medConfidential response (5/11) to Biobank's 30/10 responses to medConfidential questions on biobank

The biobank obstacles PDF says:

> "all participants have been receiving regular updates about UK Biobank"

Is that statement accurate? Do biobank have details of *all* participants? If so, why did Rory tell me biobank does not?

Given the misstep saying the BMA supported something they don't, isn't accuracy, at this point, especially important?

> "These data are held securely in UK Biobank's database"

Will biobank update the project list webpages to say which projects access physical tissue, and which projects receive disseminated data to their own premises or environments, and which are biobank-TRE only? They may be mutually exclusive, but not necessarily - I don't know how biobank protocols have evolved over time.

> "Health conditions that are managed largely outside of hospital by general practitioners (such as arthritis and other causes of pain, dementia and other neurodegenerative conditions, impaired vision or hearing, respiratory conditions, heart failure and mental health problems) have been systematically under-represented in large-scale epidemiological studies. Consequently, securing access to the coded primary care data for the participants in UK Biobank offers an unparalleled opportunity to redress this imbalance, enabling greater understanding of ways to prevent and treat a wide range of conditions."

I include this quote here mostly for everyone else to note the opposite argument being made about GP data by NHS England for Palantir. One of those arguments is wrong, and I think I agree with biobank.

> "Coded primary care data (which contains data on diagnoses, prescriptions, referrals, laboratory test results, **but not any free-text or attachments**)"

To clarify, is biobank disagreeing with Cathie Sudlow's argument in the 10/10 workshop that anything without free-text was unacceptable?

> "Coded primary care data were obtained in 2017 from TPP"

What happens to that data *today*? Or recently, for this paper for example.

> "...with considerable efforts made to make the process straightforward by providing each Practice with"... "the percentage of Practices that agreed to the data release never got above 20%." … "using a more streamlined request with endorsement from both the RCGP and the BMA. It achieved a success rate of about 10%."

It's understandable that biobank put a lot of effort into some things that didn't work, but do biobank acknowledge that it didn't work? Do biobank have any insights into why? It may be

because biobank were largely asking GPs to trust what biobank say, when what biobank say appears not to be entirely trustworthy. We'll come back to this with something missing from this document.

> "Obtaining coded primary care data during the pandemic…"

Does biobank recognise that the restriction of "COVID-19" only purposes may have been *why* the COPI notice was acceptable, and biobank is having different problems?

> "Over 200 papers have been published using these linked healthcare records, many of which have used the primary care data to investigate the role of co-morbidities and medications as determinants of severe COVID-19 and to identify ways to mitigate its impact

It may just be wording, but I would hope that *all* those papers under COPI looked at COVID-19. If there were uses outside of the remit of COPI, who is that regulator? (I understand the ICO has my original link already).

Is there a list of those 200 papers? Something like the OpenSAFELY list. The list of all papers on the biobank website is uncategorised (if you're looking, just use pubmed).

> "Following the failure of the third pilot study, UK Biobank engaged with the National Director of Transformation at NHS England (NHSE) and his colleagues, who proposed that a further request should be sent to all Practices in England directly through the NHS systems. A revised draft letter was created with input from the NHSE Communications team"

NHS England communications team getting GP data badly wrong? Surely not?

But nice of biobank to demonstrate that NHSE have learnt nothing from care.data, GPDPR, Palantir, etc. We'll come back to this, but for now, someone should suggest they read DH's independent report "Better, broader, safer: using health data for research and analysis". How does biobank satisfy the standards in it?

> On the letter: "The response to this request for coded data is being monitored via the EMIS and TPP systems providers."

Should we read anything into the fact that the numbers biobank put in the other document are missing from this one?

> “Proposal that NHSE issue a statutory notice for UK Biobank”

OK… On what basis, and for what purpose?

> “Following the termination of UK Biobank’s COPI notice for COVID-related research, no further coded primary care data are being made available for the approximately 450,000 consented participants registered in General Practices in England. Practices have been reminded by their system suppliers to turn off the data sharing agreement for UK Biobank, and no further data are currently being sought.”

Again, it might be wording, but can biobank confirm that biobank have ceased to give access to the data biobank received under COPI?

