ARTICLE:

Misunderstanding IT: Hospital cybersecurity and IT problems reach the courts

By **Harold Thimbleby**

The corruption of patient data in a hospital prompted a criminal investigation, resulting in approximately 70 nurses being disciplined, with some charged with wilful neglect contrary to the Mental Capacity Act 2005. Some nurses received custodial sentences. This paper explains the background. The paper demonstrates the inability of hospital information technology (IT) systems and management to provide reliable evidence and highlights broad problems with poor IT culture affecting manufacturers, hospitals, police, lawyers, and advisors — all the way through to regulators and legislators. Widespread misunderstandings of IT and data compromises both the provision of effective care and legal processes.

This paper includes recommendations, the most urgent being that hospitals (the UK National Health System ('NHS') and other national healthcare systems more generally) should acknowledge that IT is unreliable, and that they should procure and actively manage IT equipment with this in mind. Keeping up-to-date with legal issues relating to IT generally, as well as keeping up-to-date with cybersecurity measures should be routine.

The NHS needs to improve its IT maturity, management and policies. The police, the legal system and regulators also need a more mature approach to IT. Manufacturers are not currently providing dependable systems that are fit for purpose to operate safely and reliably in normal, complex hospital environments. All parties should engage qualified external oversight.

Introduction

This paper summarizes my insights from being an expert witness in a criminal case involving alleged fabrication of patient data by nurses.[1] I am a professor of computer science and a research fellow in digital health; the expertise I brought to this case was computer science, but more so, a scientific refusal to accept computer evidence at face value without question. For example, most evidence in this case was presented as Excel spreadsheets — but it is well-known that it is easy to delete or tamper with data in Excel without leaving any trace.[2] The data in this case showed distinctive statistical patterns suggesting deleted or corrupted data, but nobody else questioned the evidence at that stage, despite IT failure being a well-known hazard[3] that is poorly regulated.[4]

The outcome and details of the court case are in the public domain, in particular, the judge's Ruling has been published in this journal[5] and the hospital has published an external review after the event.[6] However, the aim of this paper is not to tell a story about the hospital or its nurses, but rather to tell a more worrying story: *this could happen anywhere — and probably is happening everywhere*.

In view of understandable sensitivities, this paper minimises citations to the prosecution and related evidence. I have not changed technical details or operating procedures. Numbers have been rounded and ward names changed: the exact values and names

[1] *R v Cahill; R v Pugh* 14 October 2014, Crown Court at Cardiff, T20141094 and T20141061 before HHJ Crowther QC.

[2] See 'Business records' that illustrates this issue in detail: Stephen Mason and Daniel Seng, editors, *Electronic Evidence* (4th edition, Institute of Advanced Legal Studies for the SAS Humanities Digital Library, School of Advanced Study, University of London, 2017), xii–xiii.

[3] ECRI Institute, *Top 10 Health Technology Hazards for 2015*, (No. 2. Data Integrity: Incorrect or Missing Data in EHRs and Other Health IT Systems), 2014, available at www.ecri.org/Documents/White_papers/Top_10_2015.pdf .

[4] C. Heneghan, M. Thompson, M. Billingsley and D. Cohen, 'Medical-device recalls in the UK and the device-regulation process: Retrospective review of safety notices and alerts', *BMJ Open*, 1:e000155, 2011. DOI: 10.1136/bmjopen-2011-000155, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3191575/ .

[5] Ruling in *R v Cahill; R v Pugh* 14 October 2014, Crown Court at Cardiff, T20141094 and T20141061 before HHJ Crowther QC, 14 *Digital Evidence and Electronic Signature Law Review* (2017) 67–71.

[6] Angela Hopkins, *Commissioned Review, June to September 2016 Review of the Blood Glucometry Investigations in Abertawe Bro Morgannwg University Health Board Establishing lessons learned* (Abertawe Bro Morgannwg University Health Board, 2017), available at http://www.wales.nhs.uk/sitesplus/documents/863/4.5%20Blood%20Glucometry.pdf .

are not relevant to the story. The judge's Ruling identifies the equipment (blood glucometers) which were Abbott XceedPro meters, made by Abbott Laboratories, Inc. (Abbott); Abbott was also responsible for the database, called Precision Web, which stored data from the meters. However, this manufacturer is not unusual: their products are of typical quality and design.

The judge excluded some evidence from the jury as being prejudicial under s 78 of the Police and Criminal Evidence Act 1984. Evidence was derived from flawed data, flawed IT, and flawed management of IT. Nurses were blamed for these failures. As this paper explains, the underlying causes must be properly understood as basic software issues that should have been taken seriously as and when they originally happened. Indeed, it is baffling that the hospital failed to detect *and address* the corruption of the databases that had been in continual clinical use over a period of years.

It is to be speculated as to how many other cases — internally or reaching the courts — inappropriately blame and pursue clinicians caught up in fallout from hospital IT chaos, with nobody recognizing or wanting to admit or check whether IT can cause the problems. Poor IT, and poorly managed IT, can induce clinical[7] and other error, contribute to error, exacerbate error, cause huge costs,[8] and make it hard to disentangle true causes.

Cybersecurity maturity

Cybersecurity is the protection of IT systems (including devices with embedded computers, such as glucometers and MRI scanners, etc) and data from the theft, damage, and disruption. Cybersecurity is a high-profile activity, regularly reported in the news media, often working against international malicious attackers (including state security services with considerable expertise) who seek vulnerabilities to exploit, perhaps for blackmail or terrorism, but cybersecurity incidents also occur frequently from insider attacks. Some internal users may be motivated by retribution, but many users are well-meaning and unintentionally perform unsafe or insecure operations that compromise systems. In particular, external hackers may deliberately mislead well-intentioned internal users with phishing, which tricks them into compromising their systems.

Mature IT management means staying abreast of the latest methods used by hackers, keeping track of system weaknesses (vulnerabilities) as they are discovered, and using reliable means to protecting systems against these problems. Protection can be based on software or physical means, such as conventional physical security (locked doors, access restrictions, etc) and network firewalls that isolate the IT systems from the internet and hackers operating anywhere in the world. Keeping software up to date is also essential, as hackers often exploit old errors in systems.

Secondly, management should detect problems and disruption, and have in place procedures for limiting and recovering from damage. Systems may be taken offline to isolate them from attacks, and IT data and services may be recovered by restoring lost or corrupted data from secure backups. Note that cybersecurity requires a long-term plan; it may be impractical to recover after some cyber problems if adequate measures (such as regular backup and recovery protocols) were not developed and put into practice long before the incident. Cyberattacks often have a reputational impact, and organisations need to manage media relations.

Sometimes IT problems occur during software upgrades, but the consequences can be similar. Fortunately, effective standard operating procedures for cybersecurity will detect problems, recover lost work and return the systems to an operational level regardless of how they occur. Likewise, good cybersecurity protects against problems caused by hardware failure, fire, electricity outages and many other risks.

Most of the time IT systems work very well, so it is possible to become complacent and ignore cybersecurity and the inherent problems with software errors, and also ignore the necessary investment in staff and training. Cybersecurity maturity assesses how well an organisation and individuals in it understand and effectively manage their cybersecurity. An individual might just backup a few files on their laptop in case it is stolen or breaks, but an organisation typically has a complex network and has many systems and staff to manage; their

[7] Harold Thimbleby, Alexis Lewis and John Williams, 'Making healthcare safer by understanding, designing and buying better IT,' *Clinical Medicine*, 15(3):258–262, 2015. DOI: 10.7861/clinmedicine.15-3-258, available at http://www.clinmed.rcpjournal.org/content/15/3/258.full.pdf+html .

[8] NHS trust £14m debt partly "due to patient system failure"', *BBC News*, 9 May 2018, available at http://www.bbc.co.uk/news/uk-england-gloucestershire-44052432 .

cybersecurity procedures should include auditing and staff training, as well as sophisticated off-site methods for reliable backup and data recovery. Many organisations will integrate cybersecurity management with compliance with the General Data Protection Regulation (GDPR)[9] and other data protection laws. In the NHS there are additional concerns for protecting patient data and privacy that should also be protected by cybersecurity systems.

In the case described in this paper, electronic patient data was corrupted by an authorised but misguided internal act. The deletion and corruption of data was not detected, although it created discrepancies with paper records made by nurses. The hospital and police, apparently unaware of the standard issues relating to software errors and cybersecurity, assumed the nurses had fabricated the paper records. People were unaware of the cybersecurity breach and its significance until several weeks into the court case.

High level view

Important electronic evidence was declared prejudicial by the judge, which meant the prosecution offered no evidence, and the two nurses who entered pleas of not guilty were released.[10] The problems with the data, the final understanding of which led to the evidence being declared prejudicial by the trial judge, should have been detected when the data was corrupted. This internal oversight, together with the risk of cyberattacks that may be overlooked, illustrates the scope and reach of IT and problems relating to cybersecurity and the failure to understand that software and electronic evidence is prone to error. It is alarming to think — as the case described in the paper shows — that corrupted data can remain unnoticed in hospitals for years before action is taken, and even then it is misinterpreted.

