Pharmaceutical industry priorities for the Health Data Research Service

August 2026

Contents

Introduction

Context

The pharmaceutical industry contributes £1 in every £6 of commercial R&D spend in the UK.[1] Much of that investment depends on access to health data across the medicine development lifecycle: identifying unmet clinical need, testing new medicines, and monitoring safety and effectiveness in routine clinical use.

The unique advantage of NHS data globally is its longitudinal depth and population representativeness. Unlike insurance claims data – the primary source in the US and many other markets – NHS records can span decades of a patient's life, enabling stakeholders including industry to monitor exposures and health outcomes over time and generalise findings to populations in other countries. General Practice (GP) data is the foundation of this, as GP records capture around 90 per cent of a patient's NHS interactions.

Despite this potential, the UK data ecosystem has not yet consistently delivered. Fragmented datasets, complex access processes, and unreliable linkage are well documented. These problems are a particular deterrent for an industry that can place its data-enabled research anywhere in the world. They undermine what could and should be a core strength of the UK's life sciences research environment.

Previous attempts to build an integrated national health data infrastructure in England have not succeeded. The independent Health Data Research Service Digital Ecosystem Analysis[2] (Emrys Health/Nesta for Wellcome, 2026) attributes these failures to fragmented infrastructure, loss of public trust – most starkly demonstrated by the collapse of care.data and the General Practice Data for Planning and Research (GPDPR) programme – and failure to engage research users from the outset. The UK's health data ecosystem remains, to quote that analysis , a 'federation of fragments'.

Following the Sudlow Review, the ABPI and Health Data Research UK (HDRUK) jointly called for the establishment of the Health Data Research Service (HDRS) in January 2025. The Prime Minister announced the HDRS in May 2025 in partnership with Wellcome. The government's £500 million commitment was confirmed in the Life Sciences Sector Plan.

Opportunities, considerations and risks

The HDRS represents a genuine opportunity to turn the UK's existing health data assets into a coherent national service. The HDRS Digital Ecosystem Analysis is clear that the components required already exist across the Clinical Practice Research Datalink (CPRD), Genomics England, the Secure Anonymised Information Linkage Databank (SAIL), the Secure Data Environment network, UK Biobank, and the devolved nations' Safe Havens. The gaps lie in connective architecture, service infrastructure, and governance, not in the underlying data assets themselves. Success for the HDRS therefore means building on and connecting what exists and works well, not building new infrastructure that duplicates mature capability.

Many reports have now been published on what the HDRS should deliver. Collectively they contain a large and diverse set of asks that would be difficult to sequence and prioritise without clear rationale anchored in user needs. This report provides that rationale via new insights from the perspective of the pharmaceutical industry, obtained through detailed consultation with ABPI member companies.

The global pharmaceutical industry invests where it can access high-quality and connected data cost-effectively and predictably. Companies are likely to direct investment elsewhere when data access is slow, costly or unreliable. Pharmaceutical companies also face regulatory and health technology assessment (HTA) obligations to regulators including the Medicines and Healthcare products Regulatory Agency (MHRA), Food and Drug Administration (FDA), and European Medicines Agency (EMA). These are not shared by other research sectors and set clear guardrails and requirements for how industry research is conducted and evidenced. This context is fundamental if the HDRS's service design is to be regarded as internationally leading.

The pharmaceutical industry has decades of experience in handling patient data responsibly, operates through secure, privacy protecting environments, and ABPI member companies are committed to the principles the ABPI has published,[3] which in turn align with the Caldicott Principles. A health data research service based on privacy, security and ethics by design would provide a foundation trusted by both patients and the pharmaceutical industry.

The HDRS should be understood as one node in a global network of accredited secure environments: the global nature of pharmaceutical R&D means that carrying out all research within a single country is not always possible, particularly for rare disease research, safety signal validation, and multi-jurisdiction regulatory submissions. The HDRS has the opportunity to clearly articulate the unique offering of the UK health research ecosystem in a complex global landscape.

