The Future of Patient-Generated Data for UK Health Research

Report from a Roundtable Event

2022

CENTRE FOR EPIDEMIOLOGY VERSUS ARTHRITIS

MANCHESTER 1824
The University of Manchester

EXECUTIVE SUMMARY

The uptake of smartphones and wearables has promised major opportunities for clinical care and health research for more than a decade.

In theory, people can contribute digital data from their own devices (also known as 'digital patient-generated health data' or PGHD) for population health research, with collection integrated into their daily lives. This has the potential to reach and involve much larger numbers of the public in health research; to study the day-to-day rhythms of wellbeing and disease; to collect a wider range of information about health and behaviour; and to provide information and insights back to individuals for public benefit via the same mobile devices.

Important questions that can be addressed using PGHD, and which are hard to answer by other means, include:

- To what extent should I expect day-to-day variation in my symptoms with this disease?
- What lifestyle changes, such as diet or exercise, will lead to improvements in the symptoms of my condition?
- Could monitoring of disease via my device guide care or treatment at the right time?

Examples are emerging of valuable large-scale population research using PGHD collected via consumer devices. Yet the scale of uptake and success has not yet matched the hyped opportunity. There are many barriers and potential pitfalls in the design, set-up and conduct of mobile health research studies. Successful delivery typically requires the majority of these barriers to be navigated simultaneously, perhaps explaining why we have not seen more tangible health benefits from such studies to date.

In Summer 2021, 23 stakeholder representatives, including researchers, health care professionals, members of the public, research funders and industry partners, participated in two roundtable discussions after completing a pre-meeting survey.

They considered key barriers and developed recommendations for what needs to be done, and in what way, to realise the future opportunities of PGHD for UK health research.

The **recommendations were grouped into four themes**, with interconnection between the areas.

Theme 1. ***Strengthening patient and public partnership throughout the study lifecycle***

- Patients and members of the public must be involved in designing every stage of the study lifecycle to optimise participation and long-term engagement, from direct-to-public recruitment through to sharing results with participants
- This continuous strong and inclusive patient and public partnership needs to be adequately resourced
- Building on existing best practice for patient and public involvement and engagement (PPIE), develop guiding principles for how to conduct successful PPIE in PGHD research

Theme 2. ***Advancing research methods for PGHD studies***

- Invest in methods development for population health research using PGHD, including study design and analysis methods
- Establish and share best practice for the design, conduct and reporting of studies using PGHD
- Develop and validate novel PGHD collection tools that are freely shared, allowing for future implementation across devices and platforms

Theme 3. ***Progressing technology to support PGHD studies***

- Progress towards platforms that are easily accessible and usable by researchers, and which work across devices and operating systems
- Ensure PGHD data collection tools align with FAIR principles, enabling re-use across studies and sharing across disease areas where appropriate
- Develop methods that provide confidence in derived metrics from sensor data

Theme 4. ***Developing trustworthy studies and systems***

- Develop systems that allow participants to access, control and view the flow of their PGHD, from self-administered eConsent through to linkage and onward data sharing
- For research requiring only PGHD, develop methods for secure processing of sensitive data on the user's device and associated federated data analysis methods
- For research requiring linked data, understand how PGHD can integrate into existing national infrastructure for secure data management, including enabling linkage to NHS datasets

INTRODUCTION

The increasing uptake of consumer electronic devices, including smartphones and wearables, has promised major opportunities for clinical care and health research for more than a decade[1,2]. In theory, patients and the public have the opportunity to contribute data more actively via their own devices, to supplement information collected by clinicians and researchers. This has the benefit of reaching and involving much larger sections of the public in health research; collecting a wider range of information about the occurrence and progression of disease; measuring health and behaviour in new ways, such as via touchscreens and other inbuilt technology including sensors; studying the day-to-day changes in health and wellbeing that are otherwise missed; identifying patterns of behaviour that may preceed or follow ill-health; and of opening new data-driven possibilities such as predictive analytics and just-in-time interventions.

All of these opportunities enable novel research questions to be addressed through descriptive, aetological, predictive and interventional studies (see Box 1), filling important gaps in our knowledge of health and disease. Examples are now emerging of valuable large-scale population research using consumer devices[3,4], although the integration of data from consumer devices into clinical care is lagging behind[5].

Digital '*patient-generated health data*' (PGHD) can be defined as electronic, health-related data created, directly recorded or gathered by or from patients outside of a clinical environment[6]. This might include self-reported data on symptoms, activities of daily living, lifestyle factors, as well as sensor data on physiological or behavioural biometrics.

The integration of PGHD into clinical settings has been slow due to complex and parallel challenges around patient and provider concerns, and technical and governance barriers[2,5]. It is expected, nonetheless, that technology will play a greater role in future healthcare systems, including the active involvement of people in their health and care[7,8]. Once this is achieved, *routinely-collected* PGHD may feed and serve health research, much like the secondary use of deidentified electronic health records. While that vision remains a considerable way from widespread adoption, there is the immediate opportunity to accelerate the active involvement of millions of people in health research via their own mobile devices, in turn improving the lives of current and future patients.

