all for over 100 days and many not used at all for over a week — where were they? Where were *all* the glucometers that had *ever* been used on the ward when the police came? I do not think the hospital kept an inventory that tracked where glucometers were: if it had, it would have formed an essential part of their evidence.

It is possible that the alleged non-measurements are still sitting on a glucometer somewhere, but which is still waiting to be docked. Indeed, one XML file I got during the trial showed a 4 year gap between a measurement being taken and the data transferred to the database. The alleged incidents happened less than 4 years before the trial, so perhaps the missing data is still on its way — the alleged incidents did not happen that long before the trial!

There was much internal evidence that the databases were of very low quality, and of course there was the problem that there was no forensic route from the original SQL database to the Excel worksheet; worse, by TechCo's design, there was no reliable connection between the glucometers and the main database itself. There were many failure points, for example (see also figures 2 and 3):

- A glucometer may lose data itself;
- A glucometer may not be docked;
- A glucometer may be physically lost or returned for repair;

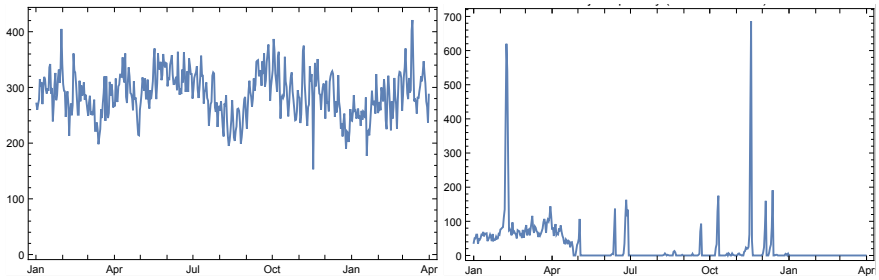

Figure 4. Side-by-side comparison of accepted glucose tests per day (left), with rejected tests per day (right) over the same period. Successful tests per day closely track patient numbers, but rejects show no obvious pattern — one would assume rejects per day to correlate with the number of successful tests per day. In particular, note long periods of exactly zero rejected tests contrasting with brief periods of very high reject rates, almost double the number of accepted tests. A simple explanation is that some records of rejected tests were manually deleted, some may have been merged, and possibly some dates were arbitrarily corrupted.

- Docking may fail, whether because of manual interference on the ward or by technical issues such as internet connectivity problems, unrecognized new servers and so on;
- TechCo's glucometers only store about 2,000 readings, yet the database shows they were used for nearly 5,000 tests: the data on the glucometers is not evidence whether nurses used the glucometers earlier than the glucometer records cover;
- … this list is not exhaustive.

Once docked, the data has a tortuous route through middleware and TechCo's own software. It can take days to get through. In particular, manual intervention is required for some data, but there was no evidence provided that any manual interventions had occurred. For example, if a glucose reading is rejected by the software, it is put into a holding folder and must be sorted out manually. Many patient IDs on the database were short-hands like 000, probably occurring because for the nurse it is far more important to get a measurement than worry about the technicalities of what happens after the glucometer is docked. Nobody had sorted out this poor data.

Such problems have not been reported on national databases to my knowledge (such as the US FDA's Maude), but there are peer-reviewed research papers that report identical problems at other hospitals with *the same TechCo devices* [so I shall not cite them]. At one hospital discussed in one paper, starting to check the database reduced poor practice such as using staff IDs for patient IDs. So what the prosecution should have concluded is not "this problem does not occur anywhere else" (i.e., the nurses must be being unprofessional) but rather "this problem routinely occurs elsewhere, but nobody worries about it enough to report it" (i.e., the nurses are behaving normally).

4.1 The collapse of the case

During the trial, the prosecution produced new evidence: an encrypted CD of gigabytes of XML files. This was the first evidence I had seen that had timestamps covering the period of the alleged incidents. (Though, as before, there was no digital signature or proof of provenance.)

The court adjourned while we analyzed this data. With effort the defence and prosecution then agreed a joint statement on the significance of this data, though some parts of the report, covering different features of the data, were written separately. The prosecution's case continued: the critical missing data was still not present.

I noted that the data had a very peculiar distribution, strongly suggesting data mismanagement, bugs or equivalent. See figure 4 for a simple example from my analysis.

TechCo was therefore called to be cross-examined over the joint report. During this cross-examination it became clear that this witness had visited the hospital before the police had seized a copy of the data. It emerged that TechCo had "tidied up" the database, and they had kept no records of what they had done.

