Prostate Cancer Research

pcr.org.uk

UNLOCKING THE POTENTIAL OF PROSTATE CANCER DATA

A report by Prostate Cancer Research to explore the needs, value and potential of prostate cancer data

Unlocking the potential of Prostate Cancer Research

Contents

Forward

Prostate Cancer Research is a research and patient information charity focused on delivering breakthrough medicines and treatments for prostate cancer. We have been shortlisted for an impact award for our patient engagement work, and we use our deep understanding of patient priorities to direct our academic and translational funding where it will have the most impact.

Putting patient experience at the heart of research and empowering patients to actively engage with research are both essential in order to deliver scientific innovation that meets patient needs and preferences. Making best use of data collected as part of healthcare provision and collecting data from patients themselves could be transformational.

We are delighted that the patients we work with feel the same. Over 2,500 patients have already pledged their support for a registry that will provide consent to link to their clinical data. In addition, through our Health Inequities Programme we are connecting with patients from different racial communities and those with low health literacy, helping to ensure they are better represented in data used to make decisions about new treatments and diagnostics.

We are excited to launch this report to share our findings. In the next step of our journey, we will be seeking additional input and ongoing dialogue with stakeholders, including patients to further develop our plans. In doing so, we will build a strong and collaborative prostate cancer data community, with patients and patient experience at its core.

PCR is actively seeking partners to help drive forward the resourcing, design and delivery of a new prostate cancer data solution. To be part of the next chapter please get in touch with the office of the CEO in the first instance.

Oliver Kemp
CEO Prostate Cancer Research
info@pcr.org.uk

Executive summary

Prostate cancer as an exemplar of data-driven innovation

Despite the therapeutic improvements of the past decade, in the UK one man still dies of prostate cancer every 45 minutes, and 250,000 die each year from prostate cancer worldwide.

Innovative thinking and new approaches are required to solve this problem, and existing treatments need to be applied more effectively.

1. What we mean by health data – data related to "health conditions, reproductive outcomes, causes of death, and quality of life" for an individual or population. Health data includes clinical metrics along with environmental, socioeconomic, and behavioural information pertinent to health and wellness.
2. Why we want to collect data – to gain greater insight into the biological, clinical, and environmental factors that impact a patient's outcome.
3. What we will achieve in the long-term – improve outcomes for those with prostate cancer by enabling and accelerating the discovery of new treatments, diagnostics, and tools.

The natural history of prostate cancer is heterogeneous. Together with the complexity of treatment and care pathways, this makes prostate cancer an ideal condition to test how applying a structured approach to the management of health-related data can unlock its potential. Making the best use of real-world clinical and patient-reported data could be transformational. However, to do this requires collaboration between patients, healthcare providers, and commercial life science organisations.

Vast amounts of health-related data are generated and collected by the NHS and other care providers and research studies. The NHS is working to deliver system-wide interoperability with a long-term vision to link health and social care data, a vision backed by a £200m investment: VISIT SITE [1]. These ambitions and the guiding principles for their delivery have been clearly articulated in the Life Sciences Vision: VISIT SITE [2], NHS Long Term Plan: VISIT SITE [3], Digital Health and Social Care plan: VISIT SITE [4], Goldacre Review: VISIT SITE [5], and Data Saves Lives: VISIT SITE [6]. There is a near term opportunity for PCR to leverage the investment that is happening at scale across the NHS in order to realise the benefits for prostate cancer patients.

LINKS

1. www.england.nhs.uk/blog/collaboration-across-the-system-to-increase-the-privacy-protection-and-speed-of-life-saving-research
2. www.gov.uk/government/publications/life-sciences-vision
3. www.longtermplan.nhs.uk
4. www.gov.uk/government/publications/a-plan-for-digital-health-and-social-care/a-plan-for-digital-health-and-social-care
5. www.goldacrereview.org
6. www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data

The value of patients

An added consideration is the largely untapped potential to actively collect patient-reported data in the community. Increasingly patients are looking for ways to engage with charities and to share their experiences to further medical research.

Although there are an increasing number of ways to capture different aspects of patients' experiences, there is currently no central mechanism to capture data from men living with prostate cancer in the UK over time.

Establishment of a registry to collect patient-reported data and linking this with clinical data (to create a "Platform" providing for access to linked data) will provide researchers, patients, and health providers with a much greater understanding of key factors that impact a patient's outcome.

