6 Jul 2025

Did HDRUK and NHS England satisfy their Data Protection Act obligations when HDRUK were (and are) processing data? (they didn't)

1. In short:
    a. NHS England says data projects appear in the data uses register – this one and ones like it do not.
    b. NHS England says that projects get reviewed by the independent Advisory Group on Data – these did not (although after questions were first asked, some did, and the majority then were "not supported").
    c. GP data projects are supposed to be reviewed by PAG, and these weren't (both at the time and seemingly again in June 2025).
    d. NHS England and HDR claims that a list of around 16 universities are data controllers, yet it is a HDR senior leader from a University not in the list who makes decisions about use and processing of patient data.
    e. NHS England's known and long ignored policy failings, combined with HDR's actions to exploit them, had the combined effect of deceiving both the public and NHS England's normal scrutiny processes – HDR knew what to say, and chose to say it irrespective of accuracy.[1] NHS England could have caught HDR's deceptions at multiple points, but they also shouldn't have had to.

   This project was likely a breach of stated processes and hence laws in multiple ways, yet NHS England's Data Protection Officer's initial "investigation" was into the most narrow question possible avoiding any topic where NHS England policy could be assessed to fail to deliver on promises to data subjects (it may be the investigation covers none of the topics so far). Indeed, NHS England policy and IG position is that project relevant processing of all medical data held by NHS England on 57+ million people involves no personal data at all and so no data protection rights could possibly be engaged, because NHSE processed the data through pseudonymisation and contracts it clearly does not enforce. NHS England is investigating itself for decisions it already believes appropriate on the sole basis that it was NHSE who made the decision. There are wider topics to address to cover whether the actions of NHSE and HDR were lawful.

2. The Pandemic GP dataset (GDPPR[2]) is a pandemic era collection of GP data for covid19 purposes only. It is the entire coded GP medical history of every patient in England (who does not have GP data opt out registered – as this is was supposed to be for secondary uses only – ie not direct care[3]). The collection and use is ongoing.

---

[1] Page 94 of the recently published NHS 10 year plan says: "The NHS never has the right to keep the public in the dark. That it so often does so reflects the centralisation of power and disregard for patient voice we identified in chapter 5." This applies to data as much as anything else.

[2] Potentially a confusing name at all for the data protection world! Formally known as GP Data for Pandemic Planning and Research, GDPPR, this was a 2020-era rapidly reconstituted version of the then-secret "GPDPR" (GP Data for Planning and Research) which was launched in 2021 and collapsed a few weeks later. Perhaps the confusion was not an accident by the relevant "information governance" staff at NHS England who have been in place consistently through this period. For simplicity, we'll call the pandemic dataset the Pandemic GP dataset which continues to be collected until today.

[3] It was used for direct care provision of covid19 vaccines, but that's out of scope of this document. The admission by NHS England that they had done this, after strenuous denials they had done this, is

3. While one could assume that HDRUK, as the "National Institute for Health Data Science" always acted competently and honestly, that is the same assumption that DHSC's pandemic PPE procurement group made about the efforts of Baroness Mone – it may be that both took advantage of that assumption.

4. There is no clarity on whether the researcher partner universities were aware that the process being followed may or may not have satisfied their own legal or ethical obligations. In this HDR was the relevant data controller even if the "consortium" names another partner. The partner universities, whether or not involved in this project, had no real choice but to follow the statements of HDR and to assume NHS England was following their stated practices with other projects. This will matter in future because of the forthcoming £600m, "HDR Service" which we'll return to at the end.

5. There is an agreement between each GP and NHS England, and a Data Provision Notice, to that effect for which the details matter later. The key fact is that NHS England has been unable to provide information to show they satisfied their own decision making obligations regarding that processing of personal data. It is not unprecedented for NHS England's Data Protection team to make promises and then break them – a glance at the list of (un)published DPIAs[4] for the Palantir FDP project shows NHSE's promise of publishing DPIAs remains undelivered in practice. When NHS England says they publish something, is is necessary to check whether they actually have. One line appeared in the 2020 NHS Digital (as then) Data Uses Register, but it does not describe the details of the "foresight" project under discussion, which only commenced in 2023 and issued a press release in 2025. NHS England may claim this fully satisfies their data protection obligations, a test the ICO must examine and to which we include various pieces of useful context below.