The judge then made a ruling: the prosecution evidence was unreliable and had to be excluded. Asked what they wanted to do, the prosecution said they had no evidence and sat down. The judge ordered the jury foreman to clear the defendants as there was no case to answer. "Set the prisoners free!" he said.

My many other arguments for the unreliability of the evidence therefore never needed to be raised or cross-examined before the jury. For example, the trial did not get to exploring the consequences of the police seizing the wrong glucometers, of the hospital not having (or appearing not to have) any inventory for tracking glucometers. The court never explored the very poor data quality or reasons behind it.

The prosecution had implicitly gambled that any problem with their evidence could undermine the credibility of all of it. They ultimately needed to prove that *absent* data was caused *specifically* by deliberate nurse behavior, and they needed to prove that absence proved nurse fabrication rather than other possibilities. I would have liked to have shown, for instance, that the poor design of TechCo's equipment could provide no evidentially acceptable relation between absence and nurse behavior, but the confession that TechCo had manipulated data (and forgotten exactly what they had done) was enough to undermine all the prosecution evidence.

If the glucometers had been designed appropriately, they would have kept better records of successful (and failed) end-to-end blood test transmission. If they had, either the police would have had a very easy job (if the nurses had actually fabricated data) or the hospital would have easily known about their poor IT systems before the police turned up. Certainly if the TechCo database had included confirmed transfer and missing data information, the police would have known immediately the evidence was unreliable — though it will be recalled that they did not notice error flags and other indicators of poor quality in the data they actually had.

This product is not for diagnostic use; all patient diagnostics should be based on results reported by the point of care instrument.

Figure 5. Copy taken directly from TechCo's database system manual. It is worth noting that the nurses were doing what the manual says: they were using the point of care instrument (the glucometer) and writing down the results.

4.2 Technical discussion

The blood glucometers and related systems were not designed to be dependable. Or at least they were not designed to be dependable for any purpose other than taking measurements and immediately displaying results. The written evidence from TechCo strongly suggested that the software in them was not of high quality: for example, they wrote "[…] we should hopefully be able to confirm that this meter also had no corrupt records." But "hopefully" is not good enough!

It is interesting that there are research papers showing the TechCo glucometers are accurate blood glucometers. The prosecution understandably mentioned this, arguing that they were therefore good devices. As a matter of fact, accuracy was not relevant to the case. Regardless of whether the glucometers were accurate, the issue was *fabricated* readings not whether they were *accurate* readings. The relevant quality criterion for the case was whether the glucometers reliably transmitted test data to the hospital's patient record system. I could find no research exploring this aspect of their reliability. From the vague written evidence from TechCo, I suspected they were not very reliable at all in this regard, or at least they had been designed (by TechCo) so that TechCo could not answer this question definitively.

TechCo's database software has a warning "this product is not for diagnostic use" (reproduced in figure 5) — and if it is not good enough for diagnostic use, why was it used for evidence? Why did the hospital even have software that was not for diagnostic use managing glucometer databases? Rejected test data (e.g., with bad patient IDs) has to be edited with this product; if that isn't clinical use, what is it?

To be charitable, the database might have been intended for maintenance of the glucometers. For example, while I was double-checking I was correctly interpreting it, I plotted performance of the glucometers against battery voltage and temperature, data the glucometers record. There were interesting digitization effects, though I could not tell whether this was how the glucometers worked or whether it was an artifact of processing the data through TechCo's software, SQL or Excel.

Surely, if the data had not been used for clinical purposes, then the police should have been very cautious extrapolating to imply criminal clinical practice: they should have used independent evidence to establish the records were reliable. Given that the glucometers did not store all relevant data, the only independent evidence was the written patient records. The police assumed they were fabricated.

The police assumed the glucometers and hospital IT systems were completely reliable, even though they knew they must require human intervention. The management of the data was not questioned by the police, and there was no evidence submitted about day-to-day management of the data. Possibly nobody was managing it. On the contrary, some evidence says the TechCo database system crashed frequently —a problem the court never examined because the case collapsed before we needed to draw this issue to the court's attention.

To take any other view than "perfect IT" would imply considering the problems of managing the data and of questioning the reliability of TechCo systems. If the police suspected a problem, they would not have sent the glucometers to TechCo. I think the hospital, like the police, just assumed the data was perfectly reliable for a criminal investigation.