WannaCry[11] is an example of a well-known cyberattack that made data unusable. With WannaCry, patient data was just encrypted (that is, made obviously unusable, and effectively deleted), but future cyberattacks will certainly maliciously corrupt data (for instance, by changing patient blood groups). Such attacks would have disastrous consequences. Healthcare cyberattacks are 'growing exponentially'[12] — with 113 million US electronic health records breached in 2015.

Much good guidance is available[13] (see also the list of recommendations at the end of this paper) — although cybersecurity is continually advancing, and up-to-date, competent professional oversight is required. For many organizations, immediate external review should be a priority; clearly, whatever cybersecurity procedures the police, the Crown Prosecution Service lawyers and hospital were using, they failed *and they did not know they had failed*.

This story is about the widespread misunderstanding of IT in healthcare, and in particular, about mismanagement of IT by hospitals, the police and the lawyers. This cannot be a single one-off example. There must be many other cases where clinicians have been blamed for hospital IT chaos, where nobody recognizes or admits, or checks that IT can be the cause of such problems. Hospitals spend a great deal of money on IT, and they have IT staff to manage it; there is a significant financial and personal investment in this being right, so it is psychologically very satisfying to blame the users (nurses in this case).[14] Indeed, IT is complicated, and it is very difficult to question it — we live in a technological world that relies on us all uncritically buying more and more IT (e.g., newer mobile telephones and tablets). Manufacturers naturally feed our belief that IT solves problems, overlooking that IT also causes some of them.

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, 4.5.2016, p. 1–88.

[10] 'Nurses cleared of wilful neglect at Princess of Wales Hospital in Bridgend,' *South Wales Evening Post*, 14 October 2015, http://www.southwales-eveningpost.co.uk/nurses-cleared-wilful-neglect-princess-wales/story-27983645-detail/story.html ; 'Princess of Wales Hospital nurse neglect trial collapses,' *BBC News*, 14 October 2015, http://www.bbc.co.uk/news/uk-wales-south-east-wales-34527845.

[11] https://en.wikipedia.org/wiki/WannaCry_ransomware_attack .

[12] Joe Davidson, 'Cyberattacks on personal health records growing 'exponentially',' *Washington Post*, 28 September 2016, https://www.washingtonpost.com/news/powerpost/wp/2016/09/28/cyberattacks-on-personal-health-records-growing-exponentially/?noredirect=on&utm_term=.3ddf7519fbe3 .

[13] Kevin Fu, Harold Thimbleby, Juuso Leinonen and Ben Ransford, 'Six Factors Essential for Mitigating Cyber Risks in Healthcare,' *Association for the Advancement of Medical Instrumentation*, 2017. Available online at www.aami.org/productspublications/articledetail.aspx?ItemNumber=5472 ; DSIWG, Data Safety Initiative Working Group, *Data safety guidance*, Safety Critical Systems Club, version 2.0, https://scsc.uk/r127B:1 .

[14] Carol Tavris and Elliot Aronson, *Mistakes Were Made (But Not by Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts* (Harcourt Inc: Florida, 2007).

The 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, and some acts that could be deemed to be criminal, even though there is no patient harm. Indeed, in this case there was no patient harm or any poor care.

The news media reported stories of many — about 70 — nurses being investigated. There was considerable political and public interest in this case. That the hospital suspended many nurses (about 70), with some being indicted, seemed to confirm that there were problems similar to the scandal at the hospital run by the Mid Staffordshire NHS Foundation Trust, discussed in the Mid Staffordshire NHS Foundation Trust Public Inquiry ('The Francis Report').[15] At Mid Staffs, perhaps 1,200 patients died as a result of poor care between 2005 and 2009.

The 70 nurses were alleged to have fabricated blood glucose readings (that is, it was alleged the nurses did not actually take any readings from patients) by writing up fabricated readings in paper patient notes. However, if the nurses had actually taken blood glucose readings, the Xceed Pro glucometers they used should have recorded the test results in the Abbott Precision Web database. For some nurses, corresponding data was missing from the database, and it was therefore presumed the glucometers had never been used and that the paper notes must therefore be fraudulent.

For vulnerable adult patients this would constitute a criminal offence. The nurses were charged for wilful neglect, contrary to s 44 of the Mental Capacity Act 2005. The implication was that the nurses were lazy and dishonest and had put patients at risk. Three nurses entered pleas of guilty, and two entered pleas not guilty. Their case proceeded to trial.

Abbott's Xceed Pro blood glucometers were used in the hospital, and they are intended to automatically upload glucose readings to a central patient record system called Precision Web, also made by Abbott. The police established that this central record system had no records of many tests the nurses had written on patient paper notes. The police concluded that the nurses had written down fictitious readings and had not bothered to do their job properly.

The police considered various ways the nurses may have made accidental errors, such as writing down numbers close to but not identical to the glucometer's actual results, a possibility that they discounted. The police compared paper records with the computer database, involving around 150,000 test records — a great deal of combined manual and computer work. Although the police considered whether nurses had made innocent errors, they never explored whether the IT systems or the glucometers may have made errors — given that the IT systems and glucometers are programmed and managed by ordinary, fallible humans, and computer errors are notorious, this was a significant oversight in the investigation.[16]

In addition to identifying the alleged fabrication (that is, paper records with no corresponding computer records), the police found evidence of many cases of poor operating practice. For example, a nurse is supposed to enter the patient's identity number (ID) into the glucometer, but sometimes a nurse will scan a 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 was clear there were many cases of this practice. (The next section, below, describes in more detail what nurses are supposed to do.)

The nurses that were the subject of the criminal investigation had followed such bad practice repeatedly.

The prosecution case was 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 been reported with Abbott's systems. Therefore, it was suggested, the bad practice was a nursing problem, not a system problem.

Eventually, as this paper describes, the case collapsed. At the end of the trial, television crews 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 entered pleas of guilty were later sentenced, some to community service and some to prison.

[15] The Mid Staffordshire NHS Foundation Trust Public Inquiry, Chaired by Robert Francis QC (February 2013), https://www.gov.uk/government/publications/report-of-the-mid-staffordshire-nhs-foundation-trust-public-inquiry .

[16] For which see chapter 6 in *Electronic Evidence*.

What 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 and to intervene quickly if a patient requires treatment. The body normally regulates blood glucose (blood sugar) levels closely, but in diabetics, the body's normal regulation fails and levels have to be corrected by insulin (if too high) or by sugar (if too low). Both high and low levels of blood glucose cause serious medical problems, including confusion and unconsciousness. In everyday life, diabetics are usually very competent at managing their blood glucose levels, but hospital patients generally need help, particularly if they have mobility problems, dementia or other cognitive problems.[17]

Using a glucometer, an outline of the typical nurse operating procedure is as follows:

> (i) Find a glucometer (typically they are stored in a cabinet).
>
> (ii) The nurse then identifies themselves to the device by scanning the barcode on their staff card or by typing their ID on the glucometer keyboard.
>
> (iii) The patient ID is scanned from their barcode or typed, perhaps copying from patient notes.
>
> (v) The patient's finger is cleaned and pricked, and a drop of blood placed on a special test strip.
>
> (vi) The test strip is inserted in the glucometer.
>
> (vii) The glucometer displays the blood glucose level (or possibly an error, such as there being too little blood for it to work).
>
> (viii) The nurse may then take immediate action to address any clinical issues.
>
> (ix) The nurse then writes down on the paper patient notes the time and reading.
>
> (x) One further step, which has no immediate clinical significance, is that the glucometer must eventually be placed in a dock (possibly after taking many readings), and then its data will be automatically uploaded to central database systems.

The above procedures are intended to give the general idea; in a clinical setting, there are other steps (e.g., the nurse should wear gloves) which are omitted here as they are not relevant to this paper.

The Abbott glucometer itself can record over 2,000 readings in its internal memory before it needs to be docked, and it will display a warning if the memory is full and, if so, it cannot be used further. The memory feature means that Abbott's glucometer can be used for batching tests: a nurse can test each patient in turn on a whole ward, respond to patients' needs, then later write up all the results using the glucometer review screen to recollect individual test details. Once successfully uploaded to Abbott's systems, the test readings should later appear in the main patient records available on ward PCs — however, this might take days or longer (see below).

Workarounds

Barcode workarounds are a well-known problem.[18] It can be hard to scan or read the patient ID, so one workaround is to type 000 etc on the glucometer keyboard, or much more easily, scan a staff ID barcode. Both workarounds obtain a number that the glucometer accepts as a valid 'patient ID' so the nurse can then use the glucometer to get a reading. Using the staff ID instead of the patient ID is called 'double tapping.' Double tapping was routine: there were about 700 double taps done with 200 different nurse IDs, accounting for 25 per cent of all nurse IDs in the database. Double tapping was to be used as evidence of bad character, within the meaning of section 98 of the Criminal Justice Act 2003.