Purpose and added value of this report

This report builds on three previous ABPI publications, adding detail validated directly by member companies. It engages explicitly with two recently published independent analyses – the HDRS Digital Ecosystem Analysis and the Imperial Institute of Global Health Innovation blueprint[4] – and identifies where pharmaceutical industry recommendations go beyond, or add specificity to, what those reports describe. Where recommendations are new or particularly emphasised by pharmaceutical companies, this is flagged explicitly.

Service design

Context

The HDRS should be designed as a professional service layer that connects and enhances existing UK health data assets, not a new centralised repository. The HDRS Digital Ecosystem Analysis explicitly warns that centralisation has ‘repeatedly failed’ due to governance objections and public opposition and recommends instead a distributed integration layer keeping data with existing custodians. The pharmaceutical industry strongly supports this approach: what matters is reliable, predictable access to linked data – not where that data physically sits.

However, the HDRS should also have the necessary powers and functionality to process and link data, since existing custodians may not always have the capacity, capability or powers to deliver these functions.

The service should be configured around a transparent, published phased implementation plan that allows companies to plan their global studies. Any transition of existing industry-facing services should result in improvement, not disruption: the pharmaceutical industry has made long-term commitments based on the current services as a minimum, and continuity is a commercial necessity.

Recommendations

Independent data linkage capability

The two most critical barriers to pharmaceutical industry use of NHS data are timely access and reliable dataset linkage. These are the two factors that most influence where industry places its data-enabled research, so treating both as top priorities is one of the clearest ways for the HDRS to attract industry investment to the UK.

Linkage requires an organisation to bring together identifiable data from datasets belonging to different custodians. The HDRS should be able to perform data linkage independently, with sufficient capacity and capability to operate reliably without reliance on NHS England (NHSE) or Department of Health and Social Care resources. This includes linkage of NHS datasets to each other and to external sources such as research registries and proprietary datasets. Achieving this will require an appropriate legal basis to hold identifiable data for linkage purposes. This is not the same as centralising data: data should remain with existing custodians, but the HDRS should hold authority to perform the linkage function.

The consequences of relying on other parties with powers to undertake linkage have caused historic problems. For example, the CPRD has long provided GP data linked to secondary care data, but depends on NHSE to undertake the linkage required. In practice this dependency proved highly problematic: for more than three years NHS Digital/NHSE lacked the capacity to carry out essential linkages for the CPRD, despite appropriate permissions and a contract being in place. During this period, pharmaceutical member companies were unable to study patients across primary and secondary care.

By holding the linkage function itself, the HDRS can remove this dependency, giving companies the confidence to run research across primary and secondary care data in the UK rather than placing that work elsewhere.

The HDRS Digital Ecosystem Analysis identifies standardised linkage services with published quality metrics as one of the most critical gaps in the current UK ecosystem. Linkage quality is currently unmeasured across assets – researchers cannot quantify linkage error or compare results across studies. By establishing national linkage services with agreed methods, transparent quality metrics, and pre-agreed governance frameworks, the HDRS would let researchers quantify linkage error and compare results across studies for the first time, closing one of the most critical gaps identified by the HDRS Digital Ecosystem Analysis.

Focus on infrastructure, not demonstrators

The HDRS should prioritise structural fixes over individual exemplar projects. Previous government data initiatives have repeatedly created pockets of excellence while leaving fundamental fragmentation unresolved; a pattern the HDRS Digital Ecosystem Analysis confirms. Comprehensive UK-wide GP data access, reliable standardised linkage, and streamlined data access processes benefit all research sectors.

Exemplar projects should only be used to test options for the design and build of the fundamental requirements for national infrastructure, with the ultimate intent that options proven to be effective are scaled nationally. They should not be used to create solutions that rely on specific or unique features in one geography or context, as these risk further reinforcing fragmentation.

Priority datasets

The four core datasets

ABPI members are agreed on the four linked datasets that form the minimum viable product for the HDRS's first phase:

1. GP data
2. hospital episode statistics or equivalent secondary care data in devolved nations
3. death registration and mortality data
4. hospital prescribing data

GP data is foundational. Without it, the UK's longitudinal population health data – its primary competitive advantage over countries relying on fragmented insurance claims records – cannot be realised. GP records encompass around 90 per cent of a patient's NHS interactions across their lifetime. Every other NHS dataset provides only a partial view.

