Dr Ben Goldacre review into the use of health data for research and analysis

This document sets outs the BMA's position in relation to the use of health data for research with regard to the terms of reference of the review.

Q1. How do we facilitate access to NHS data by researchers, commissioners, and innovators, while preserving patient privacy?

A confidential health service

1. Confidentiality allows patients to be frank with their doctors, safe in the knowledge that the doctor is under a duty not to share the information for reasons other than their own care (save in exceptional circumstances). If patients felt that they were unable to be frank with their doctor it may make it much harder if not impossible for the doctor to provide effective treatment – or patients may simply not visit their doctor at all.

2. There is, therefore, a strong public interest in maintaining a confidential health service so that patients will be encouraged to seek appropriate treatment and share information which is relevant. The obligation to protect information shared in confidence between patient and doctor is recognised in the common law duty of confidentiality and professional and ethical standards.[1]

3. There is also a strong public interest in responsible access to data for secondary uses to enable advances in medicines, healthcare and innovation from which individual patients and the public will benefit. A delicate balance must, therefore, be struck between the commitment to confidentiality between patients and their doctors and the benefits of research and other secondary uses.

4. Patient data can be used for research and commissioning purposes in three forms: anonymised, pseudonymised and identifiable. Each form of data will have its own requirements for appropriate access, handling and use with the most rigorous standards applied to patient identifiable data.

5. Our assumption is that researchers, commissioners and others should use anonymised or pseudonymised data. Gaining explicit patient consent should always be the preferred route when access to identifiable health data is sought for secondary uses.

6. We recognise, however, that seeking patient consent is not always practicable. Where consent cannot be achieved, an ethical and trustworthy governance framework for data access is needed to ensure that a balanced judgment can be made between the competing priorities of individual and societal benefits of a confidential health service and the wider benefit to society of the research or other secondary use. This will help to promote public confidence in data access.

7. Data sharing policies, at either a national or local level, need to meet the highest possible ethical and professional standards to secure the trust relationship between doctors and patients. In the context of medical research, these standards are set by NHS research ethics

[1] Confidentiality - GMC (gmc-uk.org)

committees overseen by the Health Research Authority (HRA), equivalent processes in universities and the professional regulation by the GMC of those doctors undertaking research.

8. Data sharing policies which fall short of ethical, professional and established regulatory standards would undermine the trust relationship between doctors and patients. This creates a risk of backlash and damage to public trust which would negatively affect support for data access and the benefits which follow.

Section 251 of the NHS Act 2006

9. Where access to patient identifiable data is required but where consent cannot be obtained then, in England and Wales, the provisions of section 251 of the NHS Act 2006 apply (often referred to as 'section 251 support').[2]

10. We strongly support section 251 of the NHS Act 2006 and its accompanying regulations. This legislation, and the role it provides for HRA's independent Confidentiality Advisory Group (CAG), is an illustration of how the ambition in Q1 can be fulfilled. In our view, this independent, advisory mechanism offers a solution that can navigate the tension between facilitating important medical research and commissioning and protecting privacy and confidentiality.

11. Where access to confidential data is required, 'section 251 support' can set aside the duty of confidentiality. Those seeking support under the legislation must apply to the CAG; the CAG can recommend (or not) the use of confidential data where it is not possible to seek consent and where anonymised or pseudonymised data will not suffice. Section 251 support therefore allows medical research or service planning which is in the public interest to progress, where applicants for data access have satisfied certain conditions which minimise the intrusion to patients' privacy and confidentiality.

12. The independent scrutiny of applications by CAG is crucial to help ensure public confidence. Minutes of CAG meetings are published online in the interests of transparency.

13. Under the 'section 251' regulations the ultimate decision-makers are the Secretary of State for Health and Social Care (for non-research applications) and the HRA (for research applications). In our view, it would be preferable for the ultimate authority to be vested in the Chief Medical Officer (who is accountable to the GMC) and for the Secretary of State to have a veto blocking release and use of the data.

Data Governance

14. Proper governance and security of data helps retain and ensure public trust in the NHS - and medical research which is supported by the NHS. Research shows that this trust could be significantly eroded if patients feel data is not securely accessed.[3]

15. Strong data governance models should be consistent for all access to NHS data for secondary uses. Alongside legal compliance, good governance models will combine 'check and balance' technical and managerial safeguards – which are demonstrable to the public. The overall objective is that data collection systems, and the organisations which control them, are

[2] The Health Service (Control of Patient Information) Regulations 2002.

