**The Goldacre Review, Trusted Research Environments, and NHS data**

**Background**

Seventy-three years of detailed patient records means that NHS data represents an unparalleled resource for researchers and NHS service planners. Its analysis and role in the development of data-driven technologies can lead to invaluable benefits for patients, the NHS and society more broadly. But the Government also views personal health data as central to its ambition to grow the UK economy post-Brexit by becoming a science and technology superpower, while many organisations see our data as potential sources of profit.

One concern about this agenda is how the privacy of our personal data will be safeguarded, given that there is a wide range of health data sets, accessed by a variety of data users (including commercial organisations) who work on data that’s downloaded to a range of locations. There are also signs (such as the forthcoming [Data Reform Bill](https://keepournhspublic.com/is-your-health-data-at-risk/)) that existing data protections are to be weakend by government. Even existing regulations have been insufficient to allay public concern about plans, such as ‘GPdataGrab’,[[1]](#footnote-1) to extract NHS data, largely because of fears about privacy and that commercial companies will acquire access to personal information.

Enter Ben Goldacre, Professor of evidence-based medicine at Oxford University, who the Government commissioned to review how NHS data can be stored, managed and analysed most effectively while at the same time protecting patients’ privacy and avoiding unnecessary duplication of effort. One of the main proposals of Goldacre’s review, “[*Better, broader, safer: using health data for research and analysis*](https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis/better-broader-safer-using-health-data-for-research-and-analysis#trusted-research-environments)”, is the creation of Trusted Research Environments (see below), that will allow increased access to NHS data while also retaining control and ensuring privacy.

In our view, Trusted Research Environments (TREs) are presented as something of a panacea and need to be seen in the light of competing messages from NHS England (NHSE). Although it seems TREs could offer a significant range of benefits, we are unclear about which data sets will be stored by TREs and whether some types of data will remain vulnerable to privacy breaches and commercial mining.[[2]](#footnote-2)

For example, Goldacre’s recommendation is that “all analysis of NHS patient records should move to be done in a TRE” (p.121), yet this contrasts with the evidence that Tim Ferris (Director of NHSE’s Transformation Directorate) gave to the House of Commons Science and Technology Committee on [*The right to privacy: digital data*](https://committees.parliament.uk/oralevidence/10115/pdf/). He described healthcare data as having four main uses (to inform direct care; deliver population health; plan services; and for research) and, when asked if TREs are appropriate for all four uses, his response was no, only in the case of research.

To give a further example, Goldacre suggests that the Government is in favour of TREs as the way to reassure patients about the privacy of their NHS data, to simplify data access and reduce data dissemination, and yet this does not seem to square with NHSE’s[[3]](#footnote-3) recent announcement of plans to “unlock the value of data” by creating a ‘federated data platform’ (FDP). Described as an “ecosystem of technologies and services” to be implemented across the NHS in England, the proposed FDP will both store and support access to data for operational and internal analysis purposes. The aim is to connect NHS trusts, ICSs and national organisations such as NHSE so data, including sensitive information, can flow more easily across local, regional and national levels - and seems to cut through the rationale for setting up TREs.

**The Goldacre Review**

The review, “[*Better, broader, safer: using health data for research and analysis*](https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis/better-broader-safer-using-health-data-for-research-and-analysis#trusted-research-environments)”, is extensive. We focus on just three of the issues it raises:

1. the shortcomings of ‘pseudonymisation’ as a technique to secure patient privacy;
2. the argument for a limited number of shared platforms - ‘Trusted Research Environments’ (TREs)- to allow secure access to NHS patient records; and
3. TREs and the commercial use of data.

1. Pseudonymisation

Pseudonymisation is commonly used for processing NHS data. It refers to the process of removing markers that directly identify an individual (such as their name, date of birth and address) before their data is shared with different locations or users. Although it can be augmented by other types of control, there is widespread unease that pseudonymisation can be easily reversed by analysts working on NHS records, especially when combining multiple datasets. Significantly, when providing evidence to the House of Commons [Science and Technology Committee](https://committees.parliament.uk/oralevidence/10197/pdf/) inquiry on *The right to privacy: digital data*, Goldacre acknowledged that the privacy benefits of pseudonymisation were over overstated and his own concerns about it meant that he was one of the millions of people who ‘opted out’ of ‘GPdataGrab’.