Achieving comprehensive UK-wide GP coverage requires nation-specific solutions. The HDRS Digital Ecosystem Analysis identifies specific legal barriers: Northern Ireland lacks a statutory framework permitting secondary use of identifiable data without consent, which prevents routine GP-to-secondary-care linkage for research; Scotland retains practice-level approval requirements across most regions. These are legal and governance problems, not technical ones. Resolving them through early engagement with devolved authorities is how the HDRS can secure the UK-wide GP coverage that gives the service its distinctive longitudinal advantage.

The next tier: additional datasets essential for pharmaceutical research

Beyond the core four datasets, ABPI member companies have identified the following as early-phase additions required to support the full range of pharmaceutical industry research:

- laboratory and pathology results from hospital source systems
- structured data extracted from unstructured clinical text through natural language processing
- cancer registry data linked to staging, biomarker, and treatment line information
- clinical genomic test data, along with national research genomics data, including Genomics England

Data-enabled clinical trials

Clinical trials account for more than 40 per cent of pharmaceutical R&D spend. The HDRS Digital Ecosystem Analysis documents the UK's deteriorating competitive position: despite some improvements in the ecosystem, industry-sponsored enrolment reached a seven-year low in 2024/25; the UK has fallen from fourth to eighth globally for phase III trials and is second slowest of 18 European countries for trial setup. The economic stakes are significant: clinical trials already contribute more than £7 billion in gross value added and more than £1 billion in NHS revenue annually.[5] A data-enabled clinical trials service within the HDRS should work through clinical teams, not just mass patient communications. The HDRS Digital Ecosystem Analysis confirms that clinician-mediated data-driven recruitment achieves significantly higher conversion rates.

In oncology and rare diseases, the service should also support biomarker-stratified and precision-medicine trials, where eligibility criteria require deep clinical data (staging, molecular markers, prior treatment lines) and genomic data that administrative datasets cannot provide. This gap is not addressed by current infrastructure.

Build on existing services

Existing services such as the CPRD, SAIL and UK Biobank have developed commercial access models that industry relies on. The HDRS should learn from and build on these services, improving security and standards where needed. Existing industry-facing services should be incorporated or expanded with a phased transition plan that preserves continuity. Any change should result in improvement. The pharmaceutical industry is ready to work in partnership with the HDRS and existing service providers to plan and manage this transition.

Ring-fenced budget and value return to data controllers

The HDRS should have a ring-fenced budget underpinning predictable service provision and the ability to retain surplus income for continuous improvement. However, financial sustainability requires more than charging commercial users.

The HDRS Digital Ecosystem Analysis identifies a structural problem: data controllers currently bear the costs of making data research-ready without fair value in return. This misalignment means controllers currently prioritise local operational needs over national research infrastructure – but it is correctable. By developing a method for flowing funding to the parties that provide data and make it research-ready, the HDRS can turn data controllers into willing partners and unlock hospital prescribing, laboratory and comprehensive GP data. This flow of funding should mirror the model for commercial reimbursement of clinical trials, in which internal reimbursement within NHS organisations ensures financial recognition.

Equitable access for industry and non-commercial researchers

Despite meeting regulatory standards not required of other researchers, some data custodians in the UK do not permit pharmaceutical industry access to anonymised data, or require an academic lead on industry-funded research as a way of achieving a public-interest justification. Putting non-commercial and pharmaceutical industry researchers on an equitable footing, with clear access policies making explicit that bona fide pharmaceutical industry researchers have the same rights as other approved users, would let the HDRS draw the full weight of industry research, and the investment it carries, through the service.

Data inputs and outputs

Context

Pharmaceutical companies submit regulatory evidence and conduct research across multiple countries simultaneously. While the HDRS would support in-situ analysis, there are circumstances where data export is essential – for regulatory submissions, multi-country safety monitoring and rare disease research requiring international patient populations. Supporting secure data export to accredited external trusted research environments (TREs) would enable the HDRS to serve regulatory submissions, multi-country safety monitoring, and rare disease research – the global use cases that bring high-value pharmaceutical work to the UK.