[3] Understanding Patient Data / Ada Lovelace Institute / Hopkins Van Mil (2020) *Foundations of fairness: views on uses of NHS patients' data and NHS operational data* p. 17

trustworthy. There is no established template in terms of the precise nature of data governance and safeguards which should apply to each time data are accessed or used but the following features must be included: use of technology to reduce data access; independent oversight of data access; and limits on data use. These are explored in more detail below.

16. Part of good data governance is ensuring the quality and accuracy of data. Although we recognise the issue of data accuracy is not specifically within the terms of reference for this review, in our view it is necessary to consider measures which will improve clinical coding and data entry at primary data collection sources.

Use of technology to reduce unnecessary data access

17. Researchers and commissioners usually want to link data from different datasets so that successive episodes of care and patient journeys can be analysed. Significant benefits to health can be derived from linking large scale medical datasets. Use of third-party data linkage services, where the organisation performing the linkage is independent of the organisations which hold the data, is becoming common practice.

18. Full use must be made of technical measures which reduce access to identifying information and allow large scale data linkage and anonymisation or pseudonymisation to be performed in a secure way, either:
    - within a regulated and controlled technical environment where identifiers are replaced with pseudonyms 'on landing'; or
    - by pseudonymising data at source prior to extraction.

19. Governance requirements of data linkage environments should include that the 'key' or algorithm to reverse pseudonymisation remains with the third-party linkage service or identifying data are deleted after linkage and validation. The CAG already has a role in considering the design of data linkage systems to enable large scale linkage and subsequent pseudonymisation or anonymisation. This oversight should continue.

20. To further minimise the possibility of re-identification, large data sets can and should be held in separate regional centres which are not linked and which cannot reconnect to the source of the data.

21. We are aware of the development of platforms which allow authorised persons to perform analysis without the need for data to leave the host system. Providing remote access to data in this way would provide a further level of security. While we support this principle, some of our medical researcher members have voiced concerned that these systems can lead to data corruption by error. This problem needs to be addressed if remote access systems are to be deployed.

Independent oversight of data access

22. It is essential that committees which recommend, or authorise, data access are independent - and are seen to be independent and not open to the criticism that oversight is not robust. It is sometimes the case that oversight committees for data access constitute employees or affiliates of the institution holding the data. The inclusion of externally appointed and independent lay members, patient representatives and healthcare professionals can help to avoid such criticism. The important role of the HRA's CAG is outlined above. We also support the role of the Independent Group Advising on the Release of Data, which has independent

membership, in providing oversight of data releases from NHS Digital with the minutes of meetings publicly available online.[4]

Limits on data use

23. Formal agreements, such as data sharing or data use agreements, must be used to set out the terms on which data access or disclosure is to take place. These agreements must, for example, restrict uses of the data to certain purposes and prohibit re-identification of individuals and further linkage of the data. If researchers or commissioners decide that there would be a value in interrogating the data beyond the use for which it was originally obtained, then additional approvals would be required.

24. Doctors undertaking research are subject to the professional regulation of the GMC and expected to adhere to its standards as laid down in *Good practice in research* and *Confidentiality: good practice in handling patient information.* Dentists, nurses and other health professionals are also professionally regulated, as are teachers and social workers. Other researchers are not subject to such professional regulation and the sanctions that can arise from it. This means that for many non-professionals the sanctions currently in place for deliberate breaches of data sharing agreements or conditions of use or access may be insufficient for reward to be outweighed by sanctions as a deterrent.

25. We expect researchers (and those who sponsor research) who are not subject to professional regulation to adhere to rigorous standards which are similar to those imposed on regulated professionals. One way of achieving this is to include the requirement to follow research integrity and ethical or governance policies as part of contracts of employment, for example, most UK universities already have ethical codes or policies.

26. Additional measures, applicable to all those involved in accessing or using data, should include unannounced audits or review by the approving body (with resources provided to facilitate this). For deliberate and wilful abuse of confidential data, criminal sanctions, including custodial penalties, would offer a more effective deterrent than fiscal sanctions.