> “Access to primary care data of 59 million non-consented patients for COVID-related research is still occurring within OpenSAFELY following the issue of a Data Provision Notice (DPN)”

I’ve not yet seen that figure, but assume that biobank know better than I do, that it appears that, for every practice in the country, GPs have accepted that the process agreed by JGPITC, BMA, RCGP, DH, NHS England, is acceptable to them. So clearly there is a process that works when it’s followed..

Does biobank recognise that biobank has not yet met that standard?

Principally, biobank, and relevant to the Sudlow Review as well, is that openSAFELY contains a series of governance steps that mean that even if a GP does “push a button”, there are safeguards around the process for all GPs.

> “Access to primary care data of 59 million non-consented patients for COVID-related research is still occurring within OpenSAFELY following the issue of a Data Provision Notice (DPN). It is understood that the DPN may be extended for research on all types of health conditions. This is a welcome development as OpenSAFELY allows some types of analyses to be conducted that are of enormous value for improving health care delivery.

medConfidential welcomes biobank support for OpenSAFELY. Consensual, safe, and transparent data flows are something that can benefit the NHS, benefit biobank, benefit research, and benefit patients. Hopefully Biobank at Oxford will work with others, as OpenSAFELY at Oxford has, to get to a position where it is demonstrably telling the truth, satisfying everyone and receiving data that satisfies all parties, rather than imitating the approach of the FDP.

> “Given the consistent lack of success in obtaining agreement from individual General Practices (despite guidance and support from the RCGP, BMA and JGPITC over a long period of time), and the demonstrated success and acceptability of the COPI notice, a suitable central instruction for the release of coded primary care data is needed for UK Biobank”

Do biobank currently have JGPITC and BMA support? Or are biobank claiming something biobank haven’t got (again)?

> “Such a central instruction from NHSE could come in the form of an extension of the COPI notice, the issue of DPN (as with OpenSAFELY) or some other suitable instrument. It would provide a clear and unambiguous instruction to General Practices, which they could be appropriately reassured had been properly considered and reviewed;

What would that COPI notice be for? And issued under which part of the COPI regulation?

And, if that can be answered, why would GPs press the button accepting the DPN for a dataflow under the notice since it’s likely to be identical to the current provision?

COPI did not oblige anyone to do anything, it simply said that they could, for a very specific purpose. I do hope biobank has not misunderstood what biobank were legally permitted to do with the data…

> “General Practices should also be reassured that they would have no liability or exposure related to such an instruction, which could come through a statutory confirmation or an indemnity from UK Biobank (which it is willing to provide).

On what legal basis can biobank possibly do that? And will the University of Oxford put that in writing? If biobank liability is £80k x 6000 practices, is that something Oxford is willing to stand behind, or are biobank expecting DH to pay the fines?

The current biobank webpages

> “Ensuring that GP practices are not exposed to liability
> “UK Biobank fully appreciates that GP practices take their responsibilities as a data controller very seriously and are rightly concerned about their potential risks and exposure involved in making this data available to another data controller (UK Biobank in this case).”
> “In this regard, UK Biobank has provided a number of practical assurances, including that only de-identified coded data will be accessible by researchers and that the data will remain on the UK Biobank Research Access Platform (where the data is stored within the UK, on the AWS cloud).”

> "Further, UK Biobank would be prepared to take on the full responsibility for the appropriate and lawful use of the primary care data by UK Biobank and its researchers, such that UK Biobank would assume responsibility for any regulatory issues and also ensure that GP practices would not carry any financial exposure: an effective indemnity."

What if the biobank have not demonstrated to the Profession that biobank processes for data dissemination are good enough to satisfy their obligations?

"More is more" may be an approach acceptable to biobank, and a reputation risk biobank wish to run, but biobank need to demonstrate to the Profession (and likely others) that biobank consent is better than Our Future Health, the self-styled successor to the biobank, who are currently offering a £10 shopping voucher for patient dna and lifetime medical records via a mailing list they bought off the data broker market.

> "How do I know which of my patients are UK Biobank participants?"

Biobank have answered the question for an individual GP – they email biobank, biobank email them back the list.

But that is not the process by which the list of patients for a practice flows between biobank and TPP+EMIS. Do biobank just email TPP+EMIS a list of NHS numbers and the data flows? What's the governance of that list?

I'm not suggesting biobank's answer is wrong; it's just an incomplete understanding of the questions biobank are being asked, because, as is evident from all the press, biobank don't understand the barriers biobank are facing and seem to be refusing to engage with the detail.

