

4. Implications for research and statistics on extending Secondary Uses to facilitate third party time-sensitive micromanagement of Direct Care.

For a rich research dataset, “regular” is the goal; a *monthly* administrative census of every treatment of every patient in the UK’s hospitals is sufficient. But the “calendar gap”¹ makes it clear that a monthly dataset cannot be used for micromanaging Direct Care – an activity which is not, in itself, Direct Care. The goal of a secondary uses statistical dataset is to know what happened, without necessarily knowing what is happening right now. That is the basis for the separation between direct care and secondary uses.²

Not everyone gets to do direct care, in the same way not everyone gets to be an astronaut. Sweeping the floor does help put a man on the moon, but no one credible thinks that is the same thing. What you don’t do is as important as what you do.

The (ONS decennial) Census goes to great lengths to reassure people about activities for which the data will *not* be used. If that case is not made, and believed by citizens, then the process will be fundamentally self-defeating. It is recognised that cutting corners or attacking the process would undermine the integrity of the data – rendering the entire process pointless.

UKSA takes its custodian role seriously, knowing it applies across the whole of the UKSA statistical estate. For example, the British Crime Survey asks citizens to respond with any crimes they (may) have committed – and does so in a way which maintains the confidence of the public, and the confidence of those who answer questions honestly. (It also allows citizens to lie in a manner that maintains systemic integrity.)

If administrative decisions destroy confidence in the ‘secondary uses’ of data – whatever uses those may be, and regardless of whether the data is from medical records or other sources – then the price will be paid far beyond the NHS. If NHS England can use patients’ medical histories for administrative micromanaging, why not genomic data – or anything else a well-meaning micromanager in Whitehall might think helpful next time?

Why this matters

While NHS Digital argues that data can be “anonymous in context”, it is impossible to argue this in the context where operational decisions are being made by NHS England. For, while NHS England should not know who the patient is – it has no direct care relationship – it will invariably be able to do so, because of the other information available to it. As such, for these purposes, SUS is an identifiable dataset, and dissent must be honoured.³

If the separation of operations and statistics collapses – because operational processes are masquerading as statistics – and once patients are aware that they can prevent someone micromanaging their hospital by opting out, the stampede to escape that evident risk to a beloved institution will make care.data look like a minor blip.

¹ This “calendar gap” has been clearly illustrated by the operation of Google DeepMind’s Streams app: <https://medconfidential.org/2017/medconfidential-response-to-deepminds-statements-about-their-legally-inappropriate-data-copying/>

² <https://twitter.com/thatdavidmiller/status/885427094464999424>

³ While this should already be the case, NHS bodies continue to avoid the issue by the use of perpetually-renewing ‘Section 251 support’. This tactic will shortly cease to be available.

The administrators' goal: nightly hospital micromanaging?

Every organisation working on useful collaborations with a hospital must have a clear legal basis to receive the data needed for each of those lawful collaborations. Given that monthly updates are entirely adequate, it is evident that the driving reason for making SUS *daily* is to force organisations to hand operational data to those who have no operational role.

This route is being chosen because “it is an existing dataset”, and there is no other mechanism by which hospitals can otherwise be compelled to hand data over without a public debate about whether it is truly a good idea to use medical records as evidence in political decisions.

We understand that NHS England has a drive (and/or need) to intrusively manage a hospital and so, absent a desire to get an explicit statutory basis for the data it wishes to acquire, it has to follow existing law. Officials may have to write down what information it is that they want in order to make a decision – and that information could then be published as statistics. We return to this point later, but NHS England's opposition to this is that it would then be accountable for the statistics requested, which would make policy-based evidence-making much more difficult.

The use of ‘secondary datasets’ for operational purposes is catastrophically destructive. Hospitals and clinicians understand this – even if Government administrators, who have no responsibility to patients, do not. This was a major folly of care.data.

There will always be a desire for more data. At some point, in a crisis, “yesterday” will also be deemed insufficient by those who currently believe it is all that is required. This is another (fundamentally flawed) rationale for a “national data lake” – so in the future, administrators can go fishing, even if today they claim they would not do so.

There are institutional consequences to being reckless with a longitudinal linked administrative census of all UK health care

The “Hospital Episode Statistics” (HES) are misleadingly named. They are the raw medical records of every patient in an NHS hospital; every treatment, dated, and recorded against a person's unique identifier.

Every patient is a special unique in statistical terms: everyone is identifiable if you know anything about their medical history – such as that they were hit by a car on a particular day, and it got written up in a newspaper. It is an administrative census of a level of richness and detail that population statisticians can only dream of.