In a subsequent internal disciplinary hearing, the discredited police evidence was reused as if it was unproblematic. The disciplinary hearings presented a logical fallacy to support using it:

> "in an internal disciplinary proceedings the burden of proof is a lower threshold than in criminal proceedings …
>
> The investigation looked on [the database] to verify if this blood glucose had been taken for this patient. This is not verified on [the database] … [the hearing concludes that] patient did not have 9 blood sugar recordings checked [by this nurse] … [etc]"

Indeed, words like "this is not verified on [the database]" is a recurring phrase in the disciplinary hearing transcripts. To be clear: there is no reliable database evidence to verify any fabricated tests or any other poor procedures. We know there is a lot of corrupted data, and there is a simple reason it is corrupted.

I submitted written evidence to this disciplinary hearing, including a full explanation. I quoted from the judge's ruling:

> "Professor Thimbleby has shown that the chain has various breaks where the data can be lost. None of the data now relied on is original; it was all made after human intervention by [a TechCo employee] and he has no real recollection of what he was asked to do, what ID codes he was asked to consider, and did not note it at the time. All the material is at best edited. [The evidence] has lost significant amounts of data: but there is no way to tell whether the missing files were reintegrated into the [TechCo] database, in which case the Prosecution case might have force, or simply deleted, in which case it would not. I should exclude the evidence as being more prejudicial […] and unreliable hearsay. [It] would serve only to suggest to the jury a conclusion they could not draw – namely, that absence in the searches meant those results had never been in [the database] or the reject folder."

Since my arguments were ignored, I wonder what other influences persuaded the hospital it was appropriate to use invalid computer evidence even a judge (on the basis of professional analysis and weeks' of detailed, critical discussions and cross examination) had rejected as unreliable and misleading? It is worth adding that the experts from the prosecution also fully agreed on the ruling.

In hindsight, somebody in the hospital ought to have ensured at the time and subsequently checked whether the police acted appropriately on original, uncontaminated evidence. And after the court case collapsed, somebody at the hospital

should have checked why the case collapsed and made an informed decision whether disciplinary procedures should proceed on corrupt evidence.

At any stage, it would have been easy to compare a copy of the police evidence taken against a rolled-back hospital database. In fact, it seems bizarre the police did not take a signed and dated copy of the *actual time-stamped database* ***and*** *check it*, rather than just take what happened to be there after it had been corrupted and not check it. Unfortunately, a TechCo "expert" performed the transfer of data from the hospital to the police, probably reinforcing the naïve impression of infallibility.

5 Some recommendations

5.1 What should hospitals do?

5.1.1 A hospital collected data it did not monitor, so deleting or otherwise tampering with clinical data was not detected when it happened. Basic cybersecurity should have signaled TechCo's tampering (or any other unexpected changes) as and when it happened.

5.1.2 A hospital procured equipment that unreliably recorded clinical procedures. The hospital should monitor operational data, and take steps to correct or manage problems when they occur. In this case, the hospital ignored the data it was collecting until the police seized it, and then it was too late.

5.1.3 A hospital should procure more dependable equipment and systems.

5.1.4 Hospitals should disable all IT features it is not using or not monitoring.

5.1.5 A hospital should not have clinical systems that are not designed for clinical use, nor should police use such systems except with proper caution. See figure 5.

5.1.6 Hospitals should engage regular external oversight to help avoid blindspots. IT is very complex, and it may be impossible to recognize one's own misunderstandings of it without external input. Note that the manufacturers are not the right people for any such oversight.

5.1.7 When police request data from a hospital, proper governance procedures should be adhered to. One of the problems with the case here is that the police obtained evidence without a court order, and the court had no idea what evidence the police had — hence the surprise of new evidence presented in court without any warning for the defence to interpret it. This wasted much court time.

5.1.8 Hospitals, police and courts should realize that the research literature does not tell the whole story of whether equipment is appropriate for clinical use. Clinical research papers focus on a very narrow aspect of dependability (e.g., whether measurements are clinically accurate), but real use is much more

complicated (e.g., whether measurements get processed reliably in multi-vendor systems).

5.1.9 "Data fishing" is a serious problem. After the court case collapsed the hospital resumed disciplinary proceedings that had been suspended because of the criminal investigations. Using the police evidence the hospital argued that nurses had failed to take appropriate care of patients, a conclusion drawn from the discredited computer evidence. But one can fish data (especially bad data!) to support almost any case. One type of lack of care the nurse was accused of *also* happened over 1,000 times and affected all nurses on the same ward over the period I have data for, though note that I do not have the patient data to match against sliding scales of insulin. (The corruption of data may have *created* the "evidence" for this alleged incident and may also have inflated the 1,000 similar cases I can find.) Given the high frequency of such incidents, what this one nurse was accused is routine practice. I suspect the disciplinary hearing had no idea of such widespread practice, or that they were using discredited data. It is worth saying that if this lack of care was serious, then TechCo's system should have been configured to detect it *as and when it happened*; but TechCo's system was not used for any auditing (so far as I know). I infer the hospital did not worry what nurses did until it became a public issue.