Datasets derived from patient records are growing in size, becoming more detailed, and are being accessed by a growing number of analysts, meaning that pseudonymisation is becoming increasingly unreliable as a way of safeguarding privacy. In response, there is growing reliance on a “spaghetti junction of regulations” determining who can access the data. Researchers and NHS analysts may need to obtain multiple permissions from multiple organisations to access the data they need. To make things more difficult, rules governing access are often interpreted over-cautiously because of their lack of clarity. This may lead to projects being delayed or abandoned: according to Goldacre, “important data analyses that could substantially improve the quality, safety and cost effectiveness of care are not being done” (p.143). Goldacre’s main solution to this is the development of Trusted Research Environments.

2. Trusted Research Environments (TREs)

Besides the growing complexity of datasets and regulations for access, there is currently no coherent or integrated system for the collection, storage and processing of patient data. Instead there are currently hundreds of isolated analytic siloes and data centres, to say nothing of diverse working practices amongst analysts.

Goldacre’s response is that patient data should be held by just a small number of secure analytic platforms - Trusted Research Environments (TREs) - to be used by academics, NHS analysts and ‘innovators’ (generally meaning private companies) for all analysis of NHS patient records wherever there is a risk to patient privacy. Patient data should only be transferred to other non-NHS data centres when that patient has given explicit consent (e.g. in consented clinical trials). TREs also present the opportunity to impose set standards on how commonly used datasets are managed.

According to the Goldacre review, TREs represent a more dynamic and open approach to data use while retaining control of the data. Each TRE is

“a secure environment that researchers enter in order to work on the data remotely, rather than downloading it onto their own local machine. Users can extract and download the answers from their analyses – such as results tables, or graphs – but individual patients’ data always stays within the secure environment.”

A TRE for NHS patient data[[4]](#footnote-4) would have a number of objectives, including:

* the preservation of patients’ privacy (e.g. by obstructing attempts to re-identify patients in data);
* high quality and reproducible data quality (through, for instance, supporting the standardisation of re-usable code for common data management)
* a secure computing environment and improved governance (e.g. ensuring only users with appropriate permissions can execute code on the platform)
* public logs of all actions on patients’ records to help prevent the misuse of data;
* improved patient trust by, for example, publishing governance arrangements and how decisions about data access are made;
* more efficient and collaborative approaches to data science where methods and code become open, available for discussion, review and re-use; and
* better risk management, meaning gatekeepers can be more confident about granting permissions than if data is allowed to be removed and disseminated.

Each TRE would have three components:

* a service wrapper – a common framework used by all TREs to deal with permissions, monitor access, publish activity logs and so on;
* underlying computational aspects - largely generic tasks that can be delivered by staff with generalist software and data science skills; and
* the bespoke software necessary for work with NHS data – requiring an open, sharing approach to methods, code, tools and working practices.

Goldacre seems sure that, in future, a programme like ‘GPdataGrab’ will only be launched when all data access is through a TRE, understanding that the use of TREs is now policy. But in addition to the mixed messages mentioned earlier, it is unclear how many TREs are planned. Because each new TRE brings the danger of duplicated effort and risk, the recommendation is for as few as possible, with no more than three national TREs. However, the review also refers to an unspecified number of smaller, satellite TREs. For example, local TREs could exist for the recently legitimised ICSs that are expected to use population data to improve the quality and efficiency of care services. It’s unclear how these local TREs will be built - for example whether data from several ICSs will be combined to form a local TRE – although it appears that ICS TREs should be able to request their patients’ NHS data, cut from national datasets, to be transferred to them as needed.

3. The commercial use of data

Goldacre argues that the confidence of those governing access will be increased by the use of TREs, while patients will also be more willing to share their data. As he acknowledges, much of patients’ concern about the sharing of personal data has been about access being granted to commercial organisations. TREs will address this by “providing assurance and transparency around the quality and reproducibility of commercial analyses”.

Goldacre gave his personal view on commercial access to the House of Commons Science and Technology Committee (mentioned earlier), saying that:

“I do think that if we use trusted research environments, we can do something really important, which is to separate privacy concerns from the political or moral judgment about whether private companies should be able to profit from NHS data. We can say it is possible for a large organisation to work on your data, to make money from it, and to develop new techniques, tools or devices by using that data. We can allow them to do that without at the same time worrying that they will be able to see our medical records. After we have adopted trusted research environments and competently implemented them across the community, we can grant access to many more people. Then it becomes a pure moral or political decision about who we grant access to.“

He believes that the NHS should grant access to commercial organisations but that there should be a strong strategic approach to ensuring equity in any innovations that are derived from NHS patient data – in other words, a sharing of profits. Enlarging on this point in the review, Goldacre gives the example of when a company’s development of an algorithm is dependent on richly detailed NHS datasets. These have been collected and curated at great cost over many decades by thousands of NHS technical staff, and so he believes that benefits should accrue to the NHS, even if the project does not deliver a commercial output or revenue.