There's a pathway forwards, biobank just seem not to like it.

> "UK Biobank can assure GPs regarding consent, data governance and protection (see below) and can ensure that GP practices are not exposed to liability."

Biobank say that, but don't say how. Would those assurances satisfy NHSE or DH if biobank were getting data from them?

> "Despite our previous efforts over the years, we've only had the GP data during the pandemic, which showed us its importance for understanding how to prevent and treat disease."

Is there a reason biobank don't mention the GP data from pre-pandemic that biobank seem to still use? (I agree biobank shouldn't mention it, but biobank also shouldn't be using what was received without a legal basis, as biobank may wish to confirm)

On “Flying Troika”, it’s still unclear what data they received and how, but the blasé acceptance of US shell companies suggests GPs may be the only ones doing scrutiny of biobank at all.

The other PDF

> “How the breadth of UK Biobank data leads to identification of disease subtypes”

I’m sure Our Future Health will be glad of your support for their position, but there’s a question of how those findings get fed back to everyone else. How do you propose doing that?

> “The main reason behind their lack of agreement is not known for certain but, as well as being busy and having other priorities, it appears to be largely due to concerns about data sharing responsibilities. Experience with the COPI notice reinforces this conclusion, since it removed the need for GPs to be responsible for the data sharing decision, instead relying on a Secretary of State instruction to the EMIS and TPP system suppliers to make the coded data directly available to UK Biobank. Apart from a few Practices that sought further information, no issues were raised with the primary care data being made available to UK Biobank in this way.

The context and sole purpose of fighting a pandemic probably had nothing to do with it?

> “It has been proposed that the OpenSAFELY platform could address UK Biobank’s needs (and those of other consented cohorts) by applying algorithms within OpenSAFELY to combine coded primary care data in order to provide derived health outcome codes to UK Biobank.

medConfidential has never seen the proposal as described here, and we would be interested in seeing it.

There are workable pathways for OpenSAFELY with consented cohorts, and there are many unsuitable or unworkable other pathways; biobank appear to be correct that the route biobank describe would be in the latter category and would be difficult for anyone to support for the reasons outlined by biobank.

Other pathways forward are available.

> “Leaving aside the lack of scalability of the OpenSAFELY approach (e.g. contrast the approximately 20 publications that have emerged so far, largely authored by the internal team, according to the OpenSAFELY website versus the 2,000 papers published based on UK Biobank by external researchers globally in 2022 alone)

While we’re reluctant to do anything with internal Oxford rivalries (beyond watching them[1]), neither statement in the original is backed up by citations, and we’d like to see evidenced numbers.

[1] GDBO etc

> “However, UK Biobank has demonstrated in a series of pilot studies that securing Practice-by-Practice agreement cannot ever be successful”

Except when, as biobank point out earlier, other entities have been able to do it.

> “Consequently, UK Biobank proposes that NHSE issue a DPN (or some such central instruction) for provision of the coded primary care data for its consented participants registered with Practices in England.”

COPI was practice by practice, DPNs are practice by practice. The “or some such central instruction” suggests biobank has not understood the problems of the Profession or the available tools of NHS England/DHSC sufficient to provide a pathway that satisfies everyone.

Or is Biobank proposing a mechanism by which NHSE could simply take all GP data (as part of FDP?) and hand a copy to biobank? Is that the policy biobank is willing to advocate for?

> …“would reduce data governance concerns for General Practices, and would substantially enhance the ability of UK Biobank to support a wide range of innovative health-related research in accordance with the participants’ wishes”

Perhaps data governance concerns, expressed primarily by JGPITC, could be heard, addressed, to the point that JGPITC supports this project as they have with others, many times before, and no extra-legal options need to be contemplated?

If every stakeholder retains a seat at the table, then no decision now is irrevocable. If biobank want an irrevocable mandate, which biobank don’t even have from participants, that is a higher bar than biobank have currently been able to satisfy.

In the shadow of the palantir procurement, Polly Toynbee wrote a nuanced piece about data languishing after she donated to biobank. As biobank finish up WGS of participants for access by biobank’s customers, I presume biobank board is satisfied that the “more is more” approach addresses the complexity of concerns expressed in that piece, but what happens if not?

medConfidential