NEW INITIATIVES IN THE ENGLISH NHS TO INVEST MORE IN HEALTH DATA
We welcome the publication of the Goldacre Report, ‘Better, broader, safer: using health data for research and analysis’, and the Secretary of State for Health and Social Care’s ‘Data saves lives’ initiative.1,2 The UK Government has also announced a £200 million investment to support these new strategies and their associated health data repositories.2 However, these developments are presented as technical transactions. Greater health community-wide ownership and bottom-up endorsement of data sharing by known and trusted healthcare professionals would increase their chance of success.

BENEFITS AND POTENTIAL OF ROUTINE DATA SHOWN ACROSS THE PANDEMIC
Routine health data accelerated our understanding of COVID-19 therapies and the impact of vaccination. One important enabler was the temporary relaxation of data sharing restrictions under emergency regulations that occurred during the pandemic; and subsequently withdrawn in July 2022.3 It is of note that there was no widespread campaigning or objections to this extended sharing and use of health data among practitioners or the public. Another pandemic success was the acceleration of digital maturity in UK health care; there was much more reliable cross-health service sharing of test results and vaccination data than we see with seasonal influenza, for example. Without effort and sustained investment, as proposed in these initiatives, we risk losing much of this progress.

STRENGTHS AND LIMITATIONS IN UK HEALTH DATA ECOSYSTEM
A data ecosystem is the complex series of processes that draws together data from a range of disparate but functioning medical record (CMR) systems. It is likely this complexity will grow and our future plans need to take account of this. Many of these limitations are described in the Goldacre Report; however, public mistrust in technology (as opposed to in people) and inconsistencies in health data quality could have been elaborated on further and may undermine many of the recommendations made if left unaddressed.

TRUSTED RESEARCH ENVIROMENTS
The role of so-called Trusted Research Environments (TREs) is emphasised throughout the report. TREs are highly secure data environments that provide remote access to health data to approved parties for approved purposes. A small number of TREs are planned to provide rich repositories of both biological and medical record data. Though their use within health care is under development, the aspiration is that they promote transparency, drive innovation within health, and facilitate open working. Specific TRE owners will prevent overcrowding and any duplication of efforts or confusion over who owns what among researchers. Their transparency should ensure they are not ‘black boxes’, as has been a major criticism of past technological reforms in the NHS.

CENTRALISM OR A FEDERATED MODEL FOR TREs
Centralism does not always achieve more — especially in the context of innovation. The risks are that streamlining decision-making in this way may meet the needs of some groups but not others, and that the expertise of specialist teams may be lost as a result of this standardisation. It

Table 1. UK data sources, from oldest to most recently established, that have contributed to COVID–19 research and/or surveillance and could contribute to a federated trusted research environment

<table>
<thead>
<tr>
<th>Database acronym</th>
<th>Full name or database</th>
<th>Coverage</th>
<th>Size (patients)</th>
<th>Established</th>
<th>Role in COVID-19</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPRD</td>
<td>Clinical Practice</td>
<td>UK</td>
<td>&gt;60 m, including 16 m currently registered</td>
<td>1967: sentinel surveillance</td>
<td>Sentinel surveillance and research</td>
</tr>
<tr>
<td></td>
<td>Research Datalink</td>
<td></td>
<td></td>
<td>1987: known as General Practice Research Database</td>
<td>Risk groups and research</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2012: changed to CPRD</td>
<td></td>
</tr>
<tr>
<td>QRsearch</td>
<td>N/A</td>
<td>UK</td>
<td>&gt;5 m</td>
<td>2003</td>
<td>Risk groups and research</td>
</tr>
<tr>
<td>EAVE II</td>
<td>Early Assessment of Vaccine and anti-viral Effectiveness</td>
<td>Scotland</td>
<td>&gt;5.4 m</td>
<td>2009</td>
<td>Sentinel surveillance and research</td>
</tr>
<tr>
<td>SAIL Databank</td>
<td>Secure Anonymised Information Linkage Databank</td>
<td>Wales</td>
<td>&gt;5 m records</td>
<td>2008</td>
<td>Sentinel surveillance and research</td>
</tr>
<tr>
<td>OpenSAFELY</td>
<td>N/A</td>
<td>UK</td>
<td>&gt;58 m</td>
<td>2020</td>
<td>Risk groups and research</td>
</tr>
</tbody>
</table>

m = million. N/A = not applicable.
is all too easy to create bottlenecks that block progress. A federated TRE model underpinned by converging standards but mindful of the value of expert groups could be the way forward; examples are shown in Table 1. Such a federated approach could facilitate the development of the ‘talent pipeline’ of in-house analysts that Goldacre describes while ensuring deep and domain-specific expertise remains.