To begin to understand the requirements for linking patient-reported data and clinical data we undertook the following:

- A review of the current data landscape and data initiatives in the UK
- Extensive consultation with stakeholders (including hosting 4 roundtables and 3 patient focus groups)
- Sharing of findings and recommendations with representatives of each stakeholder group for review and comment

Recommendations

1. In order to derive the maximum benefit of investment, ensuring linkage capacity should be a priority, with an emphasis on collaboration rather than duplication. Working with the planned national NHS Secure Data Environment (SDE) and the developing NHS data ecosystem should be prioritised. Resources should be focused to address the challenges in specific uses of data that are likely to persist.
2. Patients should play a key role in governance and oversight of the platform, providing input into design and delivery.
3. A primary purpose of a prostate cancer data platform should be the enablement of translational and clinical research. In designing the platform, be mindful of opportunities to support discoverability and access to tissues and samples linked to clinical and outcome data.
4. Taking into consideration the needs and preferences of stakeholders, a core minimum prostate cancer clinical data set should be defined, and an assessment made of existing data collections as source. Similarly, a core quality of life, socio-economic and lifestyle dataset for collection through a patient registry should be developed.

Introduction

Health data is a social good, access to which can enable the development of diagnostics and treatments and improve patient care.

The UK is heavily invested in delivering a health data research infrastructure that enhances patient care and supports innovation. NHS-wide improvement in the data infrastructure is disease agnostic. Creating disease or use specific data partnerships between the NHS, charities, academia and life science companies will be essential in order to unlock the full potential of health data in the future.

In this report, we highlight the challenges that academic and commercial researchers have in accessing prostate cancer data to date and examine the extent to which the roll out of the new NHS data infrastructure will overcome these challenges in the future. We have focused on identifying the ways in which stakeholders might collaborate to enhance the data landscape for prostate cancer translational and clinical research, and in particular the vital contributions that patients have to offer.

Our research and consultation with stakeholders have enabled us to describe the need and value proposition for a collaborative, multi-disciplinary prostate cancer data platform with patient involvement and engagement at its core. We also begin to envisage what the prostate cancer data platform could look like and the next steps to be undertaken to achieve this.

About prostate cancer

Despite the therapeutic improvements of the past decade, in the UK one man still dies of prostate cancer every 45 minutes.

Innovative thinking and new approaches are required to solve this problem, and existing treatments need to be applied more effectively. The heterogeneity of prostate cancer and its treatment, and the diversity of experience amongst patients is vast. Improving access to relevant clinical data, together with patient-reported data is key to unpicking this complexity, driving excellence in study design, and targeting care and treatment.

While there are many effective options for treating prostate cancer contained within the prostate gland, the side effects of treatment – of which the most common are urinary incontinence and erectile dysfunction – can have a long-term impact on the patient's quality of life, emotional well-being, and mental health. When it comes to therapies for locally advanced prostate cancer, where the cancerous cells have grown and started to spread outside of the prostate, the options are more limited and often have devastating side effects for patients including fatigue, nausea, and pain.

For the 13% of prostate cancer patients diagnosed with advanced prostate cancer, where the tumour has spread to other parts of the body such as bone and lymph nodes, treatment options are extremely limited. And for those who were diagnosed at an earlier stage but whose prostate cancer has spread to distant sites and no longer responds to hormone therapy, the prognosis is especially poor. If a new prostate cancer platform were able to help address any of these issues it would transform the futures of men with a diagnosis of prostate cancer.

Footnote: bit.ly/3Cq8t7L VISIT SITE

What prompted this report?

- Patients tell us they want to give appropriate access to their data so their experiences can better inform research in a way that respects their privacy.
- Academics tell us the lack of access to data dramatically slows research.
- Companies tell us they want support with clinical trial design and recruitment, and ensuring patient-centricity in their research.
- Clinicians tell us that improvements are needed in data linkage to improve patient care and support rapid uptake of innovations for patient benefit.

We set out to answer the following questions:

- How can we ensure that patient benefit is at the heart of the use of health data?
- What are the unique perspectives of the multiple stakeholders?
- What other data initiatives are relevant and how could we work/link with them?
- What are the potential technical solutions, resources, and operational considerations?