The longstanding promise of health research from consumer devices has not yet translated into widespread delivery, perhaps reflecting that it is harder than expected. Nonetheless, in the last decade, there have been many examples of insightful and innovative research using consumer devices, alongside informative experiences of the challenges of delivering such research successfully. This collective experience allows us to take stock, reconsider the opportunities, and importantly to understand the challenges for moving towards a shared vision as a diverse - and necessarily cross-disciplinary - field and community.

Technology will play a greater role in future healthcare systems, including the active involvement of people in their health and care.

We considered these opportunities and challenges by convening a roundtable event to hear the perspectives of a wide range of stakeholders. Contributors were invited to include perspectives from patients and the public together with representatives from academia, National Health Service (NHS), funders, the technology industry and information governance (see Appendix 1 for contributors).

We held two two-hour online workshops that included plenary and smaller group discussions, complemented with capturing of ideas on an online whiteboard. Contributors filled out a pre-meeting survey (see Appendix 2) to inform discussions, submitting up to three important research questions they believed could be addressed through large-scale collection and use of PGHD, and up to three challenges or barriers that, if solved, would help significantly accelerate national research using PGHD. The survey results and subsequent discussions have been summarised in this report, initially drafted by Prof Will G. Dixon, refined by the wider Manchester team, then circulated for edits, comments and approval by the roundtable attendees.

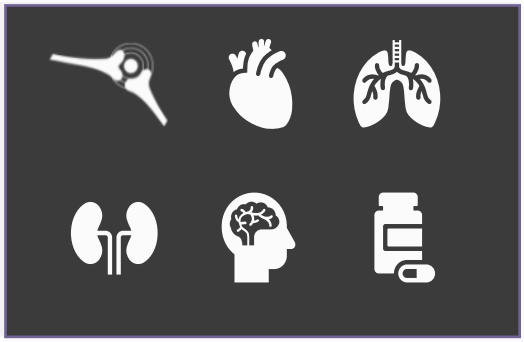
Opportunities for addressing novel research questions using PGHD

During the first workshop, roundtable contributors discussed research questions where PGHD could allow important questions to be addressed that are hard to answer using other means (Box 1). Some questions were pre-prepared to prompt discussion, and others were generated during the workshop. This list included *descriptive* and *aetiological* questions that benefit from collecting temporally-rich data integrated into people's lives, and *predictive* analytics to support timely interventions made possible by frequent self-reported or sensor data processed using machine learning techniques. Questions could be answered for an individual (i.e., *personalised*) as well as for the wider population, including understanding the impact of a (digital) health *intervention*.

Box 1. Examples of research questions addressable using patient data from consumer devices

A: Descriptive
Understanding how disease fluctuates through time, and its impact on the lives of people living with disease

e.g. "What symptoms contribute most to poor quality of life and disability in people living with multimorbidity?"



B: Aetiological
Examining causal relationships between changing health or lifestyle exposures and outcomes, such as severity of disease

e.g. "What dietary exposures lead to an exacerbation of symptoms in patients with inflammatory bowel disease?"

C: Predictive
Identifying real-time predictors of changing patterns of disease to enable just-in-time interventions

e.g. "Can changing patterns of physical activity monitored via a wearable device identify new-onset, or a deterioration in, neurological disease that could trigger an automated action for patients, carers or clinicians?"

D: Personalised
Examining aetiological and predictive research questions *within the individual*, as well as across populations

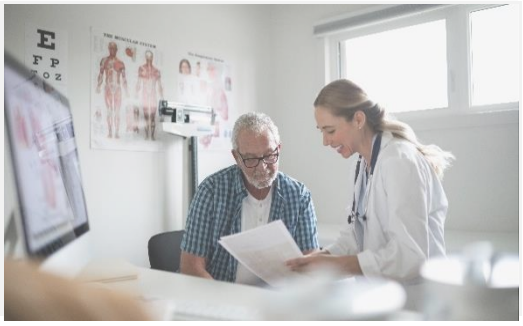
e.g. "Do any of these over the counter supplements actually work for *my* arthritis, and which works best?"

E: Interventional
Understanding how digital interventions can lead to improved health and wellbeing

e.g. "Does physical activity coaching, delivered though a mobile device, lead to a reduction in cardiovascular events?"

F: Health services research
Studying the use, uptake, delivery, costs and outcomes of incorporating PGHD into clinical care and systems

e.g. "Can patient data collected between NHS visits improve care and patient outcomes?"



Challenges/barriers to realising the opportunities at pace and scale

Many people in the clinical research community and their patient partners already understand these opportunities. They have important research questions that are amenable to mobile health research. However, there are many barriers to realising these opportunities at pace and scale.

Our roundtable pre-meeting survey generated over 40 challenges that were diverse, at times overlapping or interconnected, and often multifaceted and complex. During the meetings, the group clarified and elaborated on some of these challenges. Challenges were then mapped to the lifecycle of a mobile health research study (Figure 1), while recognising that we could also have grouped these challenges thematically or by relevant stakeholders.

