medConfidential's ~~short~~ commentary on the SMC letters about Biobank

Dear Fiona, Simon,

Encouraging easier research is not the same as wanting cheating, laziness, or 'god mode'. Biobank's Chief Executive told their participants in a personally signed note that UK Biobank "will conduct a comprehensive and forensic Board-led investigation of this incident", a Report now published that it is neither comprehensive nor forensic. Lord Vallance told the House of Lords that "importantly, this needs a complete and robust response", and DHSC insists this event is Being Taken Very Seriously – medConfidential looks forward to seeing meaningful evidence of that.

We have commented on the report elsewhere, including the negative consequences of the current situation for researchers and research incentives – but one could compare this Report against the 2014 Partridge Review, where concerns and risks were taken seriously, and meaningful changes resulted in the use of NHS patients' data. Here we pick up on the issues in your exchange, or those notably omitted. Did Biobank tell you data was still online?

The public are exceptionally forgiving of the NHS and of research, meaning public trust is eroded not by legitimate mistakes that get corrected but by deliberate word games that claim one thing and do something contradictory. Biobank's series of ineptitudes aren't important because of Biobank, but because they convinced the now-former Secretary of State to take a "pandemic only" dataset and reuse it however Biobank and others wish – it is unclear what SMC, HDR, DH/E and others expect to happen with promises made during the next pandemic.[1] Similar promises are being made by the Department of Health in England[2] around the Single Palantir Record.

"The protection worked" is one line from your exchange – except, clearly, it didn't. The supposedly secure environment had a 'download all the data' function, and the data leaked. Not just a "few bad apples", but hundreds of instances that we know of as well as 500+ that Biobank admits only privately (did they share that briefing with you? The one where they blame Biobank victims for any disclosures because they knew what they signed up for?) and an unknown number beyond that. The data remains loose on the internet as I type these words. The only reason the "Yale" "breach" is known is that the students posted their papers to github as modern researchers do, or rather they did so in 2024 when that course ran. UK Biobank senior staff knew this was possible in the environment; they kept it from others.

If rules are to mean something to participants, they must be followed – even when breaking them is advantageous. Clause 6 of the Health Bill allows the NHS to spend budget on prizes for innovation. Will prizes go to people who cheat? Would SMC give them cover?

Communications have to be true. More voices are definitely good – the NHS FDP / Palantir debate is better because ITCrowd Tom is engaged – as long as what they say is true. The UK Biobank board member who talked about UKB following the five safes framework probably feels deeply misled. Biobank allies argue it would cost £1.5m to download all the data, that's all the raw sequences and every raw image; the biggest possible figure. If that were remotely true, Biobank would be able to tell who had downloaded what simply by looking at the invoices Biobank issues – clearly it wasn't true. If one works out the cost of the detailed, rich NHS data, the questionnaires, and genome/proteome summaries, the cost is closer to the price of a pint (and not even in London). It has all leaked. The UK Biobank Report doesn't inform anyone about that, despite what Lord Vallance told the House.

---

[1] We exclude GeL from this list, but note Q123 of the Lords Sci/Tech Inquiry on Innovation in the NHS.
[2] The post-digestion entity that is currently NHS England and the Department of Health & Social Care.

In the House of Commons announcement, the Minister explicitly cited Yale University as having broken the rules. Biobank's CEO named them to the BBC as an example of the steps Biobank were taking against abuses of data, using Yale to argue that this wasn't just a China incident. That was a deliberate choice. Yet the Review is so constrained that it does not look at Yale at all.

medConfidential waited to engage until there was evidence in the public domain. Your exchange claims that the UK Biobank team were open and honest – something the report repeats insistently – but is that true? UK Biobank has insisted to some that downloads were turned off without notice in 2024, while they gave notice to users and it was a headline in the specialist press after a media interview, while there was still a working download data function in their mandatory environment. In effect they swapped one download service for another. Time and time again, Biobank makes a claim, and it's wrong.

