Annex 4B: The consolidated list of unanswered questions for UK Biobank (early March 2026)

In 2023 medConfidential met with (now outgoing?) UK Biobank CEO Rory Collins, where he asked us to take a look at UKB's processes. For a decade we had largely ignored UKB as off in their own little world, but invited to take a look, we did. Rory immediately withdrew cooperation, and over time starting in 2023, which resulted in this paragraph in the first output:[1]

> "In the absence of requested material, we superficially and briefly took a look at the material on the website. It is that public material which forms the basis for this interim summary of what we found. We would have shared what we found to biobank only, but given the lack of candour from biobank to others recently, this assessment is public."

medConfidential has now published 5 different sets of questions. Some have been answered because of the actions of others, some have not, and some have simply become irrelevant and out of date. UK Biobank have removed some information from their website which changes some context, and some of which now need edits in the light of 2 years of events. A complete list of current questions is now below based on where they were first asked.

GP data is already in Biobank – received in the 2010s when they had no legal basis

October 2023 – Biobank 1[2]

Biobank's documentation says they have GP data on 230,000 people, saying: (emphasis added)

> **"UK Biobank has been liaising with various data suppliers and other intermediaries** (including the main primary care computer system suppliers in England) to obtain primary care data for UK Biobank participants, all of whom have provided written consent for linkage to their health-related records. To date, **coded data have been obtained for approximately 45% of the UK Biobank cohort (~230,000 participants) and are now available** as part of this interim release. Details of the data providers and the coding schema used are summarised in Table 1.
>
> **UK Biobank is currently in the process of securing access to data for the remaining cohort, mainly for participants registered with EMIS practices** across England."

The last comment implies that it is TPP who is the one of the "other intermediaries" who acted as a data controller in determining data should be released to biobank. If the data controller had agreed, the current complaints about GPs not agreeing would clearly not be happening.

Table 2 of that document says biobank have 87m clinical events supplied by TPP and Vision.

1. **What data does biobank have, on what legal basis, what have they done with it, and to whom have they given access to GP data?**

[1] https://medconfidential.org/wp-content/uploads/2023/10/biobank.pdf
[2] https://medconfidential.org/wp-content/uploads/2023/10/biobank.pdf

We understand biobank has previously been told by JGPITC that what they were doing should cease.

2. **What did JGPITC tell biobank? What did both parties do afterwards?**

[2025 note: A September 2025 paper promoted by Biobank shows the data was still in use as of then]

Consented Cohort questions

3. **Is there an existing process by which a patient can (re)discover they are in the biobank? Can this apply to any other "consented" data flow?**

4. **How would biobank like to start to rebuild the perception of trustworthiness that they have undermined with their actions?**

Biobank's Application process

The complete application process for NHS England's DARS is often described as long and arduous, and examples I've seen run to tens of pages and others are longer (it varies due to different questions being asked based on earlier answers).

In contrast, the pre-2025 biobank application PDF is public and three and half pages long, and the 2025-onwards public version is only *one page*. This new 2025 form asks questions like:

> "A4.1. Can you confirm that you intend to use UK Biobank data to perform health-related research? We define health related research as any research which ultimately aims to improve human health.
> Yes Not Sure"

Decision makers can not take information into account that biobank never asks for - even if biobank's decision making on receipt of an application were to be perfect, the application form did not *ask* many questions in 2024 and seems to ask even fewer now.

5. **Does the biobank application form satisfy all the obligations of biobank to a) their cohort, b) each organisation who provides data to biobank, and c) GPs?**

medConfidential have been told, possibly not correctly, that the Board is provided with copies of all applications.

6. **Are the Board satisfied with the application process they oversee?**

7. **Relating to the answers to the previous two questions, and applications which have raised questions: what did the Board know, and when did they know it?**

8.

Biobank projects

9. **Did Biobank change anything after media coverage regarding China in 2022?**

A decade after the first research projects, biobank's website said they had around 3489 projects of which "40" are "complete". Our efforts seem to have prompted a cleanup,

10. **Have biobank kept all promises made to AGD/IGARD and predecessors?**

Biobank publishes a list of projects on their website, making available only the applicant and lay summary, alongside an application number. Those application numbers are not sequential, but have large gaps, which may or may not be a problem, or an artefact of the process somehow. We have turned the project list into a spreadsheet here.

A full analysis of all users will take investigators some time, initial examples showed why a list of latest projects presented to all biobank members via the app might be politically difficult.