5.1.10 Hospitals should routinely and regularly disclose to staff what data they are collecting and they should allow staff to see and, if necessary, to challenge it and the processes used to collect it. This means releasing data perhaps weekly if not daily. If data has no clinical role, then it should never be allowed to be used as if it had or might have done (which is what happened in this case).

5.1.11 Evidence discredited in the rigor of a criminal trial (and, in this case, very well summarized by the judge, see section 4.2) should be used to help seek the causes of the system failures, and should only be used with informed caution for disciplinary purposes.

5.1.12 Cybersecurity, improvement and culture change is at least a full time job. Avoiding predictable future problems and catastrophes will require dedicated staff and investment. Mature cybersecurity cannot be done alone, but requires collaboration across the healthcare sector and beyond — it should be a national priority, and local leaders need to be networking in this wider community just to stay up to date. This paper only discussed "simple" point of care equipment, but cybersecurity necessarily covers everything from wall-mounted emergency equipment, implants, medical apps, linear accelerators to PCs, all susceptible to hacking, ransomware, trojans and viruses — as well as all staff education (personal apps, phishing risks, etc).

5.1.13 Hospitals should also consider recommendations in all other sections in this paper (and elsewhere); hospitals should not work alone and be unaware of the activities and concerns of the wider community working to help improve IT and cybersecurity.

5.1.14 Best practice is, of course, that cybersecurity policies and implementations must be externally reviewed.

5.2 What should manufacturers do?

5.2.1 The very public discovery of VW's fraudulent IT to help their cars pass emission tests (Hotten, 2015) — which became public during the trial — should serve as a powerful reminder that IT is not just unreliable, but that it may be unreliable intentionally. VW's illegal emission levels are estimated to have contributed to tens of excess premature deaths. Manufacturers should adopt open source methods so that their code can be externally vetted.

5.2.2 TechCo testified that their equipment was CE marked, and therefore any problems *must* be the nurses' fault. With regulatory cover stories like that, there is little incentive for manufacturers to try harder. Manufacturers of clinical products should closely consider the quality of their programming, and have much better arguments for courts than that they have CE marks.

5.2.3 Hospital IT systems are very complex (through no particular fault of TechCo) and this complexity is not going to change any time soon. In view of the complexity, manufacturers must develop more defensive software — for instance with end-to-end checking, more logging and diagnostics, and with formal proofs of correctness (see section 5.5). Auditing needs to be provided, work and be used.

5.2.4 Installed software in hospitals that is not being actively managed (as happened here with TechCo's database) should be automatically reported, at least to the manufacturer who can then take remedial steps (such as reconfiguring it or notifying the hospital to audit it).

5.2.5 It was disappointing that no technical experts from TechCo wanted to appear in court, although they had provided much written evidence (though only for the prosecution it must be said). TechCo is based in a country outside of the UK's jurisdiction. Manufacturers should be eager to support investigations concerning their products.

5.2.6 Elsewhere I have written about the user interface of TechCo's systems [which I will not cite here to preserve anonymity]; the poor design of the user interface suggests that reliable operation was not a priority or perhaps not a competency for the manufacturer. The evidence presented in court further suggested the programming of the device and the database management software was substandard too —TechCo's written evidence says things like "it is not possible to say categorically that there were no corrupt records"; and, symptomatically, the TechCo database systems regularly crashed. Given the serious consequences of poor quality systems (patient harm, substandard care, pressure on staff — even prison) manufacturers should feel an obligation to put high quality professional effort into their products.

5.2.7 TechCo's systems have an "audit" feature. I beg to disagree; it certainly has a feature *called* audit, but it is manual and fallible, and generates documents that are not authenticated. Features should be named and implemented to support the conclusions most people (and courts) would reasonably draw from their names.

5.2.8 It may seem unfair to criticize manufacturers without offering any solutions. One easy thing to do, then, would be for all blood glucose tests (in fact, tests data from any equipment) to be assigned a serial number. Along with the glucometer ID, it would then be trivial to detect lost data (additionally, using digital signatures to circumvent cybersecurity problems). Lost data should be reported to the manufacturers post-market surveillance team. It would then be easy for a hospital to use best efforts to respond and recover it.