Recommendations

Accreditation of external TREs

Data exported from the HDRS should be received by external TREs meeting robust technical and governance standards. An accreditation process, developed in partnership with industry, is an urgent priority. Pharmaceutical companies operate internationally within their own accredited TRE infrastructure and can contribute to defining appropriate standards.

Regulatory and HTA compliance

Regulators including the MHRA, FDA, and EMA can request access to source data underlying evidence submissions. For the HDRS to support regulatory submissions, it should allow controlled data export for this purpose. Once a medicine is in clinical use, companies must also monitor its safety using real-world data, frequently pooling records from multiple countries to validate adverse event signals. This is a regulatory requirement, not a commercial preference, and should be accommodated by HDRS data export policies.

Cross-border dataset combination

For rare diseases and precision medicines, the UK may not have sufficient patient numbers for robust research without combining data from multiple countries. This need applies equally to non-commercial researchers in rare disease and life-course epidemiology. HDRS policies should support cross-border data use to ensure no UK patients are disadvantaged by belonging to a small domestic population.

Consented studies and clinical trials

Where UK patients have consented to their data being used in a research study or clinical trial, the HDRS should allow that data to be exported to the data custodian's accredited TRE, even if outside the UK. For international trials where companies combine data from multiple countries, this flexibility is essential.

Data access and data quality

Context

The HDRS should efficiently manage hundreds of data requests, providing access to curated, pseudonymised, individual patient-level linked datasets within commercially competitive timeframes. For pharmaceutical industry users specifically, it should also support needs unique to bringing medicines to market: auditable data provenance, point-in-time data versioning and rapid response to urgent regulatory requests.

Recommendations

Risk-proportionate data access

Pharmaceutical industry research ranges from simple feasibility estimates to complex hypothesis-driven studies. Low risk uses such as incidence and prevalence estimates, trial feasibility assessments, and minor amendments to approved applications should not require full data access committee (DAC) review. This will reduce backlogs and allow DACs to focus on projects that genuinely require scrutiny. In addition, dedicated fast-track processes should exist for urgent regulatory requests, such as safety signal validation, where a response may be required within days.

Direct access to pseudonymised individual patient-level data

Direct access to pseudonymised individual patient-level data is the primary reason the CPRD is so widely used by global pharmaceutical companies and regulators. It is essential for regulatory submissions, pharmacovigilance, cohort identification, rare disease research, and AI algorithm development. Federated eyes-off approaches, while valuable for feasibility queries and AI model training, cannot substitute for direct data access in regulatory and clinical research contexts.

Synthetic data – preserving the statistical, structural, and semantic properties of real datasets – should be provided as a standard capability across all HDRS-participating assets, not as an optional feature. The HDRS Digital Ecosystem Analysis recommends this as a system-wide requirement. Synthetic data enables rapid feasibility assessment without triggering a full access application, and is a practical prerequisite for achieving the 30-day access timelines pharmaceutical companies require.

Near-real-time and longitudinal data

Timely data updates are essential for post-market safety surveillance, cancer treatment evaluation, and clinical trial recruitment. The current two-year lag in National Cancer Registry data makes it operationally unusable for evaluating new cancer treatments within clinically meaningful timeframes. The HDRS should publish maximum lag times for each key dataset as performance standards from inception. For trial recruitment, near-real-time data is particularly critical because eligibility criteria are typically temporal – based on recent test results, medication changes or clinical events.

Longitudinal GP records, potentially spanning decades, are the UK's most distinctive competitive advantage. The HDRS should provide access from the earliest digitised records, with the full suite of coded data available to avoid excluding underserved patient populations.

Data quality assurance

All datasets made available through the HDRS should undergo validation and quality assurance checks, accompanied by metadata describing provenance, structure and quality. Common data models should be adopted to reduce variability across assets and facilitate linkage. However, source data should also be available alongside transformed datasets: variables may be lost in transformation and source access is essential for regulatory validation where evidence is scrutinised by the MHRA, National Institute for Health and Care Excellence and HTA bodies.