An umbrella project in a pandemic, misused afterwards?

6. NHS data rules were reasonable in a national emergency: Projects doing covid work in 2020 got the benefit of the doubt about what they said they would do as long as they did what they said they would do.

7. One such project was a "cardiovascular disease and Covid19" project by HDRUK's confusingly named "British Heart Foundation Data Science Centre",[5] with a selection of universities involved. This is now claimed to be an umbrella project with other projects within, which is what raises specific concerns.

---

why we continue to have zero confidence in the assurances and decision making process of NHS England's "privacy, transparency, and trust" leadership who seem to deliver none of those three things

[4] Outside the scope of this note, but to the culture of NHS England's IG team, this FOI response confirms NHS England did not feel it necessary to do a DPIA for PETs service - literally the service handling NHS numbers: NHS England decided it didn't need to measure the impact of their privacy service. The culture of the department of health in England is that it doesn't want to know - a culture HDR copied and took advantage of. https://bsky.app/profile/marcusbaw.com/post/3lsgqfzl7cs2v

[5] Which despite the name is simply a brand within HDRUK's legal entity. The privacy policy makes this clear in the first line: https://bhfdatasciencecentre.org/privacy-policy/

8. In May 2025, HDRUK promoted a press release that this project had trained an AI “Foresight” “to predict potential health outcomes for patient groups across England. This could be events such as hospitalisation, heart attacks or a new diagnosis. Predicting these events early could enable targeted intervention, shifting healthcare much more towards prevention.” There is no evidence of an academic paper as yet, despite HDRUK saying they are “championing open science” because “We promote transparency and collaboration by making all research outputs publicly available”. Public outputs appear to be limited to a webpage from an academic involved, and a truncated HDRUK press release which included supportive quotes from the two Secretaries of State for DSIT and DHSC (but does not contain everything in the press release shared privately; you’ll have to ask them for a copy).

9. The NHS England data uses register for this project says the project received 49 different datasets all linked together, and states: “sub-licensing: no”.[6] The release register contains a long list of supposed data controllers, different project pages include different lists including organisations who are not named in the NHSE data uses register, and may or may not be involved at all. We expect the only data controllers whose decision making processes are under question is NHS England and HDRUK (and the precise breakdown of responsibilities between them is unclear, which suggests the HDR assumption of their inherent legality can not possibly be reliable).

Personal data is involved

10. According to HDRUK, the model was “trained on a set of NHS data for 57 million people in England, from which personal information has been stripped away”,[7] by which they mean pseudonymisation. The press release itself says the Pandemic GP dataset was used, and the release register says the umbrella project got that data.

11. The ICO will be aware of the recent ICO pseudonymisation guidance which states:[8]

> “Is pseudonymised data still personal data?
> Yes…”

12. NHS England will confirm that they protect (“stripped away”) the name, address, date of birth, and NHS number of individuals in the dataset, replacing them with a pseudonym, and will similarly confirm that every other field in the dataset remains untouched.

---

[6] Due to how NHS England discloses data projects in their environment, these data agreements only appear in the data uses register when there is a material contractual change. There is no public indication from NHS England for how long access was granted and “data uses” register term is being strained by this failure. We have pointed this out to NHSE previously, to no effect.

[7] https://bhfdatasciencecentre.org/news-and-events/groundbreaking-ai-trained-on-de-identified-patient-data-to-predict-healthcare-needs/

[8] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/pseudonymisation/#pseudonymiseddatastillpersonal

13. If a woman has given birth three or more times, while the DOB of the child is "stripped away", the dates of the maternity events of the mother remain entirely untouched and are almost certainly unique in the dataset, and can be linked to every other diagnosis in the dataset – all health events remain linked together to all other health events via the pseudonym, one identifying event unlocks them all.