5.3 What should regulators do?

The CE marking system is discredited (Cohen, 2012; Cohen & Billingsley 2011), and fixing it to be more effective for complex computer-based systems (devices, medical apps, etc) is essential, particularly as computer-based technology is ubiquitous and taking over healthcare. Healthcare IT systems (PC, tablet, embedded, point of care, etc) support patient care and it is therefore negligent if they are not developed using equivalent processes to the rigorous processes used in pharmaceutical development (Thimbleby *et al*, 2015).

There are rigorous process in pharma because we recognize that there may be side-effects and unknown variation in patients — analogous problems to IT and cybersecurity. Randomized controlled trials (RCTs) as used in pharma may not be essential for cybersecurity, but there are other methodologies such as formal methods where correctness is proved (see section 5.5). Such methods should be used, and should be shown to be used in any certified product. Formal methods are routine in aviation (where lives depend on them); they ought to be routine in healthcare too.

Manufacturers will complain that they do not know how to use formal methods for their complex products. Well, they should start making products that are simple enough for them to understand. Logically, if the manufacturers can only "hope" to understand their own devices (e.g., see quote above; and failing to use formal methods means relying on hope alone), then hospitals and nurses are very unlikely to understand them either.

The US FDA device regulation is only concerned with the patient; staff well-being and effectiveness clearly ought to be a consideration too. In Europe and the UK (whether or not their regulation is normalized) is much more closed, obscure and informal (e.g., the role of notified bodies); in our complex world of IT threats, regulation urgently needs opening up and becoming more responsive.

5.4 What should the police do?

5.4.1 Although not discussed in detail in this paper, the police management of data exposed numerous problems. For example, one piece of evidence says that the police found that their Excel crashed analyzing the data. It is surprising the police even tried using Excel to analyze the complex relations in such a large volume of data.

5.4.2 Evidence I was given included analysis spreadsheets, which made me wonder what other edits the police had made to what was claimed to be evidence: there was no clear management of the evidence. Despite the "forensic" methods the police claimed to use, they were not available with the evidence and did not help confirm provenance. I was never given any evidence with signatures.

5.4.3 The police assumed that data (or missing data) could be used to prove poor clinical practice. This encouraged the media, public, patients and relatives to assume the trial was about poor clinical practice and reinforced a view that patients were victims of poor care. But whether or not the hospital had poor clinical practice, the criminal case was much narrower. Even if the data had been reliable (which it was not), the connection between abstract data and clinical practice would have been tenuous at best. Indeed, even if the case had been proved, there had been no patient harm and it could not have said anything about quality of care.

5.4.4 So far as I know, the police never considered assessing the internal or external validity of their evidence. A cursory look at the database would have raised many questions about it. Instead, I think they probably never took a holistic view. Accepting IT evidence on faith is problematic.

5.4.5 This was a case where alleged data discrepancies were used in evidence, so I was surprised at typos affecting data presented in the prosecution evidence — this is ironic, as the prosecution case relied on the quality of data (I am not saying the typos were sufficient in themselves to discredit the evidence). Of course I may have made some typos in my own evidence that we did not spot. In fact, I used *Mathematica* to analyze the data and automatically generate reports, tables and diagrams, etc: insofar as I can program reliably (and I more than double-checked every result), this ensured my reports were factually accurate.

5.4.6 The police seized several glucometers. Other glucometers had been used on the ward, so not all of the relevant glucometers were seized. Had the trial proceeded, this would have become a criticism of the prosecution case — it is possible that the alleged fabrications are *still* stored on a misplaced glucometer somewhere.

5.4.7 What was the ward supposed to do when they lost their glucometers to the police? I wonder whether the police did a risk analysis (there is no evidence either way): in any case, removing glucometers off a ward puts patients at serious risk of harm, at much greater risk than any of the alleged incidents.

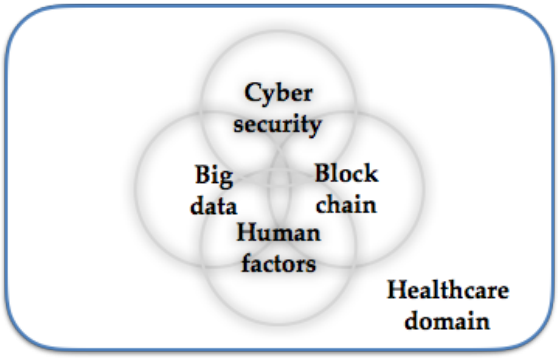

Figure 6. Convergence of powerful research challenges, particularly when applied to the complexity, risks and priorities of healthcare. Healthcare has many opportunities, it needs IT to drive quality improvement — and new technical challenges (cybersecurity, big data and blockchain, etc) need relating to *actual practice and human limitations* through applying human factors embedded in the healthcare domain. See section 5.5.