Some information had been disclosed to Parliament, there's a note to participants, and many unanswered questions that Biobank addressed only with private reassurances. Of course they were all good chaps, and this would all be fine. Colleagues misleading each other is academic research competition; misleading participants and breaching NHS contracts is a different thing – both of which, we now know, happened. Aspects of the exchange resemble a cosy common room chat not the evidence-driven, honest advocacy for which SMC is renowned, where you debate the thing you want to debate, not what actually happened. Fiona is entirely right to be frustrated.

Biobank and HDRUK share a culture that's spreading

Biobank isn't important to the public because it's Biobank, or because Biobank misleading DHSC/DSIT caused Ministers to misinform the House, but because it's the canary in the NHS data coal mine.

When the Cancer Registry gave data to a tobacco company, there was accountability and the cancer registry continued under new leadership with better governance. When Biobank gave data to eugenicists – a project directly resulting in 'designer babies' via a clinic in London (via a loophole in the law) – there is simply an insistence that blame lies elsewhere.

Similarly, the argument that 'privacy campaigners queue up' to care about properly consented studies is just flat wrong – but then SMC has no interest in hearing from medConfidential (again? See footnote 1). We too make a cost/benefit assessment and get why it's easier for SMC to listen only to those who will be supportive, as this only sometimes does goes wrong, as with Biobank – in the same way one spoonful of sewage in a barrel full of wine is rationally mostly wine and only some people will get sick.

The Biobank scandal is simple – Biobank said some things to their cohort, and agreed contracts with the NHS on that basis, and then didn't do those things, and insisted it was fine when things were clearly otherwise. We can ignore 2026 events entirely – Biobank leadership ran the same playbook about giving data to insurers. The Biobank playbook is well-worn, which is why it is hard to argue they're learning lessons.

UK Biobank continues to put very little information into the public domain or even to their participants. Nowhere have they disclosed what "categories" of data got lost – UK Biobank refuses to tell their participants or the public. Is there anything they didn't lose?

medConfidential doesn't fundamentally disagree with the position you both genuinely hold about the importance of research – but we ask an adjacent question which your exchange dances around: what do you do when the promises you're relying on turn out to be untrue? Or, more classically, it's not what you don't know that kills you, it's what you know for certain that just ain't so.

Biobank and others should keep promises to participants. The proposal for an NHS App-led recontact and potential new follow-up wave give them a way to reset. But such an arrangement doesn't scale to the entire country, despite what some institutions might want.

Biobank and their allies have made a series of choices over time, the question is what's next:

Honest mode

Our friends at UseMyData have a line "Say what you'll do; do what you say" – and with a consented dataset like UK Biobank, a project can do anything they get proper consent for. If people are happy for their data to be used in eugenics, insurers, and for download, then it's up to them to make an informed decision. There are some nuances around what that means in various contexts, but it's all entirely doable.

Everyone knows what that looks like, and everyone knows what that means. It's the default assumption that most people have about most things most of the time – it's the assumption medConfidential had before Biobank's CEO asked us to take a look at them back in 2023.

Honest mode is the default operation of direct care in the NHS – where, if your surgeon didn't tell you something material until after your surgery, you get very unhappy. Vaginal mesh being exhibit A, and there's a long, terrible list of what in finance is misselling and securities fraud. One reason the public are so disturbed by NHS staff looking up the records of terrorism victims (and then covering it up) is not that they thought the staff couldn't – the default expectation is that honest, decent people wouldn't – but that it happened. (Similarly, for data uses, the NHS National Data Opt Out should let people who don't accept the default risk exclude themselves, since, as the UK Biobank report shows, this is all personal data).

This is where everyone expects the NHS to be, and there's some pull towards that from organisations that want NHS data directly from the NHS. If a project doesn't want to meet NHS national data standards, it can go to the patients and get the data directly – but effectively no one does because too few patients will believe such promises. The NHS reasonably expects that what projects tell them will be true, and writes contracts on that basis. Contracts UK Biobank ignored.