14. We expect NHS England will argue that the personal data remains protected due to the other mitigations (safe environment, NHS England so far refusing any request to remove the multi-Gb model from the environment, etc), so it is clear that personal data was being processed. The NHS England data release register claims otherwise.

15. That processing includes processing by NHS England in providing the data to the umbrella project. The recent ICO guidance states:[9]

> "If we anonymise personal data, does this count as processing?
> Yes…"

16. (as an aside for completeness in case any part becomes relevant:
    a. as of July 2025, NHS England's IG/PTT continue to assert that pseudonymisation and anonymisation is not processing, and the NHSE data uses register contains many lines about data being "Anonymised - ICO Code Compliant".[10] The ICO Code being referred to is the 2012 Anonymisation code relating to the 1998 Act. At the ICO's webinar on the new guidance, the ICO said "It will be good to reassess" reliance upon the old code; NHS England have not.
    b. NHS England claims that the way a patient may express their right to object to unnecessary processing is via a National Data Opt Out (NDOO), an opt out that NHS England does not apply to unnecessary processing (and does not assess whether any processing is unnecessary or otherwise dissentable).
    c. NHS England does not ask researchers whether they wish NDOO to be applied – NHS England currently makes that decision solely and exclusively for itself. For pandemic purposes NDOO was not applied due to the pandemic exception (legitimately). But if this project turns out not to have been a pandemic project, any decision whether to apply or not would solely have been taken by NHSE not the researchers who would not know whether any patient had expressed a NDOO or not.[11]

17. It's notable that the press release for this project includes Peter Kyle, SofS DSIT worryingly and incorrectly claiming this is "anonymised data"

> "This ambitious research shows how AI, paired with the NHS's wealth of secure and anonymised data, is set to unlock a healthcare revolution."

---

[9] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/introduction-to-anonymisation/#doesthiscountasprocessing

[10] The "data export" excel file, "data releases" tab, "type of data" column. These entries should often read "personal data".

[11] Whether researchers should be given the choice of not using data of people who don't want their data used is a different matter the ICO may want to offer an informal opinion on.

18. The press release also has the Secretary of State for the Department of Health "harnessing trailblazing AI to radically transform our NHS – while also protecting patient data with strict security procedures" – it seems this project didn't follow those procedures, which is exactly the concern of independent parties about such projects. The trail AI blazed was what Elton John describes as "thievery".

The NHSE Information Processing Notice vs HDR practices

19. NHS England's description of how it makes decisions says:

> "If GP data is included in the application an additional review is also completed by the Professional Advisory Group (PAG). PAG consists of members from RCGP and BMA and our Caldicott Guardian"....

20. According to the gaps in the AGD minutes, this did not happen for the AI project.

21. That gap may be explained by HDR's "ways of working" process through which HDR made decisions for data access (page 8):

> "...new, standalone project proposals are submitted to the Approvals & Oversight Board for their consideration."...
>
> "The Board then decides whether each project is approved, approved with conditions, or not approved and notifies the project lead of the outcome.
>
> If approved (with or without conditions), the BHF DSC team lists the project reference and title on our webpage, together with the plain English summary (modified where necessary for clarity)."

22. The NHS England data uses register says sublicencing was not permitted, so there is no reason to expect that there would be projects that did not go through the normal NHSE processes or appear in the NHSE register. HDR admits theirs do not.

23. It appears HDR treated the data as if they could do with it as they liked. In effect, it appears HDR believed they were unrestricted data controllers.

Two decisions made by NHS England

24. In early June NHS England belatedly published their confirmation that this project was "novel" (item 10.1 from 15/5), and a week later published more minutes (item 10.1 from 5/6) confirming that *two* decisions were taken:
    a. one to suspend the foresight project pending an investigation (the primary topic of this complaint),
    b. and a second decision to terminate the approval process used for the project and send all the projects in that process through the full AGD process as normal (which is what E said they had always done) as they agreed they would do for GP data from the start.