5.4.8 There were surprising conflicts of interest. A technician from TechCo corrupted data, then selected the data from the hospital and handed it over to the police, and the police sent the glucometers to TechCo themselves to analyze and confirm whether they were functioning correctly. Independent experts should have been used throughout.

5.4.9 As TechCo's glucometers do not use an open architecture, independent experts should have been *required* to be present when the data was taken from the hospital and when any glucometer analysis was performed.

5.4.10 Knowing glucometers function correctly when analyzed tells you very little about how they might have performed during alleged incidents. (For example, they may have been serviced since the incidents.) The glucometers were not designed to answer such questions. The police should exercise more caution with test results, especially when undertaken by the manufacturer.

5.4.11 It should be routine for manufacturers to be required to disclose relevant quality control documents and risk analyses in support of any claims that their products work to specification (e.g., as required by ISO 14971, ISO 13485, etc).

5.5 What should researchers do?

5.5.1 A serious issue remains for researchers, the industry and regulators to address is that clinical trials alone are insufficient to justify the quality of computer systems or devices in normal use. The current peer reviewed literature is inadequate. We need *both*: clinical research (do things measure the clinical factors they claim?) *and* situated IT and HCI research of effectiveness in the

real complexity of healthcare (will they be used correctly and is the data reliable?). Such research needs tying up "end to end": is the final data, however the clinicians summarize or interact with it, effective for clinical use and correctly based on true clinical data?

5.5.2 While cybersecurity research has a high profile, cybersecurity is only one aspect of the potential problems and vulnerabilities of medical devices and systems. More research is needed on end-to-end dependability, from HCI to networking and multi-vendor databases, interoperability, etc. Researchers should spearhead an analogous structure to the Information Sharing Analysis Organizations (ISAOs) established for cybersecurity.

5.5.3 Formal methods is a substantial research area that has resulted in many robust approaches to software development — widely used in aviation, for example. SPARK Ada is a good place for programmers to start (Barnes, 2003). Unfortunately, there is little connection between the formal methods community and healthcare, let alone healthcare IT. This gap urgently needs to be bridged. One of the many research problems is how to migrate large, complex, buggy software (as in blood glucometers and their networking, for example) into high quality software that works "well enough" — and increasingly more reliably — until it is rigorously correct.

5.5.4 Throughout this paper I have criticized the culture of assuming IT and data is perfect. In the UK, this culture is enshrined in law: the Criminal Justice Act 2003 created the presumption that IT works correctly (Mason, 2012; Mason, 2014). Computers are deemed to be "in order" and "properly set and calibrated." So we cannot simply blame the police or the hospitals when they just reflect the wider legal culture in which they operate, the absurd presumptions of the Act of Parliament being but a symptom. While Mason (*op cit*) gives a very professional discussion, including the problems of the inscrutability of proprietary systems (e.g., those that are not open source) and the imbalance between prosecution and defence scrutiny, the challenge to researchers is to create awareness and transformation of this absurd legal position.

5.5.5 Researchers always need resourcing, and the convergence of cybersecurity, healthcare risks and costs, big data and blockchain technology closely matches many national research funding priorities, to say nothing of digitalizing healthcare and increasingly relying on (unregulated? insecure?) apps. Human factors (e.g., human error) is often overlooked, for exactly the same reasons ("loss of situational awareness") that cybersecurity is overlooked by healthcare — people are too busy doing, urgent, hard complex jobs, and this distracts attention from longer-term, broader priorities that are not immediately visible. Figure 6 visualizes this opportunity.

6 Discussion

The big picture is that nobody seems to be fully aware of the complexity and risks of IT. This results in lax legislation, lax regulation and lax procurement, and in turn lax manufacturing since no useful standard of quality can be demanded by hospitals. Unawareness in turn results in lax management, and unnoticed inconsistencies between clinical care and its unreliable monitoring.