We did a brief look at projects largely at random: the second was Babylon (bankrupt and being sold off for parts), the third seemed also to have pivoted (no obvious biobank connection), and the fourth seemed to be… superficially fine. With a success rate of 25%, this appears like a job for someone with access to US corporate registration records more able to recognise the signs of deception. Sometimes, those signs are pretty clear.

The first project we looked at was Flying Troika.

Flying Troika

No information is (still) provided by biobank on whether a project got some data in a TRE, or whether they got access to genetic samples, etc, because biobank never publishes that information.

FlyingTroika.com, registered to the applicant, describes "A pure AI driven research lab" offering "Flying Troika Deep Learning / AI / Machine Learning Research and Solutions".

They say "We are a pure research lab/ consulting service provider working on the cutting edge Deep Learning Technologies and Algorithms a variety of industries: Pharma, Medical Imaging, Finance, Insurance, Health Care, High Tech, Retail, Manufacturing, Oil & Gas, Industrial, Auto, Construction Applications.".

Their first example is "Deep Larning in Finance and Insurance", and list "OFFICE LOCATIONS/ RESEARCH & DELIVERY TEAMS/" in "New York, Chicago, Los Angeles, San Francisco, London, Edinburg, Hong Kong, Singapore, Maimi, Paris, Berlin, Madrid, Barcelona". 13 offices in 8 countries.

At the bottom of the webpage, are the customary twitter, facebook, and github logos, with a phone number given as "AI agent".

Assuming biobank ever checked the webpage, they may have noticed a couple of typos in cities and the text, but what's "Edinburg" or "Maimi" amongst fellow researchers… The other business registered to the applicant offers subscriptions to a monthly supply of stock photos.

Had Biobank checked any other databases, they'd see that the company employs only two people, with a registered office at the same address as the domain is registered, which is a house in California.

11. **Is this applicant real? What information did biobank receive, and what did biobank not ask for that they now should have?**

12. **Would Flying Troika have passed NHS England's processes?**
    a. **Do Flying Troika satisfy any obligations biobank has to NHS England if NHS England data was involved?**

13. **What diligence was done on whether biobank or NHS data was used for "Finance" or "Insurance"? Is this what biobank participants signed up for?**

14. **Does this project satisfy all assurances that GPs were given in the letter about what biobank does to ensure good governance?**

15. **Was this organisation restricted to a TRE, or did someone send medical records and potentially genetic samples to a house in California?**

16. **How can biobank know the answers to this section if they didn't ask in the application process?**

17. **How would biobank PPIE and other engagement groups feel if this was the first example of a new project listed when they saw a new biobank tab appear in the NHS app?**

18. **How does this relate to HDRUK's strident objections to TREs for cohort studies?**

We expect a range of questions to apply for other biobank users; some of which will be similar, specifics will be different.

Areas of research with biobank

The expansion of biobank registrations in recent years seems to have been largely driven by two overlapping groups, firstly are Artificial Intelligence and Machine Learning startups, and secondly are the emerging wave of "longevity" startups.

The argument seems to be that living forever reduces deaths, so the biobank team give access.

19. **Are there other companies like Flying Troika in the list of AI/ML/longevity startups that biobank has accepted?**

20. **Is this what cohort study PPIE/patient groups think that participants signed up for?**

Due to the nature of the "longevity" industry and the lack of checks, it is likely that newsworthy entities have been given access to biobank data, and biobank has not yet disclosed what they got.

November 2023 – Biobank 2[3]

The biobank obstacles PDF stated as fact:

> "all participants have been receiving regular updates about UK Biobank"
>
> 21. **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"

22. **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.

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

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

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

24. **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?**

"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.

> 25. **Does Biobank disagree that OpenSAFELY, based down the street from Biobank, *has* got practice by practice approval in recent weeks?**

"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.

[3] https://medconfidential.org/wp-content/uploads/2023/11/biobank-again.pdf

26. **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"

27. **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.

28. **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?**

May 2024 – Biobank 3: Hostile States[4]

[In 2023] you invited medConfidential to examine your processes and then withdrew your cooperation. We continued with your public materials and published as we said we would.[5] On the basis of your previous invitation, we now add questions on whether biobank continues to send British citizens' genetic and (NHS) patient data to China or other "hostile states".

You should already be aware of the recent HM Government briefing to UK Universities about hostile states targeting UK research.[6] The biobank website suggests at least 7 of the last 20 projects that biobank has approved (or at least, as publicly described as approved) were at Chinese institutions.[7]

We understand there was no meaningful change in biobank processes after the 2022 article about biobank access in China, and similarly there was no meaningful change following our 2023 questions which included concerns about biobank making data available to a US shell company run by Russian nationals. In the latter case, we understand biobank's private dismissal was "more is more"; presumably including "more" revenue to biobank in return for "more" data to hostile states, reflecting the substance of HMG's concern.