Data provenance: an essential regulatory requirement

For regulatory submissions, data provenance – the documented lineage from source system through all transformations to the final analytical dataset – is a legal requirement, not a quality enhancement. Regulators require the ability to trace this lineage precisely. Pharmaceutical companies cannot use data for regulatory purposes unless this chain is complete and auditable.

The HDRS Digital Ecosystem Analysis confirms that few UK assets outside the CPRD maintain documented transformation logic at the standard regulators require. Making end-to-end transformation provenance a condition of asset participation is what would keep HDRS data usable for high-value regulatory submissions, which depend on a complete and auditable lineage. Analytical pipelines involving AI or machine learning should document their logic explicitly: black-box processing breaks the provenance chain and disqualifies data from regulatory use.

Data versioning and archiving

Health data assets are not static: extracts are updated, linkages change and coding schemes evolve. An analysis run today may produce different results from the same code run six months later. For regulatory submissions, this is a fundamental problem – if the exact dataset version used cannot be recreated, submissions cannot be audited. Under MHRA guidelines, clinical trial data must be archived for at least 25 years.

The HDRS Digital Ecosystem Analysis finds that most UK assets operate rolling extracts where historical states are overwritten. Requiring all participating assets to implement point-in-time data versioning and archival standards, developed from the outset in consultation with industry, regulators and HTA bodies, would ensure analyses remain reproducible and submissions auditable – a precondition for regulatory use.

Protecting intellectual property

Exploratory feasibility assessments and incidence and prevalence analyses are commercially sensitive: early disclosure to competitors could undermine a company's research programme. Transparency policies should be proportionate – for example, delaying public disclosure of data use for a defined period after analysis completion. This protects intellectual property without compromising the principle of open registers, mirroring the transparency arrangements for clinical trials.

Importing proprietary code, algorithms and datasets

Pharmaceutical companies frequently use proprietary analytical code and algorithms developed specifically for their datasets. The HDRS should support import of external code, algorithms and datasets with appropriate intellectual property protection. This includes the capability to link NHS data to external datasets imported into a TRE.

Costing and contracting

Context

Financial sustainability requires a pricing model that is commercially competitive, transparent and equitable across user types – and that returns fair value to the data controllers on whom the HDRS depends. Contracting processes should be compatible with how global pharmaceutical companies operate.

Recommendations

Transactional pricing

Companies will pay a fair price for data and services when costs are internationally competitive and commensurate with quality. Transactional cost-recovery models, with published prices that enable project planning within internal budget cycles, are already familiar to companies. Adopting the same approach would let the HDRS price competitively against other markets while giving companies the predictability that draws their investment.

The pricing framework should be tiered: commercially competitive rates for industry, with appropriately subsidised access for academic and non-commercial researchers. This is consistent with the HDRS Digital Ecosystem Analysis'ssustainability model and keeps the service accessible across user types.

Pricing should also incorporate funding flow to data controllers. Done well, this reverses the misaligned incentives that have historically prevented NHS trusts and GP practices from investing in research-grade data pipelines, giving them a direct reason to build and maintain them. The HDRS should treat data controllers as partners, not simply as suppliers, compensating them for providing quality products through financial reimbursement, data engineering investment and operational benefits.