Prominent peer-reviewed papers such as Nichols (2011) "Blood glucose testing in the hospital: Error sources and risk management" reinforce the naïvety: "software ensures accurate documentation," "automation is the best prevention for errors," "smarter software is assisting with result documentation," etc. Even the recent UK *Making IT Work* report (Wachter, 2016) takes it for granted that computers work (with minor caveats on interoperability and usability) — hospitals just need to "digitalise" more, with another huge government investment (£4.2 billion) available to purchase IT. It takes it for granted there is appropriate IT that can be just purchased! The Wachter Report has a "digital maturity" scale, but it is about whether and to what extent a hospital has *adopted* levels of IT (and is paper free), not whether the IT is effective (and for what?) or even fit for purpose, which is just assumed. Every hospital is in good company, then, unwittingly "drifting into failure" (Dekker, 2011).

Ironically, if the hospital in the story here had been "paper free" (as the Wachter Report wants) there would have been no discrepancies to investigate; although there would have been no trial, the underlying IT problems would never have come to light. If we want a paper free health service, we also need to work out how to make it more reliable!

System failure only becomes apparent after there is a visible incident. In the case described here, something triggered the police investigation and that became "the incident," as discussed in this paper. In hospitals, reportable incidents usually involve patient harm or near misses of harm; in this case, thankfully, there was no patient harm but considerable staff harm. Nobody benefitted from the process.

With hindsight we can see many causes of the incident. All of them were avoidable; and avoiding only a few would have resulted in a much happier outcome. The wholly uncritical view of IT coupled with a remarkable unwillingness to consider alternative explanations for multiple IT problems related to tens of nurses are a textbook example of confirmation bias and cognitive dissonance (Tavris & Aronson, 2007); with so many nurses, underlying systemic factors, including management (Ball *et al*, 2013) and IT support (most of this paper) should have been obvious priorities to critically examine.

The causes, then, are many and complex, and other hospitals will have analogous but diverse complex IT problems. A key priority should be to have a mature cybersecurity strategy, which implies having an implementation of IT that permits having a workable strategy. Whatever the mess or causes of the mess — external hacking or just internal hacking, as here — effective procedures must be in place to detect, interpret and respond to unusual or unauthorized activity immediately. The regular

crashing of critical systems should at least have been a warning sign. Hospitals need to tighten up their cybersecurity maturity. Good guidance on data management is available elsewhere (DSIWG, 2016).

The hospital had ignored evidence of behavior the police treated as criminal; that is, the recording of tests on the database had no day-to-day significance (so far as I can tell). In any case, the process of getting data from a glucometer into the patient records was so slow and unreliable it was of no clinical use. Having correct and timely blood glucose measurements is a clinical priority but "correct" use of a point of care device and its IT system that was, unknown to anybody, not working well was not an issue, and certainly was not an issue for the hospital.

When data was deleted, the hospital did not notice. The police apparently assumed the corrupted data was perfect, and adequate evidence to charge nurses. There were many reasons that the data was unreliable, but the simplest was that TechCo themselves had corrupted large parts of it. That fact alone was sufficient for the trial to collapse, and it left the prosecution with no admissible evidence.

Once the police investigation started, my impression was that the hospital felt unable to pursue any parallel investigation, and certainly they felt unable to help the expert witness in his enquiries (in hindsight I should have sought a court order). This missed early opportunities to uncover some of the systemic problems. A parallel cybersecurity investigation, done within the hospital, would have saved a huge waste of time and huge costs to the defendants.

It is too easy to blame people at the sharp end. Nurses are at the sharp end, and a witch hunt makes a compelling story. A witch-hunt involves human things we feel we understand, and we feel angry about poor patient care and a sense of betrayal by nurses being incompetent (if they were). For everybody concerned, if you get rid of a witch the problem is immediately and visibly solved! Unfortunately, on the other hand, complex IT is hard to write a gripping story about; it really isn't very interesting, and there is no quick fix. Dekker (2009) discusses these important issues in much more detail than we can here.

7 Conclusions

This paper explored a case where unconsciously accepted unmanaged IT complexity unnecessarily led to horrible outcomes. While the hospital had a core part to play in this story, the real villain in my opinion is TechCo. In principle they could produce better IT if they wanted to; they have had many years developing hospital equipment and must know very well how to do it: their equipment could have been much easier to use correctly in a real ward, and much easier to use to *dependably* monitor patient data and ward activity. TechCo's systems failed the hospital, its staff and its patients.

Nevertheless, the data corruption this case revolved around (the nurses' alleged falsifications, the use of staff IDs, TechCo's deletion of data, and the poor software that allowed it all to happen…) could have been detected and managed as soon as

they occurred had the hospital had a mature approach to cybersecurity. Given that TechCo systems were used, the problems were unavoidable, but they could have been detected and managed. They need never have escalated to a criminal prosecution.