Were patients to be able to follow every copy of patient-level data about them that flowed across the system, they should be able to understand what data was used, and why, with “no surprises”.

The vision might be worthy, but it is perception and outcomes that matter

We understand the desire to reuse 'available' data to solve any problem, and that NHS England's adversarial management style has backed it into a corner; it is institutionally antagonistic to hospitals, and so cannot get a dataset it wants via cooperation.

There are, of course, alternate approaches to addressing any and every question to which NHS England has said it wishes to see an answer, that can be done in such a way that these problems don't even have to arise. However, NHS England will continue to ignore these alternatives and almost certainly attempt to reuse patient-level data for the purposes we describe.

So will the public *really* accept NHS England and/or Whitehall micromanagement, as a purpose? Using patient-identifiable medical records, deemed a 'legitimate' secondary use? Remember that *each* patient involved was in the hospital at the time.

[Hospital management](#) is not Direct Care, no matter how much NHS England might like it to be. Insisting that SUS becomes a nightly dataset makes it self-evidently a proxy for a 'micromanaging hospitals' dataset. If such a precedent is set with hospitals, then CPRD and Genomics England would clearly be at similar risk of individual patients' medical records being treated as a performance management resource, *especially* in times of crisis.

We emphasise: there would be no problem were managers at NHS England to request any statistic they wished, from automated systems – even daily – though they should clearly request statistics on all aspects that affect care pathways, to avoid perverse incentives. But demanding raw medical records to perform analyses for political means is, and always will be, unacceptable to patients and to the public at large.

These issues are only exacerbated while NHS England continues in its culture of coverup and thinly-veiled political intent; lurching from crisis to crisis, haemorrhaging trust and increasing the cost with every mistake and misstep.

Wider Second-Order Effects of getting this wrong

Patient-level data is invariably attractive in the short term for certain purposes. But the price of NHS England's desires would be borne by others, and affect many more besides.

Any patient-level data used for the purposes under discussion can never be considered "anonymous in context", given the "context" would be the management of the hospital. And, given the administrative chaos around NHS England, when officials are attempting to second-guess clinical decisions about patients, they will invariably find out identifying information about those patients.

Setting aside other issues of lawfulness, were any clinician to attempt what NHS England wishes to do, there would be serious questions to ask about negligence. For this reason alone, opt-outs would be required to be honoured.

Public reaction to a decision to exclude uses by DH / NHS England for managing hospitals would likely undermine the Caldicott Consent regime. In all possibility, fatally. Similarly, offering patients an opt-out from performance management alongside the Caldicott Choice would have fundamental and catastrophic effects for legitimate research.

It is in times of crisis such as these⁴ that people are most likely to do something perverse, igniting what has already become a powderkeg issue.

medConfidential has never met with Will Smart, nor have we met or spoken to the people who have spent 6 months repeatedly rewriting the “Data Lake” strategy. Perhaps none of their weekly drafts are worth the paper they’re written on. What is clear is that they also can’t work, and [meet the requirements](#) of medical ethics and [confidentiality](#).

The alternative: diverse official statistics are ungameable

In the past, quality scores were gamed – most egregiously by Mid-Staffs – because there were very few of them, and so they were manipulatable. And companies such as Dr Foster had business models around fiddling the figures. (There remain questions as to whether Mid-Staffs was “worse” than other Trusts, or whether it simply didn’t pay Dr Foster the required “consulting” money for “advice”.)

With the advent of e-records, statistics can be routinely published – possibly daily – about care pathways. Gaming a small number of annual statistics is one thing; doing so to a whole nation’s medical records is quite another.

Such processing of administrative data into statistics would be done by NHS Digital as a trusted⁵ data controller, and could and would be published as part of that process.

What statistics should there be? Over time, and provided adequate resources, any that anyone believes would be useful; on the condition that there is no requirement to collect extra data into patient records in any hospital.

As all the data would be processed by NHS Digital, there would be no extra burden on hospitals – and no new data would be going to NHS Digital, as a data controller, for *operational* purposes. NHS Digital can produce only those figures that it has been asked to produce – and data controllers would be handing data over on that single, very specific basis: the production of statistics.

If a new question needs to be asked, then new statistics can be produced and published by NHS Digital to answer it. The goal being to count and compare flows, and not to count patients.

NHS England learned about inappropriate reuse of existing data once under the care.data programme. It seems the current middle managers have forgotten that lesson. They may have to be taught it again.

⁴ Or in NHS Narnia.

⁵ cf. Onora O’Neill on trustworthiness: https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust/transcript#t-295586