The glucometer keyboard is like that of a telephone: you can type anything, for instance pressing 2 can also enter A, B or C. There are cases in the database where, instead of a patient ID (or even a nurse ID), the patient's name was entered (which is of course as slow and tedious as it would be to type on a telephone digit keyboard), and sometimes the patient

[17] *Type 2 diabetes in adults: management*, UK National Institute for Health and Care Excellence (NICE, 2015), available at https://www.nice.org.uk/guidance/ng28 .

[18] Ross Koppel, Tosha Wetterneck, Joel Leon Telles and Ben-Tzion Karsh, '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, available at https://www.ncbi.nlm.nih.gov/pubmed/18436903 .

ID was entered as MALE or FEMALE. This suggests some nurses were not well trained to use the glucometer. These are basic errors recorded over several years in the database that nobody in the hospital was picking up or correcting.

The Abbott glucometer accepts all of these workarounds (arbitrary IDs, double tapping, text such as MALE) and will still give a correct blood glucose reading. Unknown to the nurses, however, the hospital IT administrators had configured the system to reject such data, and store it separately. Manual intervention (which may never happen, or may introduce further errors) is then required 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.

Double tapping and other workarounds, since they can be automatically detected, arguably should have been detected and sorted out immediately rather than ignored. The devices can be configured to detect these problems immediately they occur. Since double tapping was so prevalent and not reported, it is probable that nurses were not aware it was problematic, nor that it was being recorded; it was what everybody did. Since data at this hospital regularly got lost, this fact would further confirm to nurses the irrelevance of docking the glucometer.

An expert witness perspective

I was invited by the defence team to be an expert witness. My first comment was that if so many nurses were alleged to have made the same mistakes, it would surely seem more likely there may be a common explanation, such as an IT failure that would affect everybody, or there may have been a cyberattack or some other IT problem. An IT technician with a vendetta could have done anything to the data. The prosecution view, however, was that if many nurses are making the same mistakes, they were doing so deliberately and, moreover, were all in it together. I wondered why the hospital had not intervened earlier (Abbott's system could have been configured to warn them) as the data covered a long period spanning over years. The nurses had apparently developed bad habits, perhaps learning it off each other. It seemed plausible to me that the hospital had not properly managed or monitored the data until the investigation and prosecution started.

I do not know, but it is possible that the hospital suffered from what is called 'baseline shifting.' At first, one nurse was suspected of fraudulent record keeping, and that would seem plausible, albeit sad. So, one nurse is now the baseline. Later, another nurse is suspected of fraudulent behaviour. Now two is the baseline: it has only increased by one. A second suspect nurse is plausible because it is only one more than the baseline, which has already been accepted. One by one, more nurses are identified. These discoveries then fit a pattern. After a point, nobody is surprised if another nurse is suspected. Had all the 70 or so nurses been suspected on the first day, perhaps alternative explanations might have been urgently considered, but once 69 nurses have been suspected, there is no surprise with the 70th.

My first task was to analyse the prosecution evidence (presented as a CD of Excel spreadsheets and, later, as data logged on blood glucometers as well as SQL and XML files) to see if the police had made any mistakes in claiming that the glucometer 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 the nurses had made minor transcription errors.

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. 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, the normal operation of Abbott's system requires administrators to make changes to data, for instance to sort out double tapping. What else could administrators or even hacking do?

That over 20 per cent of the database entries had an 'error flag' set raised my suspicions; this was becoming a more complex story than the prosecution painted.

Another worry was that a comment field on each test said 'Wrong patient' for just two entries, and *nothing at all* for the remaining hundreds of thousands of entries. This suggested to me that nobody was paying

much attention to the management of the database; indeed, the 'reviewed' flag was 'No' for all but two of the data entries. If it had been 'No' for all of them, that could be explained if the hospital did not use the field at all, but as it has been set a few times, it suggests poor data management.

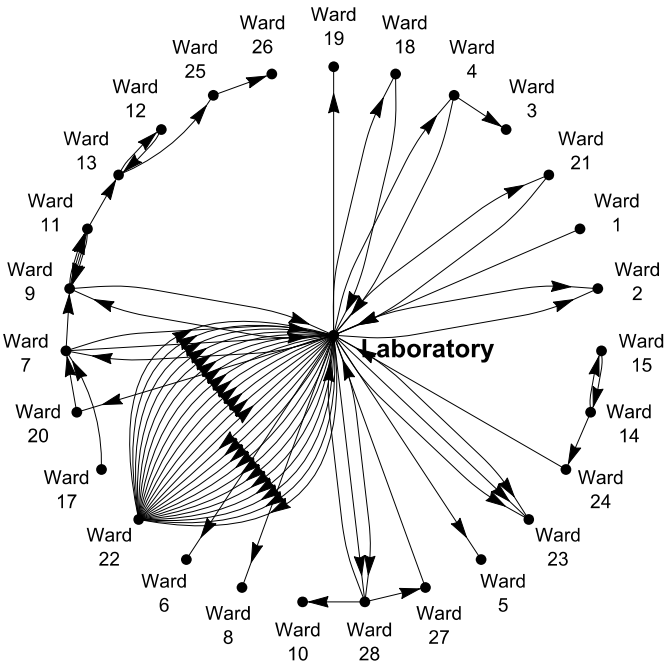


Figure 1. Recorded movement of glucometer dockings around the hospital over a period of one year. Note the centrality of the laboratory as a hub of movement, and that Ward 22 seems to have a lot of activity — 25 movements. Since wards presumably try to maintain a constant stock of glucometers, there 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' are not strictly wards at all.)

I noted that staff names occurred with many implausibly close variant spellings (e.g., differing in capitalization or spacing, or close variants like Jon and John, Justina and Justyna, Kimberley and Kimberly, or Diana and Diane *with the same surname*), sometimes but not always with the same ID. Also, many identical staff names occurred with different numeric IDs and some occurred in variants like '*name*' as well as '*name* (bank nurse)' with a separate ID. All this, and more, suggested the database was not well-managed and might not be reliable. Moreover, the poor staff data suggests that the Abbott features for only permitting authorized staff to use the glucometer might have been compromised (for instance, it is questionable as to how the device could reliably lock out unauthorized staff when the staff data itself was so poor).

Everyone took it for granted that IDs correctly identify nurses. If nurses share their ID cards (knowingly or unknowingly) then the ID recorded by the glucometer is the ID of the card, *not* of the nurse using the glucometer. Evidence is required to show that there are no duplicate ID cards and that nurses never use other ID cards. If a nurse has mislaid their ID card, as it is clinically important to take glucometer readings, it seems likely they may borrow another card just to get their job done.

No evidence was presented that the ID data was plausible. On the contrary, the database shows that 10 nurse names are recorded as having more than one ID each (whether the named nurses use these IDs themselves or whether the duplicate IDs are more widely shared is impossible to say from the database alone); in fact, the database has 80 more IDs than nurse names. There is a staff ID management problem that in my view undermines the credibility of the electronic evidence.

Each blood glucose reading in the database was associated with 35 other items of data: the patient name, number and gender, the staff identifier, the date, the ward, the temperature, the battery voltage, and so on. I suspect that the police only looked at the data they thought they directly needed for their investigation; it seems that they did not examine fields such as the error flags and alarms, or wonder why staff names were misspelt. They seemed to have no curiosity in the quality of the data.

Unfortunately for the police, proving absent data is absent for a reason (e.g., fraud) is hard when the provenance and quality of the data is in question. Moreover, Excel spreadsheets support no way to audit them: it is impossible to tell whether cells, rows or columns have been deleted, been edited, or even have never existed.

The police claimed they had used forensic methods to make copies of the database used in evidence. In fact, some evidence provided to me in several Comma Separated Value (CSV) files (an Excel data format) proved there had been manual intervention — Excel files also contained bits of police analysis along with the data. Because of their flexibility and lack of auditing, Excel and CSV are not the place to start if

you want to assure authenticity and integrity of evidence.

In addition, the Abbott database itself was in a Structure Query Language (SQL) database, so somebody had converted it to Excel, and this is a manual process. Some of the Excel worksheets had differences that suggested an unreliable process had been used to create or edit them; certainly, they showed that the 'forensic' process had failed.

The police had obtained and copied Excel spreadsheets at the hospital onto a Universal Serial Bus ('USB') external drive, (USB stick), but only then digitally signed the data and held it securely. This 'forensic rigour' came too late: the police should have made a signed copy of the original database, not a manually created dump in CSV put on a USB stick. Anything could have happened to contaminate the Excel data earlier than the forensic process. There is no way to tell in Excel whether rows or columns been deleted or edited, and this could have happened before the data was put on the USB stick at the hospital or even by the police editing what was on the USB stick before it was securely copied.

In addition, 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 the database showed 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 glucometers on the ward were seized. Note that the Abbott database records where glucometers are docked, not where the blood glucose tests are made — to know that, the test data needs to be related to correct patient and ward data.

As figure 1 summarises (from the original data), glucometers move around the hospital. If a glucometer has a fault (e.g., a dead battery), it would be returned for servicing and replaced by another, and after finishing servicing it could be sent back anywhere in the hospital. If a nurse needed a glucometer but could not find one, they might borrow one from another ward. Glucometers may also get lost, perhaps at the back of a cupboard or sent off for repair.