25. We understand there is currently no investigation to the other 100+ projects that did not go through AGD about which HDR made their own decisions – some will be 2020 covid related projects that didn't involve GP data and will be absolutely fine; some may be 2025 decisions involving GP data that HDR and NHS England made using the same process now found untenable. There is no way for the public to tell the difference in breach of NHS England's assurances about data processing pursuant to the first principle of data protection.

AGD's first look at HDR's processes – majority of projects not supported

26. After questions were raised, NHS England required HDR to send all pending new projects through the AGD process which was described in item 5.1 in the 12th June 2025 minutes. AGD looked at all pending projects that HDR had already approved, and the majority were not supported, demonstrating that the HDR process is not equivalent to NHS England' standards.

27. Those minutes are clear – at some point the covid exception metastasized inside HDR into an everything exception, which it is not. The lack of data minimisation that was appropriate in March 2020 to figure out covid response should not be a permanent revocation of data protection principles for all of health research for that organisation.

28. Those minutes also indicate NHS England did not ask the GP Profession for their view on those projects when they went to AGD, even for the projects that were supported by AGD.

The Foresight project to build an AI

29. We understand from NHS England that the most official information about the project is available from HDRUK here:
    https://bhfdatasciencecentre.org/projects/ccu078/

30. It is notable that this information is not available from NHS England, and it is entirely unclearly how anyone would know that site existed if you didn't know it was a brand of HDR, legally part of HDR, but pretending to be something different.

31. It is unclear whether the project went through *any* NHS England review, although NHS England has suggested it did not.

32. **There is no mention of the foundation AI model project in the AGD minutes or the PAG minutes as attached to those AGD minutes. It is this in particular that suggests data agreements have been broken by NHS England.** We've asked for the references, we've been told all that exists is the insufficient link above.

33. The project has reference code DARS-NIC-381078-Y9C5K and while you can search all the minutes for references, NHS England will have to give a formal answer of who knew what when. There are regular updates to AGD by NHSE as

organisations join the consortium and regular renewals are made without details of this project being minuted.[12] This lack of minutes is what makes it unclear which data rules were in place at what time, and hence whether they were followed or broken at each point.

34. NHS England was still relying on 2020/2021 disclosures as late as the mention of the project in AGD minutes from February 2024 which state:

    > “The application and relevant supporting documents had previously been presented / discussed at the GPES Data for Pandemic Planning and Research – Profession Advisory Group (PAG) on the 28th July 2021 and the 24th June 2020.”

35. It is reasonable to believe that **the project did not get presented to PAG / AGD for the advice despite using GP Data which required the PAG process, because all PAG views appear in public minutes and there is no view given** (not even a “no comments” view).

36. There is no reliable known public[13] information on what, if any, process was used instead.

37. Two lines in the press release says “The Centre also involved members of the public, who continue to contribute to approving and shaping the research” and “A BHF Data Science Centre public contributor, involved in reviewing and approving this project” suggest it was HDR, and not NHS England, who made data controller decisions about access to data. According to their website’s privacy policy, the legal entity and data controller for the activities of the BHFDSC is “Health Data Research UK (“HDR UK” “we”, “us” or “our”)”.

38. It may be that it was HDRUK who was making decisions as a data controller outside of their lawful basis. HDRUK is not a public body, but a charitable company responsible to Trustees and utterly opaque about decision making to the point of propriety concerns[14] – our experience of HDR suggests they will make claims which appear to be made from independent organisations but turn out to be another sockpuppet brand run by HDR and reporting up to the HDRUK CEO and parroting the corporate line.

---

[12] Minuted mentions: igard-minutes---28th-may-2020-final.pdf, igard-minutes-11th-june-2020-final.pdf, igard-minutes---18th-june-2020-final.pdf, igard-minutes---25th-june-2020-final.pdf, igard-minutes---23-july-2020-final.pdf, igard-minutes---6-aug-2020-final.pdf, igard-minutes---20th-august-2020-final.pdf, igardminutes-15thoctober2020final.pdf, igardminutes-22ndoctober2020final.pdf, igardminutes-26thnovember2020final.pdf, igardminutes-3rddecember2020final.pdf, igardminutes-21stjanuary2021final.pdf, igardminutes-25thfebruary2021final.pdf, IGARD+Minutes+-+1+July+2021+Final.pdf, IGARD+Minutes+-+29+July+2021+-+FINAL.pdf,IGARD+Minutes+-+5+May+2022+final.pdf, IGARD+Minutes+-+24+November+2022+finalv1.pdf, AGD+minutes+-+22+February+2024+final.pdf. All the original links: AGD 2024 onwards and IGARD 2017-2023.