The National Data Guardian, Nicola Byrne, finds this worrying. In her evidence to the [Science and Technology Committee](https://committees.parliament.uk/oralevidence/10115/pdf/), she said she would not want to see any financial contribution made by a company to the NHS to mean that their ethical obligations were “paid off”. She is also concerned that there is a risk of underestimating the value of NHS data, which cannot be understood just in fiscal terms in the short or medium term – potentially, the value lies in the knowledge that will be generated many years from now: the NHS should not get tied into commercial contracts that give exclusive rights to knowledge that is then locked away from public use.

The Goldacre review seems to assume that the public will accept commercial access to their personal data if there is a frank and open discussion with them about the ultimate purpose, along with an explanation about the important role that the private sector plays in innovation. There is little sense that patients may distrust data use driven by commercial gain and be unhappy about private companies profiting from access to their personal health data, whether or not some of that profit was shared with the NHS. There is also little sign of major engagement with the public about any of this.

Referring to the common practice of NHS organisations to outsource analytic work to commercial providers, Goldacre notes the advantages, such as economies of scale, but also the risks. For example, NHS data analysis is detailed, technical work requiring not just generalist data science skills but also deep knowledge of the clinical world and the administrative systems of the NHS. External contractors may not have this knowledge. Goldacre suggests that the NHS turns to external commercial businesses for analytic work partly because of poor understanding by NHS managers of what’s involved and an undervaluing of the NHS analytic workforce.[[5]](#footnote-5) One of the recommendations of the review is the creation of a supportive, professionalised structure for NHS analysts, akin to that of other government analyst professions with, for example, clear career paths, qualifications, technical job descriptions, and continuing professional development.

**Finally**

The government’s response to the Goldacre review is to be included in the forthcoming Health and Social Care Data Strategy, due to set the direction for the use of data in a post-pandemic healthcare system. A [draft version](https://www.digitalhealth.net/2021/06/data-strategy-put-nhs-patients-ahead-of-demands-for-profit/) of the strategy was published in June 2021, with the final version expected in Spring 2022.

June 2022

**APPENDIX 1**

**Data categories, data users and data sets**

In addition to outlining what he sees as the four main uses of data (see Background section), [Tim Ferris](https://committees.parliament.uk/oralevidence/10115/pdf/) describes healthcare data as falling into four main categories:

* electronic patient records, managed by many different organisations using different systems;
* national data registries, containing identifiable data stored at national level;
* administrative data;
* omics data concerned, for example, with genomics and proteomics.

According to Goldacre, data users and data sets include:

* National NHS analysts, typically employed within large organisations such as NHS England, each working with their own copies of (typically) national datasets such as GP data, Hospital Episode Statistics (HES) and Secondary Uses Service (SUS), and prescribing data that’s been extracted from local and national organisations and was held by NHS Digital (before NHS Digital was absorbed into NHSE).
* Local NHS service analysts employed by organisations such as Clinical Support Units, Clinical Commissioning Groups (and now Integrated Care Systems), GP Federations, NHS Trusts and public health teams based in local authorities. The way that these data sets are formulated varies widely, and so does how they are managed: for example, at the moment there are160 trusts and 6500 GPs all acting as separate data controllers.
* Academic researchers, typically employed by universities, who currently work on NHS data downloaded onto local machines;
* Commercial providers of analytic services, usually commissioned by local and national NHS organisations, who maintain their own large databases of data extracted from the NHS;
* Companies (or ‘innovators’) accessing data for research concerned with healthcare developments.

1. GPdataGrab (or, officially, General Practice Data for Planning and Research) was the recent attempt to collect comprehensive medical GP records and store these in a central database available for NHS planning and research, but also to private companies on the payment of fees. The scheme is currently paused. [↑](#footnote-ref-1)
2. See Appendix 1 for more detail about data users and types of data sets [↑](#footnote-ref-2)
3. This shortly followed NHSE taking over control of NHS digital matters from NHSX and NHS Digital. [↑](#footnote-ref-3)
4. Other types include academic TREs. [↑](#footnote-ref-4)
5. NHS analysts reported that it is not uncommon for an external provider to turn to them to provide the very work that the outsourced provider had been paid to do. [↑](#footnote-ref-5)