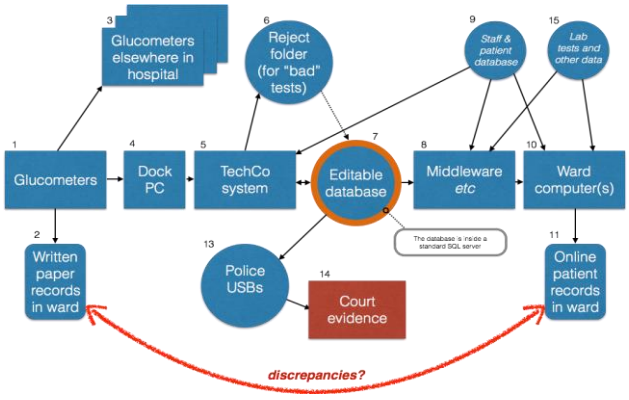


Figure 2. Diagram presented in court (anonymized) originally sized as A4; the smaller reproduction here serves to indicate the complexity of the network, but note that Box 8 in the figure contains unknown further software. The basic problem was that there were discrepancies between the paper records (bottom left) and the final computer records (bottom right). Note that not all relevant systems are Abbott'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 sent the glucometers they seized to Abbott to confirm that they worked correctly and to get copies of their data. Abbott said they worked correctly, though in my professional opinion their evidence was inconclusive on whether the glucometers had worked correctly at the material time. In the Excel data, glucometers were recorded as used almost hourly during the day on the ward, but some glucometers in the hospital were not used at all for over 100 days and many were not used at all for over a week. Nobody (so far as I could tell) knew where they were, and it is apparent that nobody used the database to locate where all the glucometers were that had ever been used on the ward so they could be correctly seized when the police attended the hospital. I do not think the hospital kept an inventory that tracked where glucometers were: if it had, it should have formed an essential part of the evidence.

Since glucometers move around, and some of the relevant glucometers were not seized, it is possible that the missing data is still on various glucometers lost somewhere in the hospital (and which had not been docked before the database was copied by the police). They may still be sitting in a cabinet somewhere.

One file in eXtensible Markup Language (XML), disclosed by the prosecution in the second week of the trial showed a four year gap between a test measurement being taken and the data successfully transferred to the database. The alleged incidents happened less than four years before the trial, so this is suggestive that the missing data might have been still be on its way through the system — the alleged incidents did not happen that long before the trial.

In my view, there was internal evidence in the database (e.g., the many error flags) that the data was of low quality, and there was the problem that there was no reliable forensic route from the original SQL database (let alone from the contents of the glucometer memories) to the Excel worksheets presented as evidence. By Abbott's design, there was no reliable connection between the glucometers and the main database itself (e.g., there was no handshaking — the glucometers do not wait for the database to confirm receipt of data). If the Abbott Precision Web database never gets data the glucometer transmits, it seems nothing in the system notices and tries to rectify or warn about the problem.

There were many failure points, for example — this list is not exhaustive (see also figures 2 and 3):

(i) A glucometer may lose data.

(ii) A glucometer may not be docked (it may not have been docked before the police seize data).

(iii) A glucometer may be physically mislaid or returned for repair.

(iv) 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.

(v) The glucometer battery may go flat.

(vi) Abbott's glucometers only store about 2,000 readings, yet the database shows some were used for nearly 5,000 tests. Hence the data on the glucometers when they were sent to Abbott for checking says nothing about the tests more than 2,000 ago: they cannot provide evidence on what nurses did earlier than the glucometer records cover.

(vii) The glucometers all look the same (and have tiny serial numbers), so you may just be looking at the wrong glucometer. I talked to some nurses and they did not realise the glucometers were different and that it could matter which ones were used.

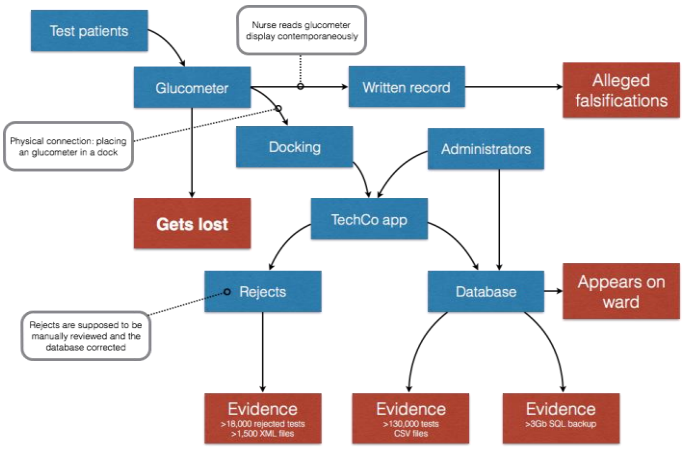


Figure 3. Diagram presented in court (anonymized) to show key sources of evidence. It was originally sized as A4 — for the present paper the details are less important than noting the complexity, in particular with no end-to-end checking of integrity. The police made digitally signed copies of Evidence 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 Abbott's modifications of data, which only became apparent after the diagram was submitted to the court.

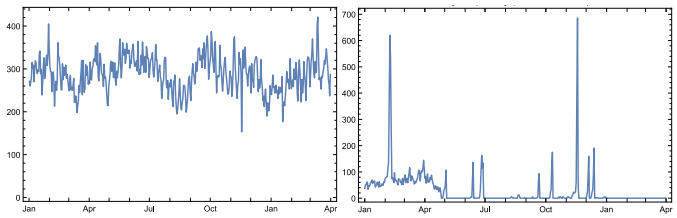


Figure 4. Side-by-side comparison of accepted glucose tests per day (left), with rejected tests per day (right) over the same period. Successful tests per day closely track patient numbers (e.g., note a dip in the August holiday period and another dip on Christmas day), but rejects show no obvious pattern, though one would assume rejects per day would correlate with the number of successful tests per day. In particular, note long periods of zero rejected tests and occasional brief periods of very high reject rates, almost *double*

the number of accepted tests. The obvious hypothetical explanation is that some records of rejected tests were deleted, some were merged, and possibly some dates were arbitrarily corrupted. The data itself does not provide any clues towards an explanation.

Once docked, the data then has a tortuous route through servers, middleware and Abbott's own software (figure 3). It can take days to get through. In particular, manual intervention is required for some data (e.g., when the patient ID is not valid somebody has to sort it out), but there was no evidence provided that any such manual corrections had occurred. Nobody, so far as I could tell, had sorted out this uncorrected data which remained in the database (technically, in a 'reject folder').

Such problems have not been reported on national reporting databases to my knowledge (such as the United States of America's Food and Drug Administration's MAUDE),[19] and the prosecution argued that because no problems had been reported it followed that similar problems did not occur elsewhere, and so it must follow that the problems were unique to this hospital and its nurses. However, there are peer-reviewed research papers that report identical problems at other hospitals using the same Abbott devices. These papers prove problems do occur but which are not reported to national reporting systems. One hospital that used their database to warn staff about poor practices saw improvements: in other words, not only is there research showing the same problems do occur elsewhere, but the same research also discusses solutions.[20] The lack of evidence in national incident databases shows that hospitals are not reporting errors, not that errors do not happen. Arguably, electronic problems are routine and hospitals do not worry about them — except for research purposes or when they go to trial.

Withdrawal of evidence

During the trial, the prosecution produced new evidence on an encrypted CD with gigabytes of XML files. This was the first evidence I had been able to analyse that had metadata and timestamps covering the period of the alleged incidents. (Though, as before, there was no digital signature, which I could use to verify the data, or any other proof of provenance.) The court ordered the expert witnesses to analyse this data and prepare a joint report under Part 35.12(3) of the Civil Procedure Rules.

The XML data was particularly interesting because the data contained timestamps, and the XML files themselves had creation dates that were plausible (the creation dates covered the relevant period — they were not dated, like the Excel spreadsheets had been, when the police created them but, rather, they were dated mostly at plausible times over the period of concern). The expert witnesses agreed a joint statement on the significance of the new 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. However, I noted that there were many indications that arbitrary data was missing, not just data related to specific nurses. I noted that the data had a very peculiar statistical distribution, strongly suggesting data mismanagement, errors, technical problems or similar. See figure 4 for one simple example from my analysis. Although I had (at least for the defence team) convincing arguments the data was unreliable, I had no good explanation *why* the data was unreliable.

A witness from Abbott was called to be cross-examined over the findings expressed in the joint report. During this cross-examination, it became clear that the witness had visited the hospital before the police had seized a copy of the data. It emerged that the witness for Abbott had 'tidied up' the database, and they had kept no records of what they had done. One should be clear that this was a naïve action, intended to be helpful, undertaken at the request of the hospital.

This disclosure led to the judge excluding the evidence,[21] and the case collapsed as the prosecution evidence was shown to have no probative value. The prosecution needed to prove that absent data was caused specifically by deliberate nurse behaviour, and

[19] MAUDE, the Manufacturer and User Facility Device Experience, https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm .

[20] Gaurav Alreja, Namrate Setia, James Nichols, Liron Pantanowitz, 'Reducing patient identification errors related to glucose point-of-care testing', *Journal of Pathology Informatics,* 2:22, 2011. DOI: 10.4103/2153-3539.80718, available at https://www.ncbi.nlm.nih.gov/pubmed/21633490 .

[21] Ruling in *R v Cahill; R v Pugh* 14 October 2014, Crown Court at Cardiff, T20141094 and T20141061 before HHJ Crowther QC, 14 *Digital Evidence and Electronic Signature Law Review* (2017) 67–71.

they needed to prove that absence proved the nurses fabricated patient records. They were unable to do this.

My many other arguments for the unreliability of the evidence were not raised or cross-examined before the jury. For example, the trial did not explore 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 (e.g., poor staff IDs), or why the error and alarm flags were set and what that implied for data quality. I would have liked to have shown, for instance, that the poor design of Abbott's equipment could provide no evidentially acceptable relation between data absence and nurse behaviour, but the confession that Abbott had manipulated data (and not remembered exactly what they had done) was enough to undermine the prosecution case.

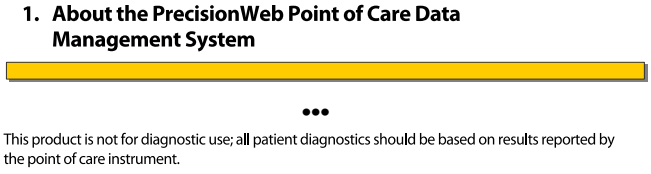
Technical discussion

The blood glucometers and related systems were not designed to be dependable. Or, to be more generous, they were not designed to be dependable given the well-known chaos of complex hospital IT systems involving many vendors, in the context of known clinical pressures, and of the known propensity of well-meaning staff to prioritise their clinical duties at all costs, even if that means performing workarounds to solve problems and overcome system hindrances.

If the glucometers had been designed more dependably, 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 the poor quality of their IT systems. Certainly, if the Abbott database had included confirmed transfer and missing data information, the police would have known immediately that the evidence was unreliable — though it will be recalled that they had not noticed error flags and other indicators of poor quality in the data they actually had.

The written evidence from Abbott strongly suggested that the software in them was not of high quality: for example, they wrote in evidence '[…] 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 Abbott devices are accurate as glucometers.[22] The prosecution understandably mentioned this, arguing that they were therefore good devices. But measurement accuracy was not relevant to this case. Regardless of whether the glucometers were accurate, the issue was recording the 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 published research exploring this aspect of their reliability — I would suggest from the experience reported in this paper that research in this area is needed.


1. **About the PrecisionWeb Point of Care Data Management System**

•••

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 Abbott's *PrecisionWeb Point of Care Data Management System User's Manual* (version QC Manager 3.0, 2009), page 1-1. The nurses were doing what the manual says: they were using the point of care instrument (the glucometer) and writing down the results.

Abbott's database software has an explicit warning 'this product is not for diagnostic use' (reproduced in figure 5). The question that I raise is, 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 being used for managing patient databases? Rejected test data (e.g., with bad patient IDs, typically caused by using nurse IDs instead of patient IDs) has to be edited with this product; if that is not clinical use, what is it?

To be sympathetic, the database might have originally been intended by the manufacturers purely for glucometer maintenance. For example, while I was double-checking my interpretation of the evidence, I

[22] For example, see Karlijn Gijzen, David L. J. Moolenaar, Jos J. A. M. Weusten, Hendrik J. Pluim and Ayse Y. Demir, 'Is there a suitable point-of-care glucose meter for tight glycemic control? Evaluation of one home-use and four hospital-use meters in an intensive care unit', *Journal of Clinical Chemistry and Laboratory Medicine*, 50(11):1985-92, 2012. DOI: 10.1515/cclm-2012-0104.

plotted the performance of the glucometers against their battery voltage and temperature. This is data the glucometers and the database record, and could be used by a hospital to help manage the glucometers — they could detect battery and other problems. The data I analysed was plausible, though there were interesting digitization effects.[23] I could not tell whether this was how the glucometers actually worked or whether it was an artefact of the police processing the data through Abbott's software, SQL or some effects inside Excel itself. The inability to tell where digitization errors occur is worrying, and raises suspicion of the 'forensic' processes used by the police. Analysing battery voltage reveals digitization *is* occurring but not *where* it is occurring. This might mean that digitization — or, other errors — also occurred in the evidence relevant to the criminal proceedings.

Surely, if the data was not to be used for clinical purposes, then the police should have been very cautious in extrapolating the evidence to imply criminal clinical practice occurred: they should have used independent evidence to establish the records were reliable evidence for their purposes. The police assumed the glucometers and hospital IT systems were completely reliable,[24] even though they knew they must require human intervention to be managed by the hospital. The police did not question the management of the data,[25] and there was no evidence submitted about the day-to-day management of the data. Possibly nobody was managing it! In addition, the Abbott systems themselves required human work to develop in the first place, and that human process, too, was unlikely to be infallible.

Evidence indicated that the Abbott database system crashed frequently — a problem the court never examined, because the judge excluded the evidence before we needed to draw this issue to the court's attention.[26] If, as was clear from the evidence, the system was regularly crashing, then this was another proof that it was not very reliable, and of course it was likely to be losing data or failing to record data when it crashed. If so, the alleged fabrication would be explained in yet another way. Such an argument would put an obligation on Abbott to explain how their systems were (if indeed they were) reliably storing all relevant data *despite* the crashes.

The Abbott systems are regulated devices and have appropriate CE marks,[27] but CE marking is a very coarse regulatory requirement and says little about dependability, particularly for use in the context of the multi-vendor networked environment of a typical hospital.

It seems the hospital, as well as the police, *unconsciously* assumed all the data was perfectly reliable, and sufficiently reliable for a criminal investigation. Nobody consciously *said*, 'we are assuming the data is reliable' — it never crossed their mind to state their assumptions; there is no English word to describe an assumption you are acting on but do not know you are assuming. The hospital, police and Crown Prosecution lawyers made critical assumptions, but they were not aware any such assumptions had been made. It did not occur to them that these assumptions, if acknowledged, could easily have been checked.

Subsequent disciplinary hearing

Internal disciplinary processes were suspended during the course of the trial. Resuming the internal disciplinary process to discipline one nurse, the discredited police evidence was reused as if it was unproblematic. The disciplinary hearings presented a misunderstanding:[28]

> 'in an internal disciplinary proceedings the burden of proof is a lower threshold than in criminal proceedings …'

A disciplinary hearing may be interested in issues that are not considered criminal, and for such issues the

[23] For example, a battery voltage of 3.4 volts may be digitized to and recorded as 3 or to 4 volts, depending on how the data is processed. The result is, instead of a continual range of voltage 3.3, 3.4, 3.5, etc, there may be strange jumps 3 to 4 that have nothing to do with the glucometer but are caused by Excel or other data processing systems. (Almost certainly numbers will be digitized in binary rather than decimal as illustrated here.)

[24] Despite the text of chapter 6 of *Electronic Evidence*, which was first introduced into the second edition in 2010. This suggests none of the Crown Prosecution lawyers or police officers were aware of this book.

[25] Tests for authentication are set out in *Electronic Evidence* at 7.128. These tests have been up-dated and improved upon since the first edition in 2007 and are now incorporated into the Draft Convention on Electronic Evidence, for which see 13 *Digital Evidence and Electronic Signatures Review* (2016) S1 – S11.

[26] Ruling in *R v Cahill; R v Pugh* 14 October 2014, Crown Court at Cardiff, T20141094 and T20141061 before HHJ Crowther QC, 14 *Digital Evidence and Electronic Signature Law Review* (2017) 67–71.

[27] The letters 'CE' appear on many products traded on the extended Single Market in the European Economic Area (EEA). They signify that products sold in the EEA have been assessed to meet relevant safety, health, and environmental protection requirements.

[28] The disciplinary hearing records were made available to me by a nurse, but are not public.

burden of proof can be lighter since the issues are less serious. However, in the present case, the rigorous standards of the criminal process discredited the evidence thoroughly. The court established that the evidence had no value, and properly understood, the data had no value even for a disciplinary process (other than, perhaps, for an investigation into the way the hospital dealt with devices and systems controlled by software).

Consider:

> 'The [disciplinary] 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 disciplinary hearing concludes that the] 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 transcripts, yet we know that Abbott's engineer had deleted this data and therefore it is inappropriate to infer anything about what any nurse actually did *from this data*. The database errors also affected about 70 other nurses similarly, so it is hard to understand this as an issue about the one nurse in the disciplinary hearing.

I submitted written evidence to the hearing, including quoting from the judge's Ruling:[29]

> '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 [the Abbott engineer] 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. CH20 [the Precision Web database evidence] has lost significant amounts of data: but there is no way to tell whether the missing files were reintegrated into the PrecisionWeb 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 than probative or I should consider it hearsay because of [the Abbott engineer's] intervention, and unreliable hearsay.
>
> […] their adduction in evidence 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 PrecisionWeb or the reject folder.'