In undertaking this project we:

- Reviewed the nature of and uses of existing relevant data collections.
- Explored case studies, from the literature presented at workshops, of existing data platforms that enable the health data from different sources (such as GPs and hospitals) and for specific patients to be brought together (data linkage).
- Hosted a series of 3 patient focus groups and 4 thematic expert workshops with key stakeholders to address specific challenges and explore opportunities for collaboration. The workshops were held under Chatham House Rules. The observations and recommendations in this report are drawn from the presentations and discussions that took place, as well as several one-to-one meetings.

"The heterogeneity of prostate cancer leads to challenges in choosing the right treatment, at the right time, for the right patient. This makes prostate cancer an ideal exemplar to test the premise that applying a structured approach to the curation and linkage of health-related data over time will unlock the full potential of current data collections to inform better treatment for patients."
Professor Rakesh Heer (Chair of Urology, Imperial College London).

The power of patients

"I hate seeing patients and families suffer - this is what drives my decision to share my data."

PCR & Patient engagement

"I am pledging my support because my experience matters. Join me because yours does too."

Stephen Fry, PCR supporter, writer, actor and presenter has kindly led the support for pledges to the new PCR registry.

- **Registries are databases containing information about individuals who are affected by a specific condition.**
- **The data in registries can be entered by patients themselves, by their doctor or by a combination of the two.**
- **A Patient Reported Outcome (PRO) is that outcome which is directly reported by a patient without the interpretation of the patient's response by a clinician or by anyone else. PROs pertain to the patient's health, quality of life or functional status.**

The ethos of Prostate Cancer Research (PCR) is to place patient views at the heart of our work. Our latest social research study has shown us that most patients want to help develop the next generation of diagnostics and treatments.

We wanted to know whether prostate cancer patients would be willing to consent to data use and to actively engage and contribute to the collection of health-related and quality-of-life data.

To assess the viability of a data platform and to prepare the ground for its delivery, we launched a pledge campaign through our patient networks. Headed by actor Stephen Fry, the number of campaign pledges rose to over 2,500 within weeks of the campaign launch, against an initial target of 1,000.

Patient insights are becoming increasingly important in the evaluation of innovations and can provide evidence relating to efficacy, safety and patient impact. The use of PROs is supported by an expanding body of literature that demonstrates associated PROs with traditional outcomes such as overall survival.

Data generated by patients in relation to their condition, such as their symptoms, social and psychological impacts and physical functioning, without the interpretation of a clinician, is a valuable contributor to realising this aim. Physical activity can be a powerful predictor of LT clinical outcomes in cancer, as can symptoms such as pain, anxiety, and fatigue. Registries have value for patients; as well as helping to drive forward innovation and research, they can provide direct benefit to patients, for example by enabling them to have a better understanding of their disease and treatment.

(Gotay et al, The prognostic significance of patient-reported outcomes in cancer clinical trials. J Clin Oncol. 2008 10:26(8) 1355-63 and tumour response (Victorson D, et al, Metaanalysis of the correlation between radiographic tumour response and patient-reported outcomes. Cancer 2006; 106(3):494-504).

In 2023 PCR will launch a new patient education and empowerment platform, "the infopool" with funding from the National Lottery Community Fund.

The infopool will help people with a diagnosis of prostate cancer to better understand possible treatment options, including the potential impact of side effects on their quality of life, and help them to make better informed decisions. While it is going to help all men with prostate cancer, it has been specially designed to address the unmet needs of those with low health literacy and people in the Black community, who are at higher risk. The infopool will help to educate people about the importance of patient data and encourage patients to engage with the patient registry.

One patient's story

Peter (pictured here with his son) is a PCR supporter. His experience illustrates the physical and psychological effects of prostate cancer.

Capturing real-world experiences, such as Peter's, via a registry could enable researchers to better predict a patient's outcome and identify the challenges and benefits of new and existing treatments.

"In 2001 when I was 59, my doctor suggested I have a full blood test. This and a following test showed that my PSA was rising steadily, and I was diagnosed with prostate cancer. I was treated with 20 doses of radiotherapy over a 4-week period. As a result, my PSA fell.

I had 6 monthly check-ups until in 2010, my PSA started rising again and I was told that it was incurable. My doctor told me that my likely lifespan with the drugs available at the time was 3-5 years. I was treated with hormone tablets, then hormone injections, and then both together. This treatment kept my PSA low for 4/5 years.