Commercially competitive timelines and service level agreements

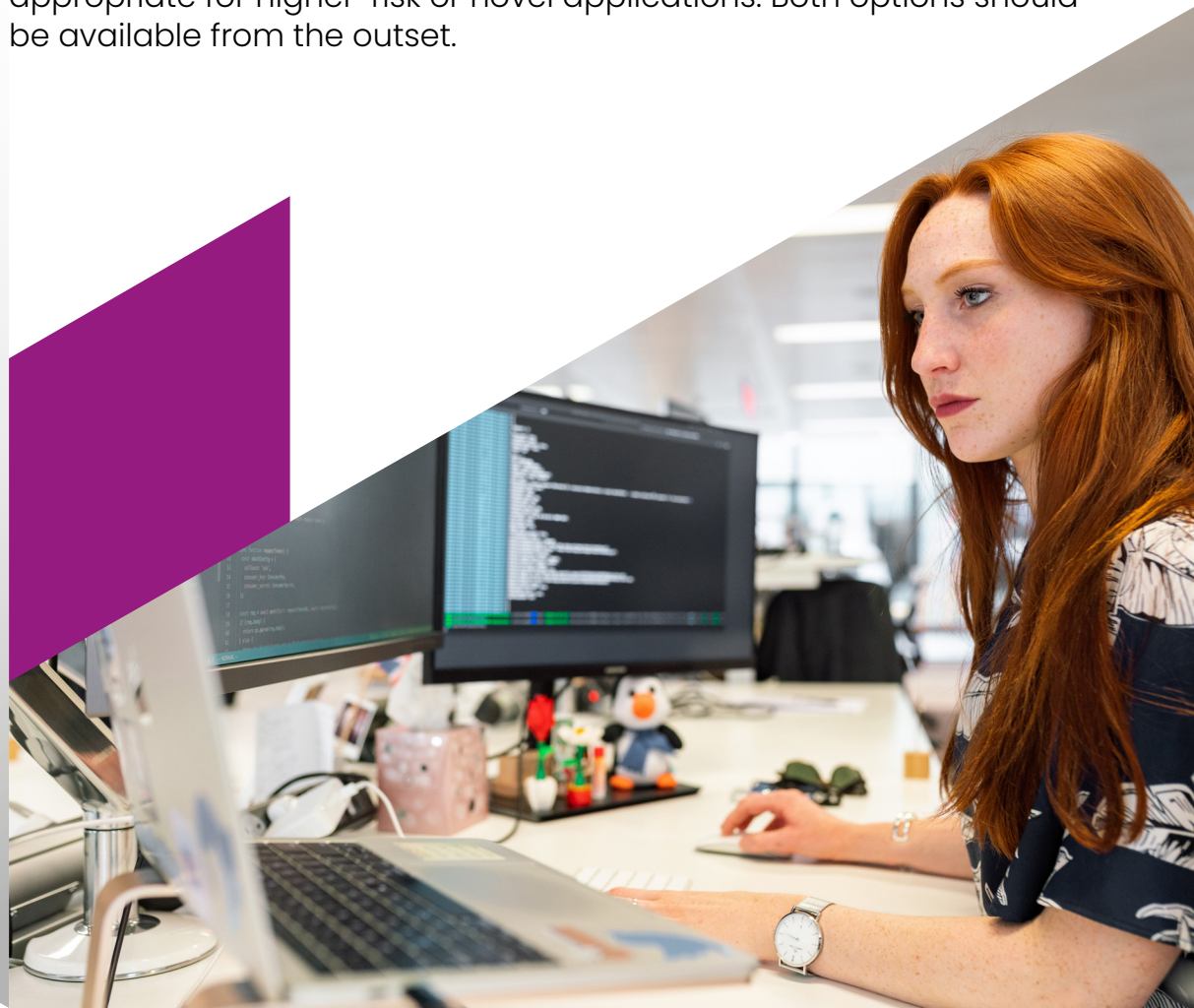
Committing to published service level agreements (SLAs) that specify performance standards and delivery timelines is how the HDRS can give companies the predictability they need to commit studies to the UK. ABPI members' median expectations, in calendar days, are:

- 30 days to receive a data access decision
- 12 days from approval to accessing data in the TRE
- 14 days from completion of analysis to release of results

The HDRS Digital Ecosystem Analysis confirms that governance unpredictability is the single most cited barrier to UK competitiveness, with some researchers reporting waits of 2.5 years from project setup to data acquisition. The HDRS should publish performance against these timelines from day one. A unified digital governance layer – with portable researcher credentials, machine-readable agreement templates and transparent project tracking – would replace the fragmented, opaque manual processes that currently make SLA commitments impossible to enforce.