NHS England started publishing Data Uses Registers after 2014, and for a decade there was an undisputed evidence base about who had what data and why – until HDRUK tore up that arrangement in Covid for their own gain. A definitive, dated, spreadsheet of what projects had access to what data when acts as a clear evidence base stating exactly who had what, in addition to the project webpage which describes (some of) the current state.

A middle ground for the NHS is a Data Opt Out that's good, which acts as a counter balance against the inevitable drift to Lazy mode.

A slow drift to Lazy mode

Lazy mode is where someone makes promises hoping never to get caught not keeping them, or where they convince themselves over time that they're doing what they said, even when it is objectively obvious otherwise. UK Biobank has a list of projects on their website, which until recently included companies that went bust back in the 2010s and supposedly kept filing their annual project updates to maintain a copy of the data. This may be why the UK Biobank website has the bare minimum about a project – why not disclose which data categories were accessible to whom?

A project saying we "will make your medical history available across the internet, with your DNA summaries, proteome, assault history, and sexual lifestyle questionnaire" is a project fully consenting adults could join. I'm not sure how many would choose to do so, but that's up to them. It's also what Biobank did with that data.

Some academics argue such open access should be the default state because it benefits academia – if a doctor types something into a patient's record, their stalker at a research organisation should be able to read it. The "open data" community has long been aware that some mixture of laziness and greed would mean someone suggested it as a cover for their "saunter down cockup boulevard".

But that's the situation Biobank has drifted into, and the situation in which statements to Parliament were made and responded to. We would hope UK Biobank's leadership explained that situation to everyone who rallied to their defence in the House of Lords debate. Hopefully they either know, or both didn't know and don't care, before making admissions published in Hansard while their sexual activity questionnaire remained downloadable from the internet, protected only by a pseudonym, attached to a special unique health fingerprint anyone can read about on their Wikipedia page. Time will tell whether those who claim to be happy with this actually are.

It would be wholly different to impose that on every person in the country as a prerequisite for healthcare. But that's the strawman Biobank and colleagues sometimes create to cover up what is at best neglect or possibly outright cheating.

Engaging Cheat mode

Cheat mode appears as the default behaviour of UK Biobank over the last few years – the litany of press controversies stands alone.

When Biobank and friends blame medConfidential and journalists for the consequences of UK Biobank's failures, they echo the character in Cuckooland (the author of which, notably, wrote some of the Biobank stories that Biobank excluded from their Report). It's easy to decry the obligations of Honest mode, to want it to be easier, to cheat in small ways, and then large, until the cartel frogmarches toward wanting to play god over patients' lives and use data however they want.

When cheating once has been undetected, the incentive is to cheat ever more. Reading page 45 of the Report, the Five Safes are described as “Safe outputs: all researchers’ publications are reviewed” or “Safe settings: data were provided under an MTA with strict requirements for data security” – which are not the definitions that the report itself links to.

Cheat mode by one organisation creates opportunities for others. We do not need to have seen any evidence of any s170 DPA18 breaches (and despite what Biobank leadership may suggest, we haven’t) to make the clear case that it is may be unwise for Peers to announce in Hansard that they are in the dataset that Biobank knows remained uncontrolled and at large on the internet – especially when some of those individuals have “special unique” health events described on their Wikipedia page allowing spontaneous reidentification.

Similarly, leaking both proteomes and data derived from genomes may finally evidence what the proteomic research community has been warning of for years – that it’s now trivial for Claude to write code to take a published “anonymous” proteome and match to the “identifying” genome that coded it; an act which may have consequences that cannot be mitigated by blaming “a few bad apples”. None of those are consequences for Biobank itself, only the participants (and other institutions).

Polly Toynbee’s piece in defence of UK Biobank argues that she doesn’t care whether Biobank were honest; she just wants the research to go ahead. Which is an entirely legitimate opinion for her to have, but it is not the deal Biobank struck with participants. (Polly’s access to the Biobank leadership is also not normal.) Polly may not have believed a single word Rory told her about safeguarding her data, but her data should be used in line with her wishes, just like everyone else's. It’s the patients / participants who should have agency here; agency that Cheat mode and God mode take away from them.