After this, I was then put on a course of chemotherapy, but I had to stop after 2 treatments due to the dangerous effect on my liver. At this stage, I was told that the only drug available to me would be effective for about 6 months. Luckily, with my son Mike due a visit from Australia, I decided not to take the drugs. When Mike was leaving, I was offered a trial of a drug too expensive to be available on the NHS. The median time of effectiveness was 12/15 months. Now three and a half years later, my PSA is still below 1."

Peter has now been on treatment for 15 months. He has taken to tennis, table tennis and croquet, and is much fitter. This has helped his outlook on life.

PCR & patient engagement

Patient views from focus groups

In exploring the potential to establish a prostate cancer data platform, we have examined the extent to which patients' willingness to improve and inform research translates to active participation and desire to share health data. The following insights and recommendations are derived from our structured discussions with men previously diagnosed with prostate cancer. The views of participants taking part in one or more of our focus groups on perceptions and preferences relating to data and tissues are set out below.

Appetite for engagement

There was a significant appetite for active engagement in data sharing. The opportunity to actively contribute data (through donating experiences, Patient Reported Outcome Measures (PROMS), and responding to surveys) was viewed as a personal benefit that would be "good for my mental health" as well as having an impact on physical health. Of the possible practical personal benefits that might be derived through data sharing most highly valued were:

- The opportunity to play a role in shaping research
- Better access to clinical trials.

"It's common sense that data should be shared for the benefit of others. I'm all for it. I can't imagine why anyone would object."

Who is most likely to engage?

Focus groups felt that those with chronic or life-threatening conditions who are well informed and who interact with patient advocacy or support groups are amongst those most likely to engage, and therefore hold positive views towards data sharing. The involvement of trusted advocacy and support groups was highlighted as being extremely important.

"I want to be seen, to share how I really am."

Increasing engagement

Communication and engagement were seen as key to building patients' trust and confidence in data sharing. The TED talk style of communication and the sharing of personal stories (of the impacts of prostate cancer and how data sharing delivers benefits) were suggested as good ways of increasing willingness to share data. The feel-good factor of committing the altruistic act of data sharing was seen as a personal benefit.

Benefits to society

Our focus groups expressed strongly the view that health data should be used to improve care and treatments for the benefit of patients and society. The support or encouragement of family (in terms of active participation in data donation, for example through PROMs) was posited as a potential inducement for participation in data sharing.

Videos which we shared that carried the NHS logo and emphasised interactions in health-care settings were well received.

"This is what the NHS is about; looking after me and looking after society."

Broad consent and the concept of "legacy data"

In agreeing to share data and access to samples and tissues, our participants highlighted the importance of knowing that their involvement would continue to add value after their death (or in the event of mental incapacity). The term "legacy data" resonated with our group, capturing their hope that their participation might deliver benefit for future generations of prostate cancer patients.

"I feel it is vital that [my data] continues to be used so that my son has the best chance of avoiding a similar fate."

Tissues and samples

All of our focus group participants were very comfortable with the notion that tissues, and samples linked to clinical data should be shared with companies and researchers. The key consideration related to personalisation and targeted medicines, with participants highlighting the opportunities that might accrue to them personally in treatment terms as a result of the willingness to share. Our participants expressed the importance of being able to provide feedback relating to the use of their data and samples. When sharing tissue for research, patients felt that there needed to be a way to prevent duplication of analysis to prevent wastage of precious materials. One way of doing this would be to share the results of tests carried out on tissue where repetition might be avoided.

"I had a prostatectomy and consented to the sharing of the tissue for research. I was later told I couldn't be included in a clinical trial because so many people had used my prostate tissue that there was none left for the testing that would need to be done for the trial."

PCR & patient engagement

Genetic and genomic data

Some participants expressed greater levels of comfort in relation to the sharing of genetic profiling of tumour tissue as opposed to germline genomic information. Participants expressed concerns that the sharing of germline genomic data could have negative (as well as positive) implications for family members that relate not only to health but to financial considerations (such as impacts on health and life insurance premiums). For this reason, the sharing of germline genomic data was viewed as worthy of specific and separate consideration.

Inclusivity of data

Our focus group discussed the importance of inclusivity in data collection, with some participants stating that they would be less willing to share their health data if any groups were under-represented in the data pool. A key priority for PCR will be ensuring equity, inclusion and diversity.