It is worth adding that the prosecution experts also agreed with the judge's Ruling. The disciplinary process, however, drew a conclusion the judge said a jury could not.

The hospital argued I had analysed evidence admitted into criminal legal proceedings, and not data belonging to the hospital. At any stage, the hospital could have compared the police evidence against their copies of the database to establish whether the problems affected the court alone and not the disciplinary proceedings as well. Either the data would be the same (in which case my arguments would be valid), or the discrepancies should have led to investigations on data management practices: why (if it was the case) were the police handed corrupt data different from the data the hospital retained? If I had been analysing different data, that again calls into question the integrity of the hospital database — if the police's forensic methods had seized invalid data, perhaps the hospital's internal procedures would have the same problems? On any interpretation, in my view, any nurse-specific conclusions drawn from the data were invalid.

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, and so on.

Prominent peer-reviewed papers reinforce this naïvety[30] — they give the impression that glucometers just measure blood glucose levels and ignore that

[29] Ruling in *R v Cahill; R v Pugh* 14 October 2014, Crown Court at Cardiff, T20141094 and T20141061 before HHJ Crowther QC, 14 *Digital Evidence and Electronic Signature Law Review* (2017) 67–71.

[30] James H. Nichols, 'Blood glucose testing in the hospital: Error sources and risk management,' *Journal of Diabetes Science and Technology*, 5(1): 173–177, 2011, available at https://www.ncbi.nlm.nih.gov/pubmed/21303641 .

readings have to be reliably networked and recorded in databases. Even the recent 2016 UK *Making IT Work report*[31] (the *Wachter Report*) takes it for granted that computers work[32] (with minor caveats on interoperability and usability) — hospitals, apparently, just need to 'digitalise' more, with a new government investment (£4.2 billion) available to purchase IT. The report takes it for granted there is good IT that can just be 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 and its management is effective or even fit for purpose, which is just assumed. Every hospital is in good company, then, unwittingly drifting into cyber-failure.[33]

Ironically, if the hospital in the story here had been 'paper free' (as the recent *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 urgently need to work out how to make it more reliable.

Failure only becomes apparent after there is a visible incident. In the case described here, something initiated the police investigation and discrepancies between paper and IT then became 'the incident'. 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. But without paper, there would have been no visible incident.

With hindsight, we can see there were many causes of the incident. All 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 so many nurses are textbook examples of baseline shifting, confirmation bias, and cognitive dissonance.[34] With so many nurses apparently implicated, underlying systemic factors, including management[35] and IT support should also have been obvious priorities to critically examine.

Other hospitals will have analogous complex IT problems. A priority everywhere should be to have a mature cybersecurity strategy, which implies having an implementation of IT that permits having a workable strategy — and there needs to be a regulatory requirement on manufacturers to supply equipment that is reliable enough to support the strategy. Effective procedures must be in place to detect, interpret and respond to unusual or unauthorized activity immediately. In the case here, data was deleted and corrupted, and the hospital did not notice. Hospitals need to tighten up their cybersecurity monitoring, which should be able to detect such incidents, whether caused by external or internal hacking. Good guidance on data management is available elsewhere.[36]

Over a long period of time, the hospital staff (nurses, management, IT management, etc) ignored evidence of behaviour that the police later treated as criminal; that is, the recording of tests on the database apparently had no day-to-day significance (else somebody would have noticed years earlier that it was not working). When data was deleted, the hospital did not notice. The police assumed the corrupt data was perfect, and formed adequate evidence to charge nurses. There were many reasons that the data was unreliable, but the simplest was that Abbott had corrupted large parts of it. That fact alone was sufficient for the trial to collapse.

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 me as an expert witness in my technical enquiries. This missed early opportunities to uncover some of the systemic problems. A parallel investigation asking the question "why is the data corrupt?", carried out within the hospital, would have saved a significant waste of time and huge costs to the defendants and to the hospital.

Unfortunately, it is easy to blame people trying to use bad technology. Blaming nurses makes a compelling

[31] Robert M. Wachter, *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, 2016).
[32] Despite the evidence set out in detail in chapter 6 of *Electronic Evidence*.
[33] Sidney Dekker, *Drift into failure*: From hunting broken components to understanding complex systems (CRC Press, 2011).
[34] *Mistakes Were Made (But Not by Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts*.
[35] Jane E. Ball, Trevor Murrells, Anne Marie Rafferty, Elizabeth Morrow, and Peter Griffiths, 'Care left undone during nursing shifts: Associations with workload and perceived quality of care,' *British Medical Journal Quality & Safety*, 0:1–10, 2013. DOI:10.1136/bmjqs-2012-001767 , available at https://www.ncbi.nlm.nih.gov/pubmed/23898215 .
[36] *Data safety guidance.*

tabloid story, involving simple human failings we think we understand. We so quickly feel angry about poor patient care, and feel a sense of betrayal by nurses being incompetent that it is hard to explore other explanations. Unfortunately, it is difficult to form a compelling story about complex and unreliable IT systems, because people will not generally understand the intricate truth.

There is no quick fix, other than getting relevant experts involved, which of course depends on recognising that experts are needed. Dekker discusses these important issues in much more detail than we can here.[37]

Conclusions

While the hospital had a core part to play in this story, we ought to consider the device and the data systems. Manufacturers have years to develop and produce better IT; they have many years to develop hospital equipment, and they should know how to do it. Arguably, then, their equipment should be reliable to use correctly in a real ward in a real hospital, along with its existing complex networks and database systems. It should be easy to monitor patient data and ward activity and adherence to standard procedures.

In contrast to the manufacturers, the hospital and nurses in particular have a very time-pressurized, clinical job to do, and it is reasonable that they rely on the quality of regulated products marketed by reputable, international manufacturers. It should not be the job of nurses (or even of disciplinary processes) to figure out what is wrong with hospital IT.

Arguably, it should not be the police's main job to figure out intrinsically faulty IT either, although the VW scandal[38] and other examples should urge a more careful approach. Nevertheless, the police have years to pursue an investigation, in contrast to the nurses who have literally seconds to do their job. The police should use their relaxed time more cautiously, something the nurses cannot interrupt their patient care to do.

The manufacturers argued in court that because their products were regulated and CE marked, any problems must be the nurses' fault. This is an attitude that misdirects attention from improving systems to blaming users. Industry needs to move from saying 'users make mistakes' to saying 'users make mistakes *so we must make better systems*.' In the case here, serious mistakes were made in the back-office by the manufacturer's representative, not by any end users or nurses, and underlying those mistakes were system design errors that allowed those mistakes to remain unnoticed.

It is clear that the regulatory environment is not fit for purpose. In contrast to IT-based medical devices, pharmaceutical regulation requires that drugs are safely developed and tested and that problems (such as side-effects) are reported. IT regulation has not caught up, despite the significant impact of IT on healthcare — not just in direct patient care, but in auditing, finance, and all areas of healthcare, even in incident reporting and, ironically, in pharmaceutical regulation. For hospitals, effective IT regulation is more critical than pharmaceutical regulation.

While hospital IT regulation may not change soon, given the well-known unreliability of IT generally, it is surprising the police were not more cautious in interpreting the electronic evidence. The data corruption underlying this case could have been detected and managed as soon as problems occurred, had the hospital had a mature approach to cybersecurity (or had the manufacturers designed the system to be monitored reliably). The police got involved in a problem that had escalated because the hospital had not been monitoring its data.

Given the widespread use of poor IT throughout healthcare, hoping for manufacturers to improve is perhaps less realistic than encouraging hospitals to prioritize mature cybersecurity and better procurement. They need to do this because effective cybersecurity measures are protective against a wide range of problems, including all of the problems discussed in this paper. Fortunately, cybersecurity already has a high profile (thanks to widespread fear of malicious hacking, and thanks to high-profile cases like WannaCry): this paper adds more reasons to take it seriously. It is not just about attacks, loss of data and downtime, but about *understanding*.

The broader problem is the uncritical acceptance of IT, from legal, regulatory, procurement and other perspectives, especially for healthcare, where billions of pounds are eagerly invested in new IT. Our culture

[37] Sidney W. A. Dekker, 'Prosecuting professional mistake: Secondary victimization and a research agenda for criminology,' *International Journal of Criminal Justice Sciences*, 4(1): 60–78, 2009.
[38] For an overview, see https://en.wikipedia.org/wiki/Volkswagen_emissions_scandal .

makes us all uncritically believe that IT and especially the latest IT is wonderful — don't we all want new things?[39] (Of course, this is how IT companies stay in business.) However, 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, and for the good of staff. Ironically, the newer IT is, and the more exciting it seems, the less tested it seems to be in the clinical environment.

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

Harold Thimbleby is See Change Fellow in Digital Health at Swansea University, Wales.

EPSRC funded part of this work under grant EP/L019272 [http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/L019272/1]. I am grateful to a huge number of anonymous people who helped with this paper and the background information and research. I am also very grateful to the nurses who had the courage to defend their innocence.