**IMPROVING DATA QUALITY AND UNDERSTANDING THE CONTEXT OF DATA RECORDING**

TREs may not be a panacea for managing and analysing health data, as great, if not greater, priority should be improving data quality and understanding the context of data recording to ensure its correct interpretation. CMRs are a complex mix of registration data, coded clinical data, and free-text. Realising the full benefits of routine data will require improvements to clinical coding. The value of clinician coding, particularly of presenting diagnosis, is all too easy to create bottlenecks that block progress. A federated TRE model underpinned by converging standards but mindful of the value of expert groups could be the way forward; examples are shown in Table 1. Such a federated approach could facilitate the development of the ‘talent pipeline’ of in-house analysts that Goldacre describes while ensuring deep and domain-specific expertise remains.

**LIMITED BENEFITS FROM EXISTING DATA RECORDING**

All too easy it is to assume that greater data quality can be used to reduce bureaucracy in primary care. Finally, by ensuring that the benefit–risks of data sharing are supported by communications from trusted clinicians. In doing so, we can be sure to strengthen the successful but sensitive health data ecosystem of the UK and minimise the risk of disruption.

Simon de Lusignan, (ORCID: 0000-0002-8553-2641) GP and Professor of Primary Care and Clinical Informatics, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford.

Meredith Leston, (ORCID: 0000-0003-0891-714X) Oxford-MRC Enterprise CASE Scholar and Alan Turing Institute Enrichment Student, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford.

Margaret Ikpo, (ORCID: 0000-0002-3442-881X) GP and Vice Chair of Professional Development and Standards, Royal College of General Practitioners, London.

Gary Howsam, (ORCID: 0000-0001-6699-5504) GP and Vice Chair of External Affairs, Royal College of General Practitioners, London.

**UNINTENDED CONSEQUENCES**

Top-down changes can cause unintentional disruptions and render a complex ecosystem dysfunctional. Too often such change fails to consider the need for local engagement. Trust is hard won, easily lost, and must be cultivated between people. The difficulties over health data being used to support wider use-cases, such as asylum seeking, reflects this.8

**IN SUMMARY**

The Goldacre Report and the ‘Data saves lives’ initiative are both welcome and timely. As Goldacre puts it best, it is our collective duty to make this work.7 We recommend strengthening this report in three important areas. First, by federating existing data expertise into these national TREs. Second, by finding ways that higher data quality can be used to reduce bureaucracy in primary care. Finally, by ensuring that the benefit–risks of data sharing are supported by communications from trusted clinicians. In doing so, we can be sure to strengthen the successful but sensitive health data ecosystem of the UK and minimise the risk of disruption.

**TRUST**

UK primary care professionals are trusted by their patients9 and show professionalism in facilitating data sharing.7 The Goldacre Report could have included more about engaging practices and providing tangible benefits to clinicians and their patients. There is scope to engage, even co-design, with practitioners, patients, and service users, for example.8

Simon de Lusignan is Director of the Royal College of General Practitioners Research and Surveillance Centre. Through his university he has funding for vaccine-related research from AstraZeneca, BSK, Sanofi, Seqirus, and Takeda; and, been members of advisory boards for AstraZeneca, Sanofi, and Seqirus. All other authors have declared no competing interests.

**REFERENCES**


**ADDRESS FOR CORRESPONDENCE**

Simon de Lusignan
Nuffield Department of Primary Care Health Sciences,
University of Oxford, Eagle House, Walton Well Road,
Oxford OX2 6ED, UK.

Email: simon.delusignan@phc.ox.ac.uk

Open access
This article is Open Access: CC BY 4.0 licence
[http://creativecommons.org/licenses/by/4.0/].

DOI: https://doi.org/10.3399/bjgp22X720965