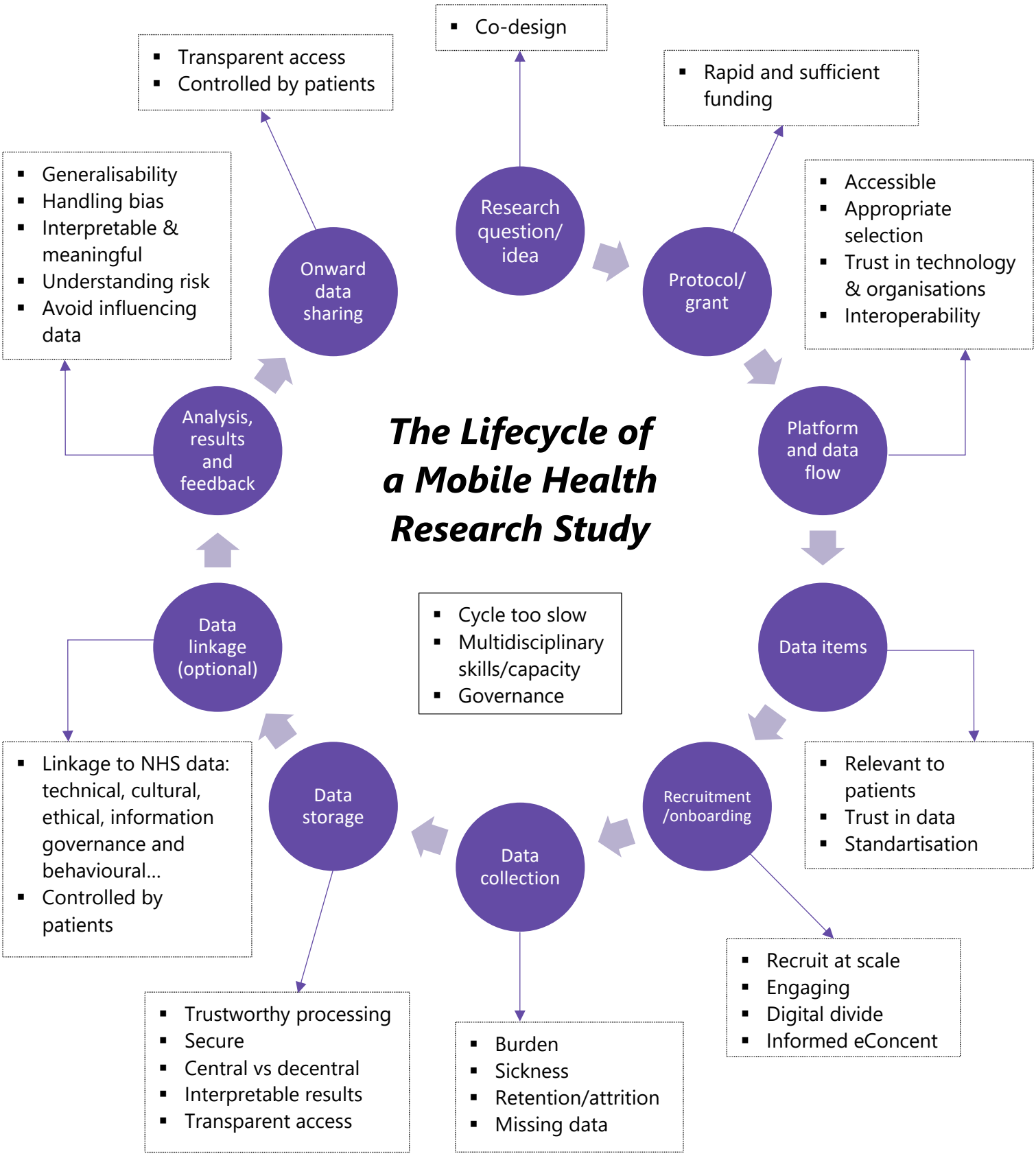


Figure 1. Challenges for the successful delivery of PGHD research at scale, listed as bullet points and aligned with the stages of a research study, shown on the purple circles.

Themes

Discussion focussed on themes considered to be the highest priority *specifically for PGHD-centred research*, rather than on challenges that were common across health research in general.

Four key themes were generated and prioritised, expanded on below:

1. Strengthening patient and public partnership throughout the study lifecycle
2. Advancing research methods for PGHD studies
3. Progressing technology to support digital PGHD studies
4. Developing trustworthy studies and systems

Theme 1. Strengthening patient and public partnership through the study lifecycle

While all health research benefits from patient and public involvement, this is especially true for PGHD research: such studies are critically dependent on the public proactively joining the study, then contributing data to the research, often repeated through time. This requires *motivation and ongoing engagement*.

The group recognised that researchers often start by deciding what data they need to answer their question, then demand that study participants provide this data without adequate consideration of the participant's perspective and motivations. If this does not change, many future PGHD studies will fail.

Key recommendations

- Patients and members of the public must be involved in designing every stage of the study lifecycle to optimise participation and long-term engagement, from direct-to-public recruitment through to sharing results with participants
- This continuous strong and inclusive patient and public partnership needs to be adequately resourced
- Building on existing best practice for patient and public involvement and engagement (PPIE), develop guiding principles for how to conduct successful PPIE in PGHD research

Research using PGHD that is not appealing or understandable is at risk of low recruitment and high attrition. Failure to consider participants' perspectives also risks addressing questions that have limited positive impact on people's lives. As articulated by a roundtable contributor, we need to learn "how to make the research interesting, engaging and vital for people to take part in". For this reason, and

perhaps even more so than in other types of research, it is important to provide feedback about the research to which people have contributed, and to maintain trust in the use of personal and potentially sensitive data (see Theme 4). These goals align closely with the NHS Health Research Authority's recent #MakeItPublic strategy[9].