Given the widespread use of equally poor IT throughout healthcare, hoping for manufacturers to improve is less realistic than hoping for hospitals to prioritize mature cybersecurity. Fortunately, cybersecurity already has a high profile (thanks to fear of malicious hacking), and this paper adds one more reason to take it even more seriously. Luckily nobody was harmed as a result of any of the issues discussed in this paper, but a mature cybersecurity approach would also reduce unnoticed data problems of all sorts, and therefore would help reduce patient harm too.

The broader problem remains our culture of uncritical acceptance of IT, from legal, regulatory, procurement and other perspectives, especially for healthcare where billions is eagerly invested in more IT stuff. Our culture makes us all uncritically believe that IT and especially the *latest* IT is wonderful — don't we all want new things? (Of course, this is how companies stay in business.) But the reality is that behind the façade of superficial wonder, modern hospital IT is too complicated for its own good, for the good of patients, for the good of staff. Ironically, the newer IT is and the more exciting it seems, the less tested it is in the clinical environment.

This culture nurtures a lax approach to cybersecurity. It created the perfect environment (bad IT, bad IT management) for accepting a superficial explanation of alleged multiple nurse failures instead of exploring underlying causes.

Acknowledgements. EPSRC funded part of this work under grant [EP/L019272]. I am grateful to a huge number of anonymous people who helped with this paper and the background information and research. I am very grateful to the nurses who had the courage to defend their innocence.

References

Ball, J. E., Murrells, T., Rafferty, A. M., Morrow, E. & Griffiths, P. "Care left undone during nursing shifts: associations with workload and perceived quality of care," *British Medical Journal Quality & Safety,* **0**:1–10, doi:10.1136/bmjqs-2012-001767, 2013.

Barnes, J. *High Integrity Software: The SPARK Approach to Safety and Security*, Addison Wesley, 2003.

Cohen, D. "How a fake hip showed up failings in European device regulation," *BMJ*, **345**:e7090, doi: 10.1136/bmj.e7090, 2012.

Cohen, D. & Billingsley, M. "When it comes to medical devices, Europeans seem to get a worse deal than US patients," *British Medical Journal*, **342**:d2748 doi: 10.1136/bmj.d2748, 2011.

Davidson, J. "Cyberattacks on personal health records growing 'exponentially'," *Washington Post*, www.washingtonpost.com/news/powerpost/wp/2016/09/28/cyberattacks-on-personal-health-records-growing-exponentially, 28 September 2016.

Dekker, S. W. A. "Prosecuting professional mistake: Secondary victimization and a research agenda for criminology," *International Journal of Criminal Justice Sciences*, **4**(1):60–78, 2009.

Dekker, S. W. A. *Drift into failure: From hunting broken components to understanding complex systems*, CRC Press, 2011.

DSIWG, Data Safety Initiative Working Group, *Data safety guidance*, Safety Critical Systems Club, ISBN 978-1519533579, 2016. http://scsc.org.uk/paper_130/Data%20Safety%20(Version%201.3).pdf

Hotten, R. "Volkswagen: The scandal explained," BBC News, http://www.bbc.co.uk/news/business-34324772, 10 December 2015.

Koppel, R., Wetterneck, T., Telles, J. L., & Karsh, B-T. "Workarounds to barcode medication administration systems: Their Occurrences, Causes, and Threats to Patient Safety," *Journal of the American Medical Informatics Association*, **15**(4):408–423, 2008.

Mason, S. *Electronic Evidence*, Chapter 5, 3rd ed., Butterworths, 2012.

Mason, S. "Electronic evidence: A proposal to reform the presumption of reliability and hearsay," *Computer Law & Security Review*, **30**(1):80–84, doi: 10.1016/j.clsr.2013.12.005, 2014.

Nichols, J. H. "Blood glucose testing in the hospital: Error sources and risk management," *Journal of Diabetes Science and Technology*, **5**(1):173–177, 2011.

Tavris, C. & Aronson, E. *Mistakes were made (but not by me)*, Harcourt Inc: Florida, 2007.

Thimbleby, H., Lewis, A. & Williams, J. "Making healthcare safer by understanding, designing and buying better IT," *Clinical Medicine*; **15**(3):258–262, doi: 10.7861/clinmedicine.15-3-258, 2015.

Wachter, R. M. *Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England Report of the National Advisory Group on Health Information Technology in England*, Crown Copyright, 2016.