Subscription and project-specific agreements

A subscription-based model, analogous to the CPRD's annual multi-study licence, should be available for regular users to cover multiple low-risk studies under a single agreement, avoiding repeated organisational contracting. Project-specific agreements remain appropriate for higher-risk or novel applications. Both options should be available from the outset.



Conclusion

The HDRS has the potential to transform the UK's position in global health data research and to attract substantial pharmaceutical industry investment. This report proposes, in specific and validated terms, what is needed to realise that potential. The recommendations are proportionate and implementable. Many are shared with other research users, others reflect the unique regulatory and commercial context of the pharmaceutical industry.

The overarching message from ABPI members is to focus on fundamentals: comprehensive, longitudinal GP data across all four nations; independent, standards-based data linkage; and predictable, commercially competitive service provision. These are the structural fixes that will benefit most researchers using the HDRS. Getting them right for industry means getting them right for most researchers.

The HDRS should build on what exists, not build anew. The CPRD, SAIL, Genomics England, SDE network, and the devolved nations' TREs represent decades of investment and established relationships. They should be harnessed and properly incentivised, including through funding flows that make data controller participation sustainable, not duplicated or displaced.

The challenges are not only technical – legislative solutions may be needed. Public trust should be built into the HDRS's design and governance from the outset, not added as a communications exercise. Care.data and GPDPR showed how quickly trust collapses and how hard it is to rebuild, but they equally show the prize for getting it right: a service the public actively supports, and that can therefore draw on the data needed to deliver.

A cross-sector strategic advisory group of representative user bodies, including the ABPI, should be established from inception to maintain focus on user priorities, surface implementation challenges early and ensure the HDRS roadmap is published and kept current. The pharmaceutical industry is committed to being a partner in this service's success and stands ready to work with stakeholders to design a UK service that will finally achieve the goal of improving research outcomes for patients at home and around the world, now and in the future.

Relationship to previously published work

This report builds on three previous ABPI publications – the joint ABPI/HDRUK paper,[6] (January 2025); the data-enabled clinical trials report,[7] (March 2026); and the joint Trade Association report on unlocking NHS data for research,[8] (February 2025) – and engages with the Wellcome-commissioned HDRS Digital Ecosystem Analysis[2] and HDRS Blueprint.[4] Where recommendations are consistent with those analyses, this is noted. Where they add specificity or reflect pharmaceutical company experience not captured in published reports, this is flagged explicitly. All recommendations have been validated with ABPI member companies.

Endnotes

1 ABPI, 'Headline pharmaceutical industry statistics', no date
2 Wellcome; Emrys Health; Nesta, 'HDRS Digital Ecosystem Analysis', 19 May 2026
3 ABPI, 'The principles for analysis and use of health data by ABPI members', 2022
4 Ghafur S, Waldock WJ, Leis M, Acharya A, Howitt P, O'Shaughnessy J, Darzi A, 'Building the future of UK health data: A blueprint for the Health Data Research Service', Imperial College London, 2025
5 ABPI, 'UK industry clinical trials: translating actions into impact', 2 December 2025
6 ABPI, 'Options appraisal to deliver a health data research service in England', 26 January 2025
7 ABPI, 'Globally competitive UK-wide data-enabled clinical trials: the time is now', 22 March 2026
8 ABPI, 'Unlocking NHS data for research - how to improve the regional Secure Data Environment network', 18 February 2025

About the ABPI

The ABPI exists to make the UK the best place in the world to research, develop and access medicines and vaccines to improve patient care.

We represent companies of all sizes which invest in making and discovering medicines and vaccines to enhance and save the lives of millions of people around the world.

In England, Scotland, Wales and Northern Ireland, we work in partnership with governments and the NHS so that patients can get new treatments faster and the NHS can plan how much it spends on medicines. Every day, our members partner with healthcare professionals, academics and patient organisations to find new solutions to unmet health needs.

www.abpi.org.uk

abpi

The Association of the British Pharmaceutical Industry

A company limited by guarantee registered in England & Wales number 09826787

Registered office: 2nd Floor Goldings House, Hay's Galleria, 2 Hay's Lane, London, SE1 2HB

RMI-0208-0826

Trusted Information Creator

Patient Information Forum