Risks of re-identification

27. Risks of re-identification must not be under-estimated. A feature of large scale linked medical databases is the richness of the resultant datasets – even after anonymisation or pseudonymisation processes have occurred. The risk of data 'triangulation', or the 'jigsaw effect' where data held in different places is pieced together, is increased with the availability of data on social media sites. Thorough risk assessments to mitigate against accidental or deliberate re-identification are necessary prior to any releases of data from these datasets. In addition, software can tell when de-identification appears to be being attempted and stop it. The area of re-identification risk requires ongoing monitoring so that additional safeguards can be considered in order to keep pace with technological advances. These safeguards should match those for accessing confidential data by employees not involved in that care pathway as an equivalent data access breach with the associated sanctions.

Q2. What types of technical platforms, trusted research environments, and data flows are the most efficient, and safe, for which common analytic tasks?

[4] Note that, unlike CAG, IGARD does not have a statutory function.

28. The SAIL (Secure Anonymised Information Linkage) system in Wales has a long-established record of providing anonymised linked data to support health research. Use of a 'twin track' approach means there is separation between identifiable demographic data and clinical data, a design which poses minimal risk to confidentiality.

Q3. How do we overcome the technical and cultural barriers to achieving this goal, and how can they be rapidly overcome?

29. To tackle the cultural barriers that exist to the use of patient data in research, patients and the public must not just be participants in medical research but co-producers of it, participating in the identification of research priorities, the questions that should be asked, the process for gathering that data and in promulgating the results and learning from them. The role of doctors should be to facilitate and aid this level of participation and, where necessary, explain what is proposed at each stage of the process. Thus co-production (academics, professionals, patient and public involvement etc) in governance bodies and research projects (throughout the process from identifying the question, solution, data collection, analysis, interpretation and dissemination) will mitigate data misuse or spurious use.

30. We have carried out research with doctors about what they see as practical barriers to providing data for research. From the GPs perspective the concerns below were highlighted:

- numerous requests from a variety of sources;
- reviewing whether the researchers have the necessary approvals takes up a lot of time which can be difficult to manage alongside heavy workloads;
- GPs very often did not get the information they need – details of the project, legal basis for disclosure, whether the project has REC approval, what information is needed;
- GPs were frequently asked to provide researchers with the contact details of patients with a particular condition so that the researcher can contact them (which of course is a breach of confidentiality);
- handling requests takes a lot of time for which GPs are very rarely reimbursed;
- GPs were nervous about the confidentiality aspects and not knowing what the information governance arrangements were for holding and using the data once they released it;
- when GPs did provide data, they very rarely got any feedback.

Q4. Where (with appropriate sensitivity) have current approaches been successful, and where have they struggled?

31. There has already been well-publicised debate about the failure of the *care.data* programme in 2014. In short, the failure centred on the fact that the government's proposal to collect GP data and link it to hospital datasets did not align with public expectations about privacy and confidentiality. There was concern about the governance of the data held by NHS Digital (known then as the Health and Social Care Information Centre, HSCIC), specifically a lack of clarity about who would have access to the data and how these decisions were managed by HSCIC. Although the proposed collection had a lawful justification under the Health and Social Care Act 2012, it was clear that in this scenario the law conflicted with what the public deemed to be acceptable, resulting in a damaging loss of trust.

32. It was therefore disappointing when, in 2017, NHS Digital published a memorandum of understanding (MoU) with the Home Office and Department of Health and Social Care which permitted data sharing on those suspected of immigration offences. In its heavily critical

report into the MoU, the Health Select Committee said that NHS Digital's actions were inconsistent with the principle of confidentiality and risked undermining public confidence.[5] The MoU was suspended in 2018.

33. The theme of trust is critical if the benefits of data access are to be achieved. If patient and public trust is to be built and maintained in NHS Digital, as the national 'safe haven', it is important that the government and policy makers do not attempt to change established presumptions about confidentiality which are understood by both patients and healthcare professionals.

34. Using anonymised and pseudonymised data in the first instance and only using identifiable data when it is clearly necessary and within the appropriate legal and governance framework would also help build trust. Government and statutory organisations should be subject to the same restrictions as other potential data users.