What's needed

Patients and the public need to become an equal partner in the prioritisation, design and conduct of PGHD research.

Patients and the public need to become an equal partner in the prioritisation, design and conduct of PGHD research. Although seemingly intuitive, this explicit recommendation is vital if PGHD studies are to be successful because recruitment and ongoing engagement are paramount.

This partnership may include individual patients or patient and public groups and organisations. Furthermore, it is essential that the research community should recognise the importance and benefits of strong patient and public partnership, rather than it being seen as an unnecessary burden or a 'nice to have'.

Successful partnership should include:

- Bringing together researchers, technology partners and patients at the earliest stage of identifying the best ideas for PGHD research studies that are both feasible and have significant potential benefit, thereby building mutual trust and securing support at later stages
- Requiring meaningful patient and public involvement at the early and late stages of projects, for example mandating descriptions of involvement within funding bids or ethics applications, recognising patient-led research and patients as co-investigators, and adequately resourcing patient and public involvement throughout the study lifecycle
- Understanding what motivates patients and the public to engage with the study during the design phase, for example selecting important and meaningful outcome measures or designing feedback loops (without compromising scientific rigour)

- Discovering the most appropriate set of data to collect, balancing what data the researchers need to answer the research questions, with what motivates the participants, and what data they are willing to contribute through time
- Actively considering how under-represented groups can best be included in PGHD research, for example by considering access to technology and the costs of connectivity
- Transparent communication throughout the study with research participants and the public, including the study aims and proposed outcomes – specifically how the research will make a difference; a description and justification of what data are being collected and why; understandable descriptions of data flow, security and protection, and how access is controlled; simple and proportionate consent processes; provision of feedback of data and insights in ways that do not affect scientific rigour; a data usage report, or 'data receipt', confirming who has accessed the data (see Theme 4); and clear and understandable communication of the study results

It was noted that future success will require guiding principles to support researchers in how best to involve patients and public contributors as they move through all stages of the study lifecycle.

Theme 2. Advancing research methods for PGHD studies

Epidemiology has shifted from studying communicable diseases to investigating the causes and consequences of non-communicable disease.

New PGHD sources will allow a different time-scale to be studied: we can now measure between- and within-day changes in exposures alongside fluctuations in disease. This may herald a further shift towards studying disease progression (where symptoms are dynamic through time), rather than the presence or absence of disease.

Key recommendations

- Invest in methods development for population health research using PGHD, including study designs and analysis methods
- Establish and share best practice for the design, conduct and reporting of studies using PGHD
- Develop and validate novel PGHD collection tools that are freely shared, allowing for future implementation across devices and platforms

Digital health interventions may be delivered via people's own devices, the exposure to which is measurable in new ways such as screen-time, clicks and 'likes'. New opportunities such as novel, multimodal time-series data require an extension to existing research methods.

Yet, the underlying principles of good population health research still apply and must be revisited for this new era of PGHD. This will allow us to learn when we can (or cannot) rely on the results from such research. Robust critical appraisal and interpretation of study findings using novel methods will be important for all stakeholders, including patients and the public, clinicians, researchers, journal editors and regulators.

What's needed

New PGHD sources will allow a different time-scale to be studied: we can now measure between- and within-day changes in exposures alongside fluctuations in disease.

The lifecycle of a study using PGHD shown in Figure 1 can help us think through what is needed to advance digital epidemiological research. Frequent, longitudinal PGHD collection integrated into people's everyday lives allows us to re-think **study design**. Within-person designs, such as case-only designs and n-of-1 studies, and micro randomised trials to develop and evaluate components of personalised adaptive interventions are particularly opportune for PGHD but need further development. For example, the need to extend methodology for how to analyse daily ordinal time-series data, or for sample size calculation for a series of n-of-1 studies to detect with confidence a change for an individual and identify distinct sub-groups of responders. Study designs need to be more agile to address the current gap between fast technology change versus the slow pace of evidence generation, while ensuring scientific rigour is retained. Methods development and evaluation will inform future best practice for study design and ensure investment is linked to an ability to answer the question at hand.

New **PGHD collection tools** will move us from an era of physical examinations and questionnaires to active self-administered data entry using touchscreens, microphones and cameras, plus passive data collection using sensors such as accelerometers, gyroscopes, GPS and photoplethysmography. The development and validation of these instruments takes on even greater importance as devices differ in their hardware (e.g. screen size), operating systems and software applications. The design and usability of PGHD collection tools is important as researchers seek to balance the quantity and

comprehensiveness of data versus the data entry burden and participant retention. PGHD collection tool development thus needs interdisciplinary teams skilled in user interface design, software engineering, psychometrics, behavioural and social science, as well as clinical researchers and epidemiologists. To facilitate large-scale uptake, funding to support novel tool development and validation could have a requirement to make the tool widely available and easy to implement in data collection platforms, for example via a Software Development Kit, to benefit the wider PGHD research community and ultimately the general and patient populations that they serve.

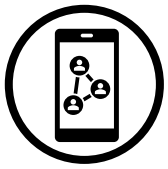
Developing methods to tackle, or at a minimum to understand, inclusion and diversity in PGHD studies will be key.