Demanding God mode

Simon opens your exchange by posing the question “who decides?”, which is then promptly ignored – as if researchers alone decide what happens to data they have. A view that is not altogether uncommon in some circles.

God mode is when the researcher decides to do whatever they wish, with no restrictions – those who have a copy of the data can do with it what they will. Biobank’s failures give lie to another research fiction too – that if only there were no rules, then feeding all the data to a smart postdoc fuelled by caffeine would cure everything.

The UK Biobank datasets have been floating around for years. So if a dataset with maximum richness and no rules is the requirement, where are the papers? All of the data has already been fed to Claude by any researcher worth their salt who has a copy, but there’s a complete lack of anything substantive to show for it. Where are the outputs?

When HDRUK ignored the Covid-only restrictions and did whatever they wanted with the data they had access to – they ran 100+ projects before they were caught – what they have to show for it seems to be precisely zero. Where are the discoveries?

Meanwhile, RECOVERY followed all of the rules and succeeded.

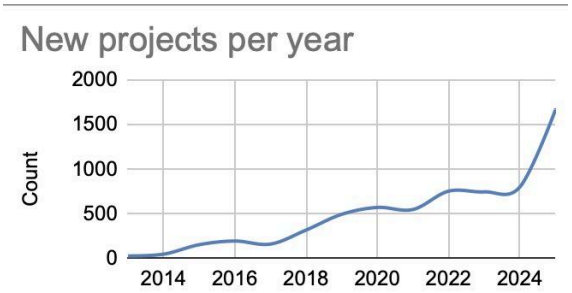
Some researchers dream of a discovery that earns them a trip to Stockholm. Others claim they will come up with something so novel and so ground-breaking that patients across the NHS must be treated differently, immediately – without consulting their responsible doctors, and without any burdens like publication, peer review, external input, or even validation.

So a single PhD with a caffeine overdose should be able to change how patient treatments are made in clinic tomorrow – as an aside, this is how changes can be made to the Palantir / FDP Cancer360 product – as if all of the things that made RECOVERY safe should be abandoned because one academic thinks they know better? And no-one should be able to test it? Doctors will remain clinically responsible though...

Those who follow the rules do good work, while those who cheat in one area fail in another and seem to have nothing to show for their cheating besides a lack of embarrassment at getting caught. Welcome to Cuckooland.

God mode is epitomised by the Scottish Health and Wellbeing census which misled parents about the "sex survey", proudly advised by HDRUK and ADRUK. Researchers are also right to be worried about the Adolescent Health Study. Will AHS leadership ask teenagers and parents to hand over their most sensitive mental health experiences and then put them on the internet because promised safeguards are a barrier to research?

The leadership of that project worked for Biobank, and has now gone on to do something "a bit like UK Biobank, but in a different age group, so in young [people]". We observe that Biobank's change in approach and acceptance of new projects followed the MRC meeting that resulted in the AHS and Our Future Health studies being born, instead of a new wave of Biobank.



Biobank leadership repeatedly panicked in the face of competition, and participants felt the consequences. What matters to their participants is how they choose to respond.

The most recent admission of that approach is "embryo screening" for IQ, which validated their work on Biobank – work Biobank says is fine and still "current". As the most high profile and politically connected Biobank project in recent years, perhaps Rory will invite them for a research conference keynote? With a nice SMC explainer? No?

God mode thinking inspired a monologue from a former leader at both HDRUK and Biobank:

> "...through emergency legislation, there were improvements in access to linked data for crucial Covid-related analyses during the pandemic, with both legislative and regulatory changes that improved access. However, several of those have been withdrawn,despite clear demonstrations of benefit and the potential for similar data access mechanisms to similar datasets to actually provide benefits of equal, and indeed much greater, magnitude, to prevent, diagnose and treat many other conditions that one might also regard as global pandemics or UK-wide epidemics or pandemics—for example, heart disease, stroke, cancer, diabetes, arthritis, dementia, mental health conditions and more."