35. The government does not build trust by appearing to by-pass public scrutiny and debate. A recent example of this is the Cabinet Office consultation on the expansion of the National Fraud Initiative data matching powers and new Code of Data Matching Practice. This consultation has potentially significant consequences for use of NHS data and was given an extremely short deadline with limited circulation.

Q9. If data is made available for secondary research, for example to a company developing new treatments, then how can we prove to patients that privacy is preserved, beyond simple reassurance?

36. We are in an era where the public has greater awareness about data use (and misuse) with recent prominence on, for example, the introduction of GDPR and the Cambridge/Analytica/facebook scandal. This means that more engagement with the public about data access is expected and required in order to build and maintain trust.[6]

Transparency and clarity of information

37. In order to build trust, the provision of clear, and easily accessible, information which tells people what is happening to their data is essential, including who has access and why.[7] For complex data flows, inclusion of a simple data flow diagram is recommended.[8] The information must include details of the safeguards to which the data will be subject.

38. The importance of transparency cannot be over-emphasised when considering commercial companies' access to health data. Research with the public consistently reinforces this theme.[9] The public generally view commercial access to health data with suspicion – even

[5] https://publications.parliament.uk/pa/cm201719/cmselect/cmhealth/677/67708.htm#_idTextAnchor033

[6] One of the key findings of recent public engagement work was that the public should have a say in how NHS data is used: https://understandingpatientdata.org.uk/what-do-people-think-about-third-parties-using-nhs-data

[7] NHS Digital's data release register is an example of this: Register of approved data releases - NHS Digital

[8] Transparency must also extend to being clear with the public about the meaning of terms such as 'confidential data', 'pseudonymised data' or 'anonymised data'. Without understanding what these terms mean it is likely to be difficult for members of the public to reach a decision on where there stand on use of data.

[9] Understanding Patient Data / Ada Lovelace Institute / Hopkins Van Mil (2020) *Foundations of fairness: views on uses of NHS patients' data and NHS operational data* pp. 24, 29, 61

when data are anonymised or pseudonymised.[10] Data should not be shared with third party companies unless it has been approved as part of the initial data request. Public discomfort about commercial access, or involvement, can be lessened where it is made explicit that the purpose is for health research or the organisation is working closely with the NHS.[11] Commercial organisations must be expected to adhere to the same standards of transparency and confidentiality as academic and NHS organisations and staff.

Involvement of patients and the public

39. With further reference to the CAG, meaningful patient and public involvement (or PPI) is an important factor considered by the CAG when reviewing whether a proposed unconsented secondary use is in the public interest. Applicants must demonstrate that the use of confidential information without consent had been discussed with, and supported by, relevant patient groups. These discussions should occur at the start of the project, not as an 'add on' at a later stage. Good PPI allows those wishing to use the data to take account of the attitudes of those (or their representatives) whose data is being used.

40. Involvement and engagement with relevant patient groups provides the 'social licence' to support important research, or other secondary uses, from which patients and public will benefit. In view of the persistent tension between access to data and preserving confidentiality and privacy, the support from relevant patient groups can significantly strengthen the public interest argument in permitting unconsented data access.

Mechanism for opt out

41. An important feature of fostering public trust is to provide patients with an easy way to opt out of the use of their confidential data (in all but exceptional circumstances). The level of public awareness and understanding about either the National Data Opt-out or the 'Type 1' opt-out is unclear.[12] Any future discussions about the use of health data must recognise the importance of public understanding about rights of opt-out and, equally, the enormous benefits which come from the use of data in health research. It is usually the case that small numbers will opt out and, for the majority who would be happy to participate a failure to offer the option is likely to build mistrust.

For further information please contact:
Sophie Brannan
Senior Policy Advisor (Medical Ethics)
sbrannan@bma.org.uk
Tel: 020 7383 6816

[10] Wellcome Trust / Ipsos MORI (2016) *Public attitudes to commercial access to health data*, p.13 A significant minority of people (17%) objected to private companies having access to health data under any circumstances.
[11] Wellcome Trust / Ipsos MORI (2016) *Public attitudes to commercial access to health data* pp. 10, 84
[12] A 'type 1' opt-out prevents confidential information being shared outside a patient's GP practice for purposes other than direct care.