Population health PGHD studies will inevitably have different patterns of recruitment and attrition compared to other study designs. **Recruitment** to PGHD studies requires us to learn how to initially attract participants, and to successfully support them through eligibility criteria, self-administered eConsent[10] and baseline data collection. Developing methods to tackle, or at a minimum to understand, inclusion and diversity in PGHD studies will be key since device ownership is skewed towards the more affluent and younger populations. It is important to consider the cost implications of study participation, such as connection costs, and to ensure this research area does not widen the digital divide. Once people have shown initial interest, maintaining mutually beneficial **engagement** is a priority. Global smartphone statistics show one in four people abandon apps after using them only once, while only 32% of users return to an app 11 times or more. Good engagement can be supported through feedback via the device, yet we need to ensure feedback does not generate information bias or misclassification. Attrition tends to happen not at random, hence it is important that the community learns where possible selection biases *into and out of* a study impact on study validity[11]. Prior to conducting large-scale studies, it may be necessary to perform feasibility studies to test and evaluate recruitment and longitudinal engagement. Although these early feasibility studies may not provide an immediate return on investment, subsequent large-scale studies will consequently become more efficient, reproducible and reliable.

The research questions in Box 1 will benefit from advancements in **data processing and analysis methods** as well as improvements in their substrate of richer data. Patient-generated symptom data will often generate a time-series of ordinal data that has floor and ceiling values and missingness. It is not yet clear how longitudinal data of

this type should best be summarised. How, for example, should a flare in pain be defined? Sensor data, e.g. from accelerometers and gyroscopes, generates further challenges of signal processing and standardised metrics, especially when consumer devices use proprietary algorithms to generate apparently similar metrics such as step count. Validating the outputs of (processed) sensor data collected in free-living environments can be difficult if there is an absence of 'ground truth'. And the challenges continue: as machine learning methods develop predictive algorithms of, say, a disease flare using all available PGHD, does that algorithm need to be explainable? Can it change through time? How should it best be implemented within self-management or clinical care, and what is the appropriate governance for its use in these different settings? Future methods development for personalised digital interventions could consider how we best identify what intervention should be delivered for whom, when should the intervention be delivered, and when should it end.

In time, as best practice is defined, it will be important to have **guidelines** for the conduct, analysis and reporting of PGHD research studies as a reference for those judging grant applications, reviewing publications of research findings and interpreting research results.



Theme 3. Progressing technology to support PGHD studies

An early step of the research study lifecycle (Figure 1) requires the research idea to be converted into a data collection system. Here, tools to collect the right data items are hosted on a platform alongside additional 'modules' that manage other aspects of the study such as consent and set-up, and feedback to participants. The system then needs to manage the data, either on-device or securely transmit data to become accessible to researchers.

Key recommendations

- Progress towards platforms that are easily accessible and usable by researchers, and which work across devices and operating systems
- Ensure PGHD data collection tools align with FAIR principles, enabling re-use across studies and sharing across disease areas where appropriate
- Develop methods that provide confidence in derived metrics from sensor data

Many studies develop their own apps or configure existing systems, which requires software and data management expertise within the research team, or external technology partnership. Finding the right tech partner can be

challenging for academic researchers and hard to evaluate by funding bodies, while multiple providers of study software leads to non-standardised and non-interoperable data, variable quality of associated modules, and less reproducible research. Systems, or apps, may run on only one operating system, introducing the selection biases described above. Embedding (novel) validated PGHD collection tools into existing platforms is not commonly done, yet will be needed if the community wishes to have high-quality interoperable data.

Continuing the current *status quo* will result in more non-interoperable data, inefficient investment in multiple platforms of variable quality, high attrition of participants, lower confidence in secure data management and, ultimately, lower quality research and less public benefit.

What's needed

Accessible, easy-to-use and configurable study platforms that support research across devices and operating systems are a priority.

Accessible, easy-to-use and configurable study **platforms** that support research across devices and operating systems are a priority. Research teams without major technical expertise need to be able to set up a well-designed mobile health study, ensuring they collect the right data using the best possible tools. These data collection tools must be embedded within software that also supports all necessary interactions with the participant including self-administered consent (see Theme 4) and feedback, thereby also supporting optimal user engagement (recruitment and retention).

Robust, secure data management must also be supported. Apple's ResearchKit exemplifies several aspects of this requirement: it provides open source predesigned screens to "make it easy to create a beautiful research app that's customized for your study and enjoyable for people to use". However, it only supports research on Apple devices (whose owners typically have higher salaries) as well as requiring researchers to set up the back-end database. Tech companies are unlikely to make software for rival operating systems, hence the community needs to solve this difficult challenge. Future solutions may be achieved through a set of standards against which platforms are judged, a library of reusable and validated data collection tools, the development of - and support from - a research design service including technical advice, or even the national provision of an (open source) study platform.

Future platforms need to support research across any disease area, including studies of multimorbidity and multiple long-term conditions. Remote monitoring and digital epidemiology are cited as a major opportunity in this area, which requires 'developing measures to collect, link, store and share appropriate data and outcomes for multimorbidity'[12].