Apps and wearables

Participants expressed concerns about the potential sharing of data from Apps and wearables, and sought clarity on the types of Apps and the nature of data to be collected. The "one-way" nature of data sharing was seen by some as problematic and there was strong support for the proposition of active feedback and participation.

Ethical use of data

Patients underlined the need for assurances relating to the robust interpretation and ethical use of data. In relation to genetic/genomic testing, the use of test results by insurers to load premiums or decline policies was raised. The existence of the moratorium on the use of genetic tests by insurers was seen as positive, but concerns remained that there is no cast-iron guarantee that the moratorium might be dismantled at some point in the future.

Our focus group participants stressed the importance of ensuring that the "right" breadth and mix of clinical data is collected from the outset, suggesting examples of symptoms and conditions that can impact treatment choices, treatment responses, and overall quality of life. Examples of key co-existing conditions that are important to include are mental health and osteoporosis.

Data sharing

The sharing of data between health care providers sparked a discussion of whether the data might be used to deny access to treatment. The concern centered around Covid-19 and press coverage of decisions relating to de-prioritising of patients with underlying conditions and disabilities for treatment in intensive care units and for access to ventilation.

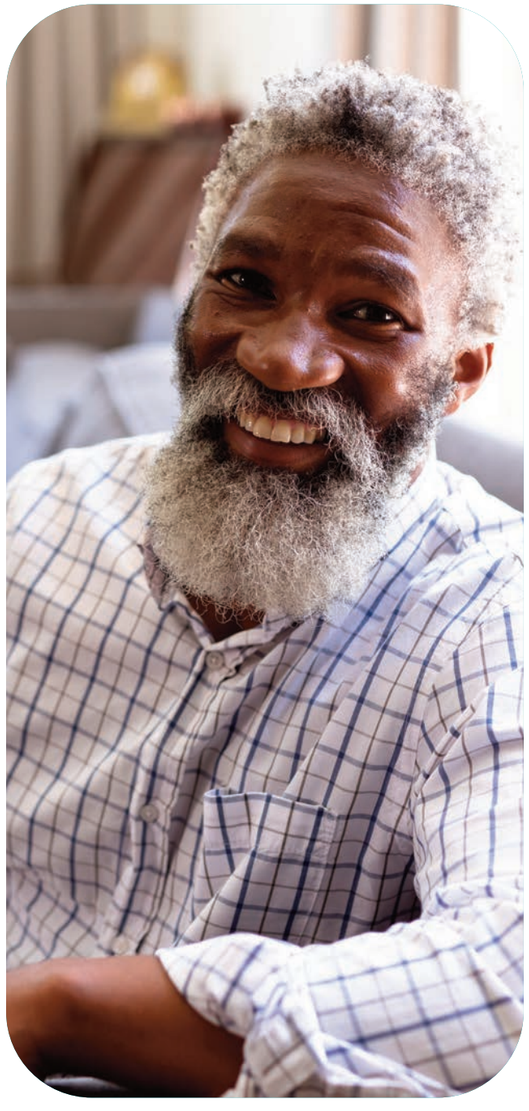
PCR & patient engagement

Sharing of anonymised data with pharmaceutical and biotechnology companies was supported by all participants in our focus group, with comfort levels being on a par with researchers, charities, and the wider NHS. The potential for companies to ultimately profit from data-enabled product development was not considered contentious, though the need to ensure suitable plough back of financial benefit to the NHS was strongly supported.

The "why" of sharing with pharma and biotech was identified as being important, with discussions echoing our earlier findings (survey July 2021) that level of positivity towards working with pharma and biotech is linked to the level of understanding/familiarity an individual has of the industry.

Privacy and trust

Health data is sensitive and personal. Whilst most people are generally supportive about the use of their data for research and public benefit they also want to be sure that use of their data will not compromise their privacy. Half of our focus group participants were very comfortable with sharing their data, and half were "a little uncomfortable". Those expressing negative views relating to data sharing highlighted concerns about privacy and data security. The aggressive exploitation of data by marketers and search engines (Google) suggested that "privacy is an illusion". The decision to opt-out of health data sharing for one participant had been down to the perception that anonymised data is "easy to unscramble".



The consent model used in traditional research studies does not necessarily serve for consenting to the use of health and social data in research. A consent model that provides for an ongoing dialogue or relationship between the platform and the participant, with a strong governance framework inclusive of patient representation are likely to be essential in terms of maintaining trust.