This paper has been revised and extended from an earlier version published as H. Thimbleby, 'Cybersecurity problems in a typical hospital (and probably all of them),' in Mike Parsons and Tim Kelly, editors *Developing Safe Systems, Proceedings of the 25th Safety-Critical Systems Symposium* (Centre for Software Reliability, Safety-Critical Systems Club, 2017), 415 – 439.

http://www.harold.thimbleby.net

harold@thimbleby.net

Appendix

Sample recommendations

The list of recommendations below is divided into sections of recommendations primarily focused on hospital, police, manufacturer, regulator and researcher interests. Cybersecurity and IT are dynamic, ever-changing areas, and any specific list of recommendations is bound to become obsolete and eventually become counter-productive.

There is no particular order of priority in the following recommendations. They should all be read. The recommendations should inspire your thinking: if you spot errors or omissions, then you are thinking about the issues, and that is the aim (and please email me any insights). The recommendations should not be read as criticisms — they are to encourage constructive thought; good cybersecurity is not easy.

(1) IT governance. Everyone (hospitals, police, lawyers) should put in place processes to continually review and update their cybersecurity and IT maturity. Hospitals should have board level cybersecurity expertise and responsibility — remember that a cybersecurity incident can put a hospital out of action in seconds. Note that the GDPR makes much of this a legal requirement. Elsewhere we have published advice[40] but it must be emphasised that is dated 2018, and continual review of cybersecurity is essential, for instance by using the resources at the UK GCHQ National Cyber Security Centre, including their training and certificated courses.[41]

(2) Risk assessment and safety cases. All projects (whether installing new IT or investigating IT) should include an explicit IT and cybersecurity risk assessment. Procurement should insist on seeing appropriate documentation from suppliers.

(3) Continual review and training. IT is used to support an organization and to help deliver its products. For example, manufacturers require IT to function, and their products contain IT. Police require IT to function, and their investigations may involve IT. Hospitals require IT to function, and patient safety depends on reliable IT systems. Both aspects of cybersecurity and

[39] Harold Thimbleby, 'Trust me — I'm a computer,' *Future Healthcare Journal*, 4(2):105 – 108, 2017. DOI: 10.7861/futurehosp.4-2-105; this includes lawyers, for which see *Electronic Evidence*, chapter 6.

[40] 'Six Factors Essential for Mitigating Cyber Risks in Healthcare'.

[41] https://www.ncsc.gov.uk/guidance.

IT maturity must be continually reviewed, and not only reviewed but supported with adequate training.

(4) External oversight. Internal reviews are inadequate, as they may merely perpetuate existing blindspots. Those people in positions of authority should formally engage professional *external* and independent oversight, and collaborate closely with colleagues nationally and internationally to learn and share best practice. Shneiderman makes a strong argument for independent external oversight of all algorithms, not just healthcare systems.[42]

(5) Resourcing. Cybersecurity, improvement and culture change is at least a full-time job. Avoiding predictable future problems and catastrophes requires dedicated staff and investment. Mature cybersecurity cannot be achieved alone, but requires continual 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 discusses '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).

(6) Digital and investigative skills. All investigations should understand statistics, data fishing, and question IT reliability — is the data reliable; does it pass statistical tests? One can use data (especially bad data) to support almost any case if the baseline is ignored.

(7) Thinking out of the box. IT is developing rapidly and changing the rules. Reliance on law or regulation to inform IT decisions is naïve. Cyberattacks succeed by surprise, which means to defend against IT problems, we must defend against people who do not play by rules. Ironically, the poor understanding of the problems created by the software systems that had to be managed manually initiated off this prosecution did not surprise anybody, because nobody was alert.

What hospitals should consider

(1) Hospitals should continually monitor all data for cybersecurity anomalies and attacks. Basic security protocols, if in place, would have raised alarms at Abbott's tampering with the data (or any other unexpected or unwanted changes) as and when it happened.

(2) It is routine for hospitals to perform clinical audits; IT audits should be routine too; ideally audits should be undertaken with external oversight. Hospitals should audit their data and monitor it. In the case described here, a hospital collected data it did not monitor, so deleting or otherwise tampering with the clinical data was not detected when it happened.

(3) Hospitals should procure equipment with reliable IT. In the case here, a hospital procured equipment that unreliably recorded clinical procedures. The legal department should be aware of the failings with software, and ought to ensure such failings are properly dealt with in the formation of contracts. There are many standards for helping improve procurement processes.[43]

(4) Hospitals should disable all IT systems and features they are not using. They should not install software that is not used or monitored, and they should resist police attempts to seize data which they cannot verify is accurate.

(5) A hospital should not have systems supporting clinical work that are expressly designed as not for clinical use; they should only be used with proper caution. Investigations should never rely on such systems without independent verification that they are appropriate for the specific purposes of the investigation.

(6) When police request data from a hospital, proper governance procedures should be followed. One of the problems with the case in this instance is that the police obtained evidence without oversight; neither the hospital nor the court had any idea what evidence the police had.

(7) Hospitals, police and lawyers 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 more complicated (e.g., whether measurements get processed reliably in multi-vendor systems).

[42] Ben Shneiderman, 'Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight', *Proceedings of the National Academy of Sciences*, 113(48):13538-13540, 2016. DOI 10.1073/pnas.1618211113 .

[43] Brian Donnelly, *International Code of Practice for Providing Technology Enabled Care Services* (Troubador Publishing, 2018).

(8) Hospitals should routinely and regularly disclose to staff what data they are collecting. They should allow staff to see and, if necessary, to challenge it and the processes used to collect it. This means releasing data regularly. If data has no clinical role, then it should never be allowed to be used.

(9) Evidence discredited in the rigour of a criminal trial should be used to help understand the causes of the system failures. It should not be used with in disciplinary processes. A likely failure in the case here was that the hospital did not do a thorough post-trial review: what went wrong, what could be learned?

(10) Hospitals should be aware that trials can collapse or terminate for any reason before all evidence is presented. For example, in the present case, once it was known that the database had been corrupted the trial collapsed, and further evidence (such as problems with staff IDs, wrong glucometers seized, and so on) was never raised. Working with the consultants and expert witnesses after a case will reveal information and insights that never reached the trial.

(11) 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.

(12) Performing reviews after incidents is important, but such reviews should be externally chaired or assessed by experts. Internal reviews that followed this prosecution have responded by noting that 'we have since replaced […]'s system with a wifi-based system, so there is no problem' — yet the problem was nothing to do with the network, wired or wifi, but with failures in management and a basic understanding of IT.

(13) When something go wrong, *everything has gone wrong* — every defence failed to stop the incident. This is the vital message of James Reason's Swiss Cheese Model[44] — things go wrong when *every* defence fails. Nurses are one defence, but so too are management systems, training, supervision – and IT systems. Hence if things fail, then do not just treat it as a personnel disciplinary issue. Another of Reason's insights is that if something goes wrong, if anybody could have made the same mistake and it leads to the same incident, then there is a problem with the system and not that particular person. So if 70 people make the same mistake, what we might call 'Reason's Law' says it is a system problem. (Of course, some nurses in the 70 might have been negligent, but it is unlikely all of them were). This 'Reason Law' is implemented in the NHS's *Just Culture Guide*, where it is called the substitution test (would anyone else substituted from the same peer group be likely to do the same things?)[45]

(14) There is growing acknowledgement that hospitals should move away from a *blame culture* to a *just culture*.[46] With a rapid focus on blame, hospitals may fail to learn more general lessons, and hence fail to improve; the important insight is that when clinicians are blamed, that is not the end of the story, and it is rarely the place to start. The temptation to discipline or sack an individual creates the impression the problem is solved, but doing so discounts the underlying causes (such as IT failures) and the importance of addressing the causes and not just the symptoms. Just culture is usually associated with patient safety investigations, but it applies to any investigation, and could have applied in the present case, and I believe would have resulted in a more constructive process.

(15) Review all IT governance and resourcing.

What the police should consider

(1) The police failed to have a critical approach to data analysis. There were many simpler explanations for the alleged fabrications by 70 nurses: for example, computer problems, a vendetta in the back office, poor data management, faulty equipment, faulty networks, and more. The police should routinely assess alternative hypotheses, and provide evidence they have done so.

(2) The police never considered assessing the internal or external validity of their evidence. A cursory analysis of the database would have raised many questions about its quality.

(3) The police management of data exposed numerous IT problems. For example, one piece of evidence indicated that the police found that their Excel crashed analyzing the data. It is surprising the

[44] James T. Reason, *Human Error* (Cambridge University Press, 1991).

[45] NHS Improvement, *A Just Culture Guide*, 2018, available at https://improvement.nhs.uk/resources/just-culture-guide .

[46] Sidney Dekker, *Just Culture: Balancing Safety and Accountability* (Ashgate, 2007).

police even tried using Excel to analyze the complex relations in such a large volume of data.

(4) There was no clear management of the evidence. Evidence I was given included spreadsheets (including police analysis *in* the spreadsheets), which made me wonder what other edits the police had made to what was claimed to be evidence. Despite the 'forensic' methods the police claimed to use, they were not presented with the evidence and did not help confirm provenance.