Data collection systems must take a 'patient-centric' approach in order to avoid asking about the same symptom multiple times for different conditions, for example pain, fatigue or breathlessness, and thereby risking high attrition. Elsewhere in healthcare, common data standards are fundamental to the interoperability, usefulness and efficiency of data. Routinely collected health data are adopting controlled vocabularies like SNOMED and communication standards like HL7 FHIR, yet PGHD remains in its infancy: standards are not widespread and may not even exist. There is an important opportunity for new PGHD collection tools to align with **FAIR principles**: ensuring that they are Findable, Accessible, Interoperable and Reusable[13].

Future PGHD studies need to consider how data items can 'interface' between diseases and across different platforms to optimise efficiency while retailing quality.

Future PGHD studies need to consider how data items can 'interface' between diseases and across different platforms to optimise efficiency while retaining quality. For wearables, derived metrics from sensors also need to move towards transparent and reproducible research. Possible solutions include reference algorithms that are open-source with unrestricted rights to use, that the community of practice maintain (e.g. activity identification from accelerometer data), and validation of proprietary 'black-box' output metrics (perhaps where accuracy is accredited by a third party).

Joining together the novel validated PGHD collection tools from Theme 2, with FAIR principles and accessible platforms presented here in Theme 3, one can imagine a future searchable library of data items, associated tools and metadata that can be selected to efficiently build the front- and back-ends of a research study. If multiple studies were to be hosted on a single or linked platform(s), it would also be possible to identify live studies to which people could participate according to their characteristics and the inclusion criteria of the studies. This would solve a longstanding problem of empowering patients and the public to contribute to health research studies – in this instance from the comfort of their own home.

Theme 4. Developing trustworthy studies and systems

Health data is personal and sensitive, whether derived from electronic health records, self-reported information on diagnosis, symptoms and quality of life, or derived from passive sensors. As PGHD creates an increasingly granular picture of individuals and expands outwards to other sensitive data types, such as geolocation, the need for secure and trustworthy systems increases. The public need to trust that any data they collect and share is being used and protected appropriately.

Key recommendations

- Develop systems that allow participants to access, control and view the flow of their PGHD, from self-administered eConsent through to linkage and onward data sharing
- For research requiring only PGHD, develop methods for secure processing of sensitive data on the user's device and associated federated data analysis methods
- For research requiring linked data, understand how PGHD can integrate into existing national infrastructure for secure data management, including enabling linkage to NHS datasets

There has been important progress in national infrastructure to support UK health data science through the development of Trusted Research Environments (TRE). These environments manage researcher access to deidentified health data, built on the principle of the Five Safes framework, and with robust and independent accreditation, monitoring and auditing[14]. A key goal is to provide "transparency for public and patients as to who is accessing the data and for what purposes". At present, this has been focussed on EHR and administrative data and less on PGHD, or how PGHD can be linked to these other data sources.

After the collection of PGHD on consumer devices, the data can remain on device (Figure 2A), or can be transmitted and stored centrally (Figure 2B). Centralised data could be on a server hosted by the organisation conducting the research, or held by the organisations that provide consumer services (e.g. a fitness tracking company). It might also be managed by an accredited service like a TRE (Figure 2C), where additional benefits may be derived through data linkage.

It is important that people are able to control the flow of their various data types for health research, as before always being clear who is accessing the data and for what purpose.

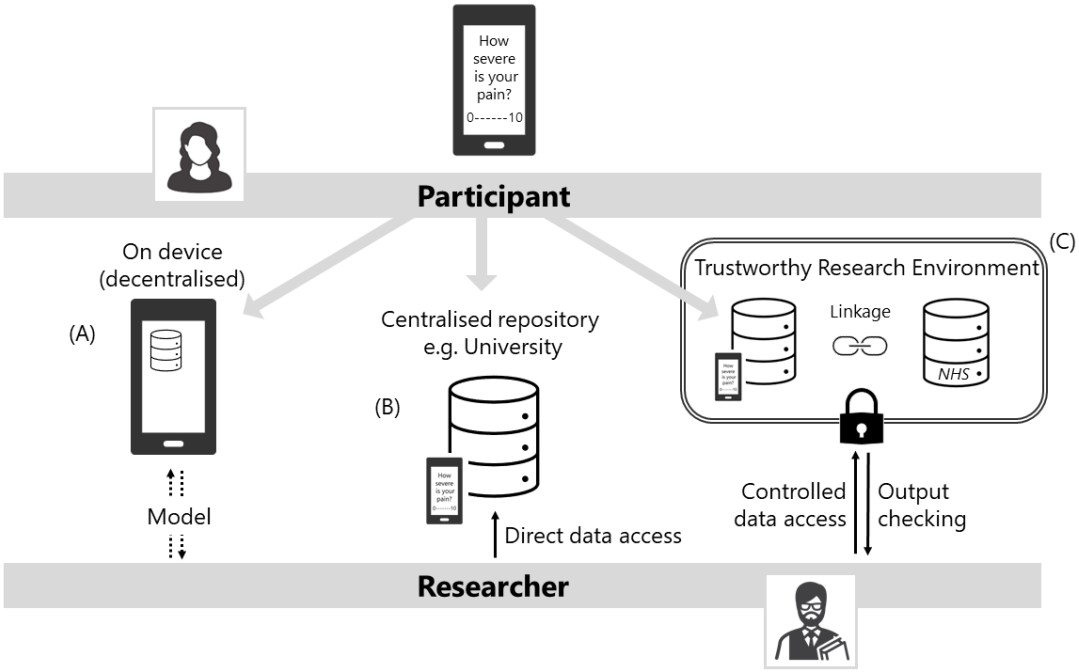


Figure 2. Options for data access and analysis of PGHD. PGHD can remain on device (A), be transmitted and stored centrally by the host organisation (B), or be managed in a Trustworthy Research Environment, potentially with linkage to other data sources such as NHS data (C).