---

[4] https://medconfidential.org/wp-content/uploads/2024/05/biobank-hostile-states.pdf
[5] https://medconfidential.org/wp-content/uploads/2023/10/biobank.pdf and https://medconfidential.org/wp-content/uploads/2023/11/biobank-again.pdf
[6] https://www.theguardian.com/technology/2024/apr/26/foreign-states-targeting-sensitive-research-at-uk-universities-mi5-warns
[7] https://www.ukbiobank.ac.uk/enable-your-research/approved-research

No one should be satisfied that there is still no clarity on exactly what patient data Biobank has disseminated to what researchers where, despite this being a constant question for multiple years. This is an answerable question to which there should be no space for concern.

With biobank's lack of transparency, and with biobank continuing to sending British citizens genetic and patient data to China, and the recent HMG briefing, we have some questions:

…on Data

29. **What data fields (from which suppliers) has biobank received payment for and consequently sent to *China* in the period of concern to HMG[8]?**

30. **What data fields (from which suppliers) has biobank received payment for and consequently sent to *Russia* since March 2022?**

31. **We have previously asked whether biobank satisfies obligations to NHS data providers in an application form which does not ask the questions necessary to satisfy those obligations. Is Biobank satisfied that it will pass an NHS England audit?**

32. **Given HMG concerns about "commercial" priorities of states to which you have granted access, does biobank satisfy all obligations to the NHS and HMG about public benefit beyond new treatments being developed that the NHS and the UK public purse must pay high prices for (as with Kaftrio)?**

…on Transparency

As biobank only makes available the bare minimum of information on projects – title, date, institution, (now status) lead investigator, and a description[9] – there is no information in the public domain whether data was disseminated beyond biobank's control to the recipients publicly listed.

33. **Which projects are restricted to what biobank terms a "trustworthy research environment", and how does biobank ensure that individuals who *use* the environment are only those whom biobank has approved?**
    a. **Who decided that the biobank environment is trustworthy?**
    b. **Does the current opacity have the confidence of biobank cohort representatives?[10]**

… on GP Data

34. **Has biobank disseminated to China any of the English GP patient records it "acquired" with an "unclear" legal basis from TPP either a) before or b) after the BMA asked biobank to regularise the legal basis?**

35. **Do current biobank practices around risk satisfy all promises to the biobank cohort and other institutional data providers?**

[8] Since that period is not public, perhaps since the 2022 China article?
[9] https://www.ukbiobank.ac.uk/enable-your-research/approved-research
[10] For the avoidance of doubt, the opacity does not have the confidence of medConfidential, but it is easy for you to rectify this.

In July 2024 UK Biobank stopped disclosing new projects on their website until sometime in mid-2025. UK Biobank had continued approving new projects, it *simply didn't disclose them.*

April 2025 – Biobank part 4[11]

It was unclear in 2023 whether Biobank checked the details of applicants, and it was shown in 2024 that Biobank does not check the details provided by applicants.[12]

Rory's public and private comments in response suggest that the only way to convince him that Biobank did what Biobank has said it has done – gave Biobank data to a project run by eugenicists – is for the eugenicists to use the data Biobank gave them for (at best) the purposes of promoting eugenics (or in the case of Biobank's user[13] itself, commercialising their ideas); perhaps UKB will offer them a research keynote at the Biobank annual conference?

36. **Reflecting the shared culture between Biobank and HDRUK, does Biobank's current leadership agree with Biobank's former Chief Scientist at the launch of their HDR/Sudlow Review saying Biobank should be "used as widely as possible" and access should be granted in "days" because Biobank have "one of the best systems" for access? Do those comments accurately reflect Biobank's practices?**

You'll now be aware from reporting by The Guardian, or from reading the biobank website, that UK Biobank sent data to an organisation *after* it had been sanctioned.[14] It has already been shown that Biobank does not check addresses, and it is now clear Biobank does not check sanctioned entities beyond the country of origin of the applying organization.