[13] We don’t know of any private information either.

[14] We have long held concerns about HDRUK’s activities.

39. Despite being the organisational applicant listed in the NHSE Data Uses Register, HDRUK itself is not listed as a data controller in the register entry, although this may be due to incompleteness by NHS England as their SDE project entries contain less information than their dissemination entries. It has never been assessed whether this oversight is fully compliant with NHSE's stated policies, but it is clearly not best practice that NHSE has never corrected until now.

Processes get gamed for gain

40. We do not know when those decisions were made, how they were made, what the decision makers were told, what advice was requested or given, nor what rules were in place at any relevant time or how they changed over time. However it does look like HDRUK decided to set and mark its own homework on decision making.

41. After questions were asked in recent weeks, this delegated authority arrangement was immediately terminated according to minutes of the NHS England Advisory Group on Data meeting on the 5th June 2025, although it is unclear whether HDRUK believe they have other channels to use data without it appearing in NHS England's data uses register. As the minutes make clear, the permanent termination of any ability for HDR to make their own (new) decisions (instead they come to AGD) is entirely separate to the initially-temporary suspension of the Foresight project for the duration of any assessment of what that project did. However, the permanent termination shows that the process agreed to by DH/E was not appropriate in 2025, even if it may have been appropriate in March 2020. COPI notice expiry in 2023 may be the boundary between the two but the foresight project (and others of concern) happened after COPI exceptions had wound down, and HDR continued using the process until NHS England suspended it as above.

42. HDR does not publish any reliable information about what projects were approved when. Their internal process was wide open to being gamed, which in 2020 was reasonable due to an emergency and an assumption everyone was working to mitigate the emergency. Yet the gaming of processes analogous to Baroness Mone in the Covid PPE VIP lane must be seen very differently when they continued to happen into 2025.

43. If HDR were not gaming the process, it will be up to NHS England to evidence and explain why the number of "covid only" projects that went through a process for OpenSAFELY to use GP data (and which had normal scrutiny) are so very different to the number that snuck under the HDR umbrella project (without normal scrutiny). (as an aside, this is why shadow/weaker processes for data access are of widespread concern because they always get gamed, as Biobank have found).

44. Indeed, the HDR/BHFDSC newspage (copyright HDRUK) about the announcement says:

> "Through rigorous approval processes, the British Heart Foundation Data Science Centre at Health Data Research UK made it possible for the researchers to access and work in the SDE."

45. That page does not refer to NHS England at all. It is as if HDR believes it could make decisions about who could access GP data; the very thing Government has announced it wants to do, but currently has not yet done.

46. The government announcement is for “consented cohorts” like the eugenicist-defending UK Biobank, but HDR are also arguing to expand to “unconsented cohorts” of which the prime example is pandemic GP dataset. A dataset they have already seemingly misused. At the launch of the HDR/Sudlow Review, the report author cited Biobank as having “one of the best” processes for getting data in “days” to eugenicists and other users. The “ways of working” for the HFDSC says on decision making speed:

    > “The current period from submission of a proposal to outcome notification is - on average - one month and the BHF DSC team is looking at ways to reduce this timeframe.”

47. These ways of working contradicts the big clarifier box on the NHS England website saying:

    > “**The applicant organisation of a DSA is the organisation that submits the application to NHS Digital and is not always the organisation receiving or determining the purposes and means by which the data are processed. The data controlling organisation determines the purposes and means by which the data are processed.”

48. We note HDRUK’s BHFDSC page about “whole population datasets” says HDRUK are already “Accelerating data access: We provide a single, streamlined project proposal and review process across the UK nations”.