What's needed

People need to **be in control** of what data they collect and who they share it with for what purpose, where their options are presented in an understandable way. Stages of control include, firstly, the ability to complete self-administered consent for PGHD collection that meets all elements of informed consent including informedness, voluntariness, and comprehension[10].

People should then be able to share and link their own consumer data, such as past tracked activity data, even if it is stored within a company's server. Thinking beyond individual research studies, collected data may be useful if made more widely available to others, especially if data items are collected in line with FAIR principles to enable pooling to create a larger cohort of participants. This step of onward sharing could amplify the value that is derived from the contributions of study participants, yet participants should be able to guide with whom they are willing to share their data and for what purpose, and be confident that their choices are respected.

The sensitivity around partnership with the technology industry using health data means transparency about use and purpose is particularly important. Data control processes thus need to be flexible to handle initial participation, incorporation of other data streams, data linkage and onward sharing, and dynamic to support where preferences change through time. Furthermore, they need to provide transparency by giving information back to the participant about how their data has been used.

Personal data stores - using standard, open and interoperable data formats - and their associated access controls may provide a solution to some of these challenges.

A traditional model for health data research has been 'data to researcher', where data is (sent to and) stored on local servers. TREs support a 'researcher to data' model where researchers are provided with remote access to a secure data environment. A third, emerging option is **federated learning** without the data leaving the device. In this model, a person's study data remains on their device without centralised data collection, with only aggregated results becoming available more widely[15]. If research uses only PGHD (i.e. requires no linkage to other data sources located elsewhere), then this emerging area shows important potential for trustworthy health research. Future work should help the community understand its role, educate and support for its wider use.

Federated learning is not expected to solve all problems because certain research questions will require **PGHD to be linked** to other data sources where patient-reported and clinician-reported data items complement one another. Routinely collected health data are not widely held on individuals' own devices, but instead held within the NHS or as deidentified datasets within TREs – thereby ruling out a federated learning, 'on-device' solution. It makes sense that the linkage of PGHD to NHS data uses the existing and expanding infrastructure of TREs which solve many problems of the trustworthy use of health data. For example, once within a TRE, auditing of data access could support public reports of compliance and even direct feedback to individuals, thereby enhancing transparency. How, when, by whom and under what conditions PGHD can be securely and accurately linked to ultimately deidentified EHR data held within a TRE remains an important challenge. As this challenge is tackled, it will be important to bring together key NHS organisations, researchers and the public to work through the legal issues and (importantly) expectations of all parties as to ownership and custodianship of the linked data.

Novel health data collection and analysis methods currently operate at the boundaries of (or beyond) what's covered clearly by existing rules and regulations[16]. For example, certain data collected from non-medical settings might now be used in health-related analytics, such as inferring health conditions based on patterns of physical activity. Such activities don't fall neatly under the General Data Protection Regulation (GDPR) as it relates to health data or by other guidelines and regulations. Researchers may increasingly find themselves working in regulatory grey areas which might lead to issues that diminish trust and patient wellbeing. It is important that regulatory authorities play an active role in providing clarity around responsibilities for PGHD for use in both clinical care and research.

CONCLUSION

Population health research using PGHD collected through people's own devices is a long-standing promise that, so far, has failed to deliver at scale. Raising the ambition and quality of this opportune and exciting area will require the establishment of an **interdisciplinary community**. This community will develop the composite foundations for future high-quality research, in turn supporting a wider group of interested - but so far inexperienced – researchers by defining and supporting best practice. New state-of-the-art exemplar PGHD research studies could showcase and extend best practice. Investment is needed for these four themes of public involvement and engagement, study design and methods development, data and platforms and governance, with expansion to also support the early and late resourcing required for PGHD studies that falls outside of traditional funding cycles. This would include funding for the prioritisation and co-design with patients in the lead up to funding bids, and support for data management, participant feedback and onward data sharing beyond the end of a specific project's funding.

We need to learn how expertise from academia, the technology industry (e.g. software engineering, data management, user interface design), funders and others can come together for mutual benefit, while recognising the vital importance of maintaining transparency and public trust. Increasing porosity between academia and the tech industry may catalyse joint working to solve real-world problems using PGHD, and help navigate the malaligned timescales of academic research and funding with faster-moving and more changeable industry roadmaps. This could be achieved through a national Centre of Excellence or network, linked into other wider health data research institutes, networks, initiatives as well as external partners.

People living with disease have many unanswered questions that could be addressed through mass participation in health research studies using their own devices. There are many challenges to address, but this is a worthy endeavour with major opportunities - all deliverable now - to improve the lives of current and future generations.