(5) Data discrepancies were used as evidence, so I was surprised at typographical errors affecting data presented in the prosecution evidence — this is ironic, as the prosecution case relied on the quality of data. Of course, I may have made some typographical errors in my own evidence that I did not notice. (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.)

(6) The police seized several glucometers. As other glucometers had been used on the ward, 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. The implication is that the police were inadequately aware of clinical practice and did not engage with the hospital to help seize the correct equipment.

(7) What was the ward supposed to do when they lost their glucometers to the police? The police should do a risk analysis, because removing glucometers from a ward puts patients at serious risk of harm (ironically at greater risk than the alleged incidents).

(8) There were surprising conflicts of interest.[47] A technician from the manufacturer corrupted data, then selected the data from the hospital and handed it over to the police, and the police sent the glucometers to the manufacturer to analyze and confirm whether they were functioning correctly. Independent experts should have been used throughout. In fact, the police used consultants with narrow terms of reference; if the consultants had worked with the expert witnesses, the case might have been avoided.

(9) As the manufacturer'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.

(10) Knowing glucometers function correctly when analyzed tells you little about how they might have performed in the past. (For example, the glucometers may have been serviced since the incidents, so they work well now but did not before they were serviced.) The police made no attempt to establish whether the glucometers worked correctly at the material time — and the logs the glucometers recorded seemed to offer no help here.

(11) It did not appear that the *ACPO Good Practice Guide for Digital Evidence* (March 2012, v5) were followed by the police.[48]

What regulators should consider

(1) The very public discovery of VW's fraudulent IT to help their cars pass emission tests[49] — 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 (particularly of healthcare products) should adopt processes that reassure users or the public more generally that they are reliable, for instance by using open source methods so that their code can be externally vetted.

(2) Formal methods should be required in design and development of hospital IT systems.[50] Manufacturers may complain that they do not know how to use formal methods for their complex products. They should start making products that are simple enough for them to understand. Logically, if the manufacturers can only 'hope' to understand their

[47] Linda Geddes, 'Evidence of failure,' *New Scientist*, 3175:22–23, 28 April, 2018.

[48] Available at http://library.college.police.uk/docs/acpo/digital-evidence-2012.pdf .

[49] Russell Hotten, 'Volkswagen: The scandal explained,' *BBC News*, 10 December 2015, http://www.bbc.co.uk/news/business-34324772 ; *Electronic Evidence*, 6.139.

[50] There is significant literature on formal methods. For instance, see, M. D. Harrison, M. Drinnan, J. C. Campos, P. Masci, L. Freitas, C. di Maria and M. Whitaker, 'Safety Analysis of Software Components of a Dialysis Machine Using Model Checking', *Proceedings International Conference on Formal Aspects of Component Software*, Lecture Notes in Computer Science, volume 10487:137-154 (Springer Verlag, 2017). DOI: 10.1007/978-3-319-68034-7_8.

own devices (e.g., failing to use formal methods means relying on hope alone), then non-technical hospitals and nurses are very unlikely to understand them either.

(3) It is notable that regulation is primarily concerned with patient health; staff well-being (and hospital effectiveness) clearly ought to be a consideration.

(4) The CE marking system, which is used to indicate products meet the relevant requirements of the European Medical Device Directives, is discredited,[51] 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 rapidly 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.

(5) In our complex world of IT threats, regulation urgently needs to be more open, and more responsive to legitimate concerns.

(6) There are rigorous processes 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 as used in pharma may not be the essential methodology for cybersecurity, but there are other methodologies such as formal methods where correctness is proved. Such methods should be used, and must 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 (where lives depend on them).

(7) To become a nurse to use equipment like blood glucometers requires a minimum of three years degree level training, plus registering with the Nursing and Midwifery Council. You can then become a registered nurse. You also need to pursue continual professional development throughout your career. As a registered nurse, you have a legally protected title. It baffles me that to develop anything a nurse uses, you could start this afternoon with no qualifications whatsoever, yet doing a job like programming a glucometer is doing something as life-critical as nursing — actually, more so, as your glucometer could harm millions of patients. If you are building equipment that professionals use, surely you ought to have qualifications to confirm your competence? This requirement would be a simple regulatory change, and would be welcomed as it would give qualified developers higher salaries and give manufacturers assurance they were building dependable systems.

(8) I would change the law, so that if a manufacturer cannot provide a complete log of all data it has processed (with appropriate forensic checks), then the presumption is that the manufacturer is liable for lost or corrupted data.

What manufacturers should consider

(1) Manufacturers should employ (or out-source to) developers with adequate qualifications. See point (7) above.

(2) Hospital IT systems are very complex (through no particular fault of the manufacturer cited in this article) and this complexity is not going to change. 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. Auditing needs to be provided, work and be used.

(3) Installed software in hospitals that is not being actively managed (as happened here with Abbott'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).

(4) The equipment in this case was CE marked. This implied that any problems must be the nurses' fault. With regulations in place of this nature, there is little incentive for manufacturers to try harder. Manufacturers of clinical products should closely consider the quality of their programming, and the whole system of CE marks should be more transparent, and lawyers should be willing to test the efficacy of products with CE marks.

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

[51] Deborah Cohen, 'How a fake hip showed up failings in European device regulation,' *British Medical Journal*, 345:e7090, DOI: 10.1136/bmj.e7090, 2012, available at https://www.bmj.com/content/345/bmj.e7090 ; Deborah Cohen and Matthew Billingsley, 'Europeans are left to their own devices,' *British Medical Journal*, 342:d2748, 2011. DOI: 10.1136/bmj.d2748, available at https://www.bmj.com/content/342/bmj.d2748.long .

by the relevant international standards ISO 15197, ISO 14971, ISO 13485, etc).

(6) It was disappointing that in this particular case, no technical experts from the manufacturer wanted to appear in court, although they provided much written evidence to the prosecution. Manufacturers should be eager to support investigations concerning their products.

(7) The widespread poor design of user interfaces suggests that reliable operation was not a priority or perhaps not a competency for the manufacturer. The evidence presented in court suggested the quality of the programming of the device and the database management software was an issue, and the 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.

(8) Abbott's systems have what they called an 'audit' feature. It certainly has a feature *called* audit, but it is manual and fallible, and generated 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.

(9) It may seem unfair to criticize manufacturers without offering any solutions. One easy thing to do 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 security problems). Lost or corrupt data should then be routinely reported to the manufacturer's post-market surveillance team. It would then be easy for a hospital to use best efforts to respond and recover it.

(10) There are published papers on the performance of hospital equipment; manufacturers should follow this literature and update and respond to research insights.

What researchers should consider

(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?

(2) While security research has a high profile, security 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.

(3) Researchers should lead an analogous structure to the Information Sharing Analysis Organizations have already established for cybersecurity.

(4) Formal methods are 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.[52] One of the many research problems is how to migrate large, complex, error prone 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) Modern ideas for dependable distributed systems could be applied to the problems of getting data around hospitals. Distributed ledgers may not be the solution in the present case, but distributed ledger *thinking* could be very productive.

(6) Throughout this paper I have criticized the culture of assuming IT and data is perfect. In the UK, this culture is enshrined in law: it is presumed that IT works correctly.[53] Computers are deemed to be 'in order' and 'properly set and calibrated' – yet the position is even more absurd than described above, as Mason explains:[54]

[52] John Barnes, High Integrity Software: The SPARK Approach to Safety and Security (Addison Wesley, 2003).
[53] *Electronic Evidence*, chapter 6; Stephen Mason, 'Electronic evidence: A proposal to reform the presumption of reliability and hearsay,' *Computer Law & Security Review*, 30(1):80–84, 2014. DOI: 10.1016/j.clsr.2013.12.005.
[54] Stephen Mason, 'Artificial intelligence: Oh really? And why judges and lawyers are central to the way we live now – but they don't know it,' *Computer and Telecommunications Law Review*, 2017, Volume 23, Issue 8, 213 – 225, 222.

> 'The presumption also illustrates the hypocrisy at the heart of English law. Lawyers write clauses for contracts relating to the use of software code that require the user to accept that the software is not free of errors. Such contract terms are considered so normal that nobody appears to understand this fundamental contradiction between the presumption and the acceptance of flawed software code as being normal.'

We cannot simply blame the police or the hospitals when they reflect the absurd and hypocritical legal presumptions from the legal culture in which they operate. While Mason 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. Until that changes, everything else is pushing against the tide.

(7) 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 digitizing healthcare and increasingly relying on (unregulated? insecure?) medical apps. Blunt end human factors (e.g., design error by manufacturers) 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.

(8) Healthcare IT is very complex, beholden to confidentiality (patient data, and commercial confidentiality), lack of data, poor data, and numerous incompatible vendors and independent initiatives (such as national funding for AI for cancer diagnosis, but not for dementia), and results in workarounds and more. Meta-research needs to help figure out what an improved healthcare IT infrastructure would look like *and how to get there safely from here*.