37. **When Biobank says "all researchers are checked against international sanctions lists", what does that actually mean?**

38. **What data did Biobank make available to this project?**
    a. **What data did the users download? (if different)**
    b. **Did Biobank make available the GP data received directly from TPP prior to 2018 (for which we understand NHSE is legally responsible)?**
    c. **How much notice did Biobank give BGI that downloads of data would be ending?**
    d. **Was the notice BGI received any different to that of any other users?**

39. **Why did Biobank leadership warn users that Biobank would be restricting downloads before it happened?[15]**

---

[11] https://medconfidential.org/wp-content/uploads/2025/04/biobank4.pdf
[12] Such as your eugenicist users operating from the same front address as anonymous QAnon front companies…
[13] Incorporated in Wyoming where there is no public information on who owns the company for Biobank to have checked in advance of The Guardian's reporting.
[14] https://www.theguardian.com/technology/2025/apr/15/revealed-chinese-researchers-access-half-a-million-uk-gp-records
[15] Including in press interviews that resulted in articles published prior to the deadline?

medConfidential supports Biobank's move away from high risk downloads, but any TRE has to be trustworthy and satisfy the spirit and practice of five safes (if that is the model you wish to use). We have seen a letter to Biobank from some of your users suggesting that the RAP is not an effective TRE, saying explicitly:

> "a nefarious user could trivially export data from RAP in several different ways (as the UKB access team have confirmed)."[16]

40. **Do Biobank's stakeholders believe the current RAP is compliant with "safe outputs" expectations for a NHS SDE, or the other tests of a five safes environment?**

41. **Does Biobank believe that it has an implementation of the five safes model that is robust against the militaries of "hostile states" operating within your RAP as disclosed by the Guardian?**

[2026] in this slot were various questions about what became the Biobank Direction published in early 2026 which we described as "Biobank want GPs to lie to patients" (and DH has agreed that is acceptable).

June 2025 – Biobank 5 (jointly to HDRUK)[17]

The fifth letter was jointly to HDRUK and Biobank asking them to show the letter to their patient group. Biobank did show the letter to their participant group who considered it and substantively replied. (HDRUK did not reply)

February 2026: UK Biobank have deleted lots of information from their website and told DH to demand GPs lie to patients (and DH agreed)

When UK Biobank relaunched their website around July 2025, they disclosed they'd kept approving projects without listing them publicly, disclosed previously secret pre-2024 projects that they'd never published before, and to maximise opacity they blocked web archivers from preserving the evidence — forcing medConfidential to step in and maintain our own archive and use technical means to ensure transparency (including on the Internet Archive Wayback Machine).

As an example, one project page which was public then vanished when the new site launched, and quietly reappeared weeks later. How many other projects were similarly hidden, and for how long? Meanwhile, basic terminology on the site remains unexplained: what does "closed" mean in the context of "last updated"? A simple date for when a project closed would go a long way — but UK Biobank hasn't provided one, for similar reasons that they refuse to say what data a project was

[16] https://docs.google.com/document/d/1LIUyHEq6_7x1dD8_7Cq936BRVnrcN83iZL0wyijW-eA/edit?tab=t.0
[17] https://medconfidential.org/wp-content/uploads/2025/06/biobank5.pdf

able to access and from where they connect. “Shadow applications” keep appearing and then disappearing.

Having been very keen to announce that Biobank users should use their online RAP environment (but UK Biobank absolutely refuses to say which projects they exclude from that requirement; we infer it’s a lot of them?). UK Biobank doesn't properly track which countries users actually connect from — they apparently accept that TikTok is US-based, when the named company is registered in the Cayman Islands and its staff work in China. Clearly application forms are still failing to be checked. (RAP access matters not just for Biobank, but for Our Future Health, who make explicit claims about country limitations that UK Biobank doesn't even attempt to make).

Going forwards towards mid-2026?

medConfidential will continue to watch how organisations keep their promises. As UK Biobank and others have found, shortcuts may work for a short period but sustaining them across institutional transitions is difficult. Even if decisions are made for short term gain, the truth will eventually come out, and the NHS will have to clean up the mess.

Biobank is a consented dataset – it should do what the cohort agreed to. medConfidential usually offer deference to those who give informed consent to do anything they choose, even when objectively unwise, but that consent has to be informed and respected by all sides. medConfidential has an evidence based level of zero confidence that the current leadership of Biobank will be open and honest with others, so we have to provide the information Biobank leadership refuses to. That is a direct consequence of the failure of Biobank’s Board and management.

Did Biobank participants think they were signing up to have their genome and medical history used as a real world dataset for undergraduate teaching in China? Biobank clearly argue yes – UKB claims that their cohort knows exactly what they signed up for and Biobank can do what they want. They made that claim with the insurers, and then changed their mind. Perhaps other topics will prompt other reflections.

medConfidential
March 2026