REFERENCES

1. The Academy of Medical Sciences, “Our data-driven future in healthcare” (2018).
2. M. Deering, “Issue Brief: Patient-Generated Health Data and Health IT” (2013), (available at https://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf).
3. M. V. Perez *et al.*, *N. Engl. J. Med.* **381**, 1909–1917 (2019).
4. C. Menni *et al.*, *Nat. Med.* **26**, 1037–1040 (2020).
5. I. Sim, *N. Engl. J. Med.* **381**, 956–968 (2019).
6. M. Shapiro, D. Johnston, J. Wald, D. Mon, *Prep. Off. Policy Planning, Off. Natl. Coord. Heal. Inf. Technol. Res. Triangle Park. NC RTI Int.*, 35 (2012).
7. “The NHS Long Term Plan” (2019), (available at www.longtermplan.nhs.uk).
8. P. Mistry, The digital revolution: eight technologies that will change health and care (2020), (available at https://www.kingsfund.org.uk/publications/digital-revolution).
9. NHS Health Research Authority, “Make it public: transparency and openness in health and social care research” (2020), (available at https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/research-transparency/make-it-public-transparency-and-openness-health-and-social-care-research/).
10. Sage Bionetworks, “The elements of informed consent: a toolkit” (2020), , doi:10.3928/0098-9134-19940801-10.
11. K. J. Rothman, J. E. Gallacher, E. E. Hatch, *Int. J. Epidemiol.* **42**, 1012–1014 (2013).
12. The Academy of Medical Sciences, “Cross-funder multimorbidity research framework” (2019), (available at https://acmedsci.ac.uk/file-download/49628715).
13. M. D. Wilkinson *et al.*, *Sci. Data*. **3**, 160018 (2016).
14. UK Health Data Research Alliance, “Trusted Research Environments (TRE)” (2020), (available at https://ukhealthdata.org/wp-content/uploads/2020/04/200430-TRE-Green-Paper-v1.pdf).
15. A. Sadilek *et al.*, *npj Digit. Med. 2021 41*. **4**, 1–8 (2021).
16. Ada Lovelace Institute, “The Data Will See You Now” (2020), (available at https://www.adalovelaceinstitute.org/wp-content/uploads/2020/10/The_data_will_see_you_now_Ada_Lovelace_Inst_Oct_2020_Final-min.pdf).

APPENDIX

Appendix 1: Roundtable attendees

Name	Title/ Role	Organisation
Justin Bingham	Chief Technical Officer	Janeiro Digital
Laura Dickens (Meeting 2 only)	Associate Director Industry Partnerships	MRC UKRI
Rob Harle	Research Team Lead	Google Fit
Bruce Hellman	CEO	uMotif
Neha Issar-Brown	Director of Research	Versus Arthritis
Benjamin James	Public Contributor	n/a
Rhod Joyce	Deputy Director of Innovation Development	NHSX
Lynn Laidlaw (Meeting 2 only)	Public Contributor	n/a
Dave Leon	Professor of Epidemiology	LSHTM
Federica Marinaro	Research Informatics Engagem ent Manager	Alzheimers Research UK
Mike McConnell	Senior Clinical Lead, Mobile Health & Connected Care; Clinical Professor of Medicine	Google Health & Stanford University
Catherine Moody (Meeting 2 only)	Head of Population Health	MRC UKRI
Leanne Morrison	Lecturer in Health Psychology	University of Southampton
Aidan Peppin	Senior Researcher	Ada Lovelace Institute
Jo Roach	Executive Director of Platform & Products	Our Future Health
Cathie Sudlow	Professor of Neurology and Clinical Epidemiology	University of Edinburgh & Health Data Research UK
Colin Wilkinson	Public Contributor	n/a
UoM team		
John Ainsworth	Professor of Health Informatics	UoM, Centre for Health Informatics
Will Dixon* (convenor)	Professor of Digital Epidemiology	UoM, Centre for Epidemiology
Elaine Mackey	Research Information Governance Manager	UoM, Centre for Epidemiology
John Mcbeth	Professor of Chronic Disease Epidemiology	UoM, Centre for Epidemiology
Niels Peek	Professor of Computer Science	UoM, Centre for Health Informatics
Sabine van der Veer	Senior Lecturer in Health Informatics	UoM, Centre for Health Informatics

*Will Dixon has received consultancy fees from Abbvie and Google

Appendix 2: Pre-roundtable survey

Topic 1: Research opportunities

Patient-generated data unlocks new opportunities to support UK health research, for example involving much larger numbers of the public in health research, and collecting richer and more diverse data about disease and patterns of behaviour. This in turn allows us to address research questions that have previously been difficult to address

Please provide up to three important research questions you believe can be addressed through the use of patient-generated data.

For example:
Can we enable the early detection of disease (e.g.) through passively monitoring changing patterns of physical activity collected via sensors in mobile consumer devices (smartphones and wearables)?

Note: these questions can require patient-generated data alone, patient-generated data linked to NHS data, or linkage to additional data

1. 2. 3.

Topic 2: Challenges and barriers to realising the opportunities at pace and scale

Members of this roundtable event have important collective experience of the challenges and barriers to achieving impactful research using patient-generated data at scale, across the whole lifecourse of a research study from design through delivery and recruitment to the implementation of results. Issues might include selecting what data items to collect and why, using what tool on what platform, establishing successful cross-disciplinary partnerships, how to recruit and retain participants, issues of data standards and interoperability, identifying best practice in research design and analysis, or governance considerations about controlling data access and wider data sharing.

From your perspective, please list up to three major challenges or barriers that, if solved, would help significantly accelerate national research using patient-generated data

1. 2. 3.