If you had asked a schoolchild why they were acting differently in April 2020, they could have given you an age-appropriate response. RECOVERY worked, as the entire planet focussed on a single challenge. To claim the same need for “emergency legislation" on the "global pandemics” for the conditions listed is demonstrably disconnected from reality. They are chronic health conditions; they will benefit from more research, but the argument that public wishes should be ignored is as disconnected from reality as other ‘theories’ around the pandemic. Institutions must be consensual, safe, and transparent because they expect to be seen as trustworthy into the future. It’s unclear what UK Biobank expects.

Hard mode is Consensual, Safe, and Transparent

The NHS is expected to run in Hard mode: to be honest, not to cheat, and to cope with patient expectations that may be contradictory. If a researcher wants data from the NHS, they have to accept the constraints that come with it. Institutions (especially new ones like HDRS today) will always be criticised for having any rules or constraints – researchers and especially industry will say if only the rules didn’t apply to them they’d do wonderful things. Promises made to patients have to be kept, and some complaints are indistinguishable from the line “I’m not good enough to succeed at my goal, let me cheat”. If a user wants Easy mode, the private sector will sell them data without constraint,[3] but there’s a reason more detailed data comes with constraints…

People have busy lives, and the NHS should ‘do the right thing’ with secondary uses of the records created for direct care. The opt-out arrangement means the NHS takes most of the responsibility, and those who don’t accept the risk shouldn’t be exposed to it. That’s not enough, however, for those who believe ever more data is necessary – a demand that can never be satiated.

As the previous National Statistician found, public bodies may think they can use data however they want, irrespective of the public’s views – but at some point they have to go back to the public with questions, and then the lack of public confidence destroys the economic statistics. The public has to believe that an institution’s word means something if one expects them to provide honest answers. The “mandatory, questionnaire-based, whole-population” census is under five years away.

Your discussion cited the concerns of ALSPaC, where the researchers go back to their cohort periodically; a cohort that has been actively involved for most of their lives (all of their lives for those who started at/before birth). As with doctors treating really rare diseases, the ALSPaC researchers get birthday cards from their research subjects, and get invitations to their weddings.

Saying ‘no downloads’ while simultaneously allowing arbitrary downloads is anathema to everything ALSPaC has spent decades working towards. Every staff member there would viscerally feel the betrayal of a Five Safes environment claiming only “Safe outputs: all researchers’ publications are reviewed” (other outputs are not checked, and participants’ data was leaked to the internet). We know how Biobank’s management team felt about the cover-up – they appear mostly offended they got caught. ‘Careless people’ in positions of power have created a culture where Biobank leadership has drifted from Honest to Lazy to Cheat and into God mode.

The test is not whether Biobank’s actions were ‘defensible’, or whether their counsel wrote a report absolving the organisation that pays him to review decisions. The test is whether UK Biobank did what they told participants, and what their contracts required. Biobank made promises and assurances to its participants and to NHS England, and then broke them.

[3] Or the London SDE, but that’s a mess for a different day.

There's no debate on whether UK Biobank misled their users; the argument is only whether it matters. It's unclear where you come down on that. Everyone is welcome to whatever position they wish to take – but there's a difference when one imposes the consequences of those actions on others, while claiming it is for their benefit.

As comms professionals, you're used to being brought in only at the end of a process to 'explain' to the public why they must accept the toxic decisions that have already been made – or where success is dependent upon material omissions, and collapse occurs when there is awareness; a model which has proven catastrophic time and time again ('Comms mode' has been tried, and is currently the only approach of the FDP project, who forget that patients have views that must be respected; lazy, cheat or god modes all eventually fail. It may be in the interests of a patient to have that blood transfusion, or have a particular course of treatment, but they decide what they want, and the NHS has to deal with that. Much of the FDP debate comes down to a god-mode insistence that DH/E can and will do it anyway, and patients should just be quiet and accept it – noting the author of that piece also sits on the Biobank board with data still online).

As the PHG Foundation notes, "Trust does not depend on the absence of risk. It depends on honesty about where risks exist, clarity about how they are mitigated and accountability when safeguards fail." UK Biobank may be able to claim their users accepted all those risks; the NHS cannot.

Sam