49. We are aware of no national scale GP dataset which would allow them to do that; although perhaps HDR have done it anyway as BMA imply.

Questions for the ICO and NDG

50. The ICO is not the everything regulator. It is the NDG’s role to assess whether the NDG Caldicott principles were satisfied (but DPA requirements can not be satisfied if they were not) or the agreements between the Department of Health / NHS England and GPs.

51. The Caldicott principle of “no surprises” seems demonstrably to have been broken which suggests the “lawfulness, fairness and transparency” principle, as well as the “purpose limitation” and “accountability” principles can not have been satisfied. Arguing in 2025 that a covid-only project approved in 2020 allowed the training of a “foundation model” AI from 2023. The lack of an academic publication and the extensive effort that went into a press release suggest this was not an accident.

52. **Does this entire mess and lack of clarity satisfy the Caldicott Principles?** If it does not, it can not have been compliant with the Data Protection Act.

53. This project could have been done safely, it could have been done following the agreed processes – there is not even disagreement that it wasn't. Everyone agrees it didn't go through the agreed necessary process, the question for the ICO is whether that was a breach of the Data Protection Act, or whether the Act permits NHS England to offer free-for-all even when it says it will not.

54. Does HDRUK's use of their many sockpuppets (PEDRI, DARE, PRUK, BHFDSC, "the alliance", etc, who all report into HDR's leadership) to lobby in favour of HDR satisfy the Caldicott Principles?

55. HDRUK have made public claims (most recently at the Westminster Health Forum on the 2nd July for which a transcript is forthcoming) that they have responded to medConfidential's concerns. While they may have said things in response, we have not seen any response of HDR to AGD's inability to support the majority of HDR's new projects. HDR also claimed that they must be trustworthy because they have written a data use register standard; although it is unclear why HDR did not use it for the 100+ projects for which information is not in the public domain.

There is no urgency, but there is importance

56. medConfidential understands that the AI HDRUK's researchers built remains in a locked folder inside NHS England's Five Safes Trusted Research Environment. As such, while data may have been processed inappropriately, it remains under the active control of NHS England even if NHS England claims someone else is the data controller (somehow?).[15]

57. That the research project has made no assessment of how it would prove the contents of a multi-Gb model does not contain embedded personal data suggests it could never have been used in practice. Nor did they make any assessment of efficacy or safety. So was it just a research project at heart? If so, what detailed ethical scrutiny did the research project receive from their sponsor "for the duration of the research, taking account of developments while the research is ongoing"?

58. This debacle matters because the Secretary of State for Health has announced, to the glee of HDRUK and others, that he wishes to take decisions about data access for GP data away from the Profession and make the decisions himself (after the "merger" of NHS England into DHSC) and create a "Health Data Research" service to encourage projects like this. In practice, it looks like HDRUK assumed that some future state was already in place and did whatever they liked with the personal data in their possession – the GP records of 57 million people – because they believed they could; because they believed that they could use this data to train an AI. Analogies to other AI data debates abound…

---

[15] We note this argument is diametrically opposed to Biobank's argument that Biobank have no responsibilities after Biobank have given the eugenicists continuing access to the Biobank TRE running on servers donated to the Biobank charitable entity as a tax writeoff.

59. When the Secretary of State makes a promise about how data will be used, “covid only” or not used, can he simply change his mind and not tell anyone? Can NHS England simply ignore such promises when inconvenient? Those are general questions which regulators don’t answer, but this case is a specific example to examine what happened and whether it was lawful.

60. Several weeks ago when first raising concerns, we understood NHS England were going to do an internal audit of all pandemic GP dataset projects both under this project umbrella and under others, and whether they should have been shown to the Profession Advisory Group in advance as agreed. Despite any protestations of HDR or NHSE’s PTT team, that audit is still necessary until NHSE and HDR wish to give full undertakings that there aren’t other projects…[16]

medConfidential
coordinator@medConfidential.org

---

[16] In which case we’ll have to write a part 2.