"I am 100% satisfied [with respect to privacy protection] because the identifying data is destroyed [prior to sharing with commercial companies]."

The changing health data landscape

Every UK citizen has a ten-digit NHS number that remains with them from cradle to grave. This number is unique and associated with any health care they receive in their lifetime. For each patient, there is both health and medical information contained in electronic health records (EHRs - detailing primary care) and electronic medical records (EMRs - detailing episodes of secondary or tertiary care).

For some patients, genomic or genetic data and data held in registries or social care settings may also be available. Vast amounts of health-related data are generated and collected by the NHS and other care providers, academic and commercial research studies and increasingly by individuals, for example through wearables and devices, and this data is currently stored in a multiplicity of data silos (see Annex 2). The shortcomings of the multiplicity of approach to health data collection and curation across the NHS, and solutions to address these, have been highlighted by the Goldacre Review (April 2022). VISIT SITE [1]

The NHS is currently working to deliver system-wide interoperability and there is a long-term vision to link health and social care data. Secure data environments were identified in the Data Saves Lives strategy (June 2022) as the default way that NHS health and social care data will be accessed for research and analysis in future. These changes in the data research infrastructure will offer new opportunities for collaborative opportunities that bridge the gaps and maximise the value of health and social care data for specific use cases.

Core to the strategy is provision for access to data through Secure Data Environments (SDEs). The strategy sets out ambitions for a radical programme of reform that will reduce the number of platforms in use across the NHS through the creation of a small number of national broad, and disease agnostic SDEs. The National SDEs will form a network with subnational SDEs (working to the same framework), and the system will likely operate a single "front door" for access to NHS data in the future.

These changes will deliver efficiencies in data linkage for Trusted Research Environments (TREs) such as disease data platforms (such as the British Heart Foundation Data Science Centre) and specific, pioneering initiatives established to help bridge gaps in existing systems.

An example of such an initiative is the National Pathology Imaging Cooperative to bridge the gap between pathology and genomics data. VISIT EXAMPLE [2]

1. www.goldacrereview.org
2. REF: go.nature.com/3Td0GRK

Case studies

The British Heart Foundation Data Science Centre is a partnership between Health Data Research UK (HDR UK) and the British Heart Foundation (BHF).

It accesses data within the NHS Digital Secure Data Environment, funded by the NHS England Data for R&D Programme.

The BHF works with patients, public, clinicians, researchers and NHS organisations to help them carry out research relating to prevention and treatment of all diseases of the heart and circulation. The centre started in January 2020 (£10million investment over 5 years).

The platform provides for linkage of:

- National, population-wide (coded) health data
- Unstructured data, for example imaging, medical free text and electrocardiograms
- Data from apps and wearables.

The platform aims to support:

- The development of methods, tools and protocols through collaborative networks
- Recruitment of patients into clinical trials
- Linkage of large "omits-rich" cohorts to electronic health records.

The UK Multiple Sclerosis Register serves as an exemplar of the value and practice of linking patient reported data with routinely collected clinical data. Registry MS Register (Swansea University / SAIL).

The MS Register was the world's first register, for any condition, to combine information direct from patients, with clinical and NHS data.

Funded by the Multiple Sclerosis Society and launched in 2011, the register currently has over 17,000 members who are asked to regularly update information on their treatments and the impact MS is having on their lives. Over 800,000 questionnaires have been completed and the register now contains 7 years of longitudinal data.

People with MS who are treated at one of the MS Register's 40 NHS partner sites can give their consent for their medical information to be securely transmitted to the Register.

Their clinical details can be 'linked' to their questionnaire responses.

Secure Data Environment (SDE)

A SDE is a secure computing environment that holds data and provides remote access to health data for approved researchers. TREs enable the highest standards of information governance, transparency and security by removing the need for data to be physically shared between users. The concept of TREs embeds the "Five Safes Framework", a set of principles originated by the Office for National Statistics (ONS).

Safe data: data is treated to protect confidentiality.

Safe projects: research projects are approved by data owners for the public good.

Safe people: researchers are trained and authorised to use data safely.

Safe settings: SecureLab environment to prevent unauthorised use.

Safe outputs: screened and approved outputs are non-disclosive.

Artificial Intelligence (AI):

Computer systems developed to perform tasks that normally require human intelligence.

Machine Learning (ML):

A subset of AI that can build mathematical models based on a set of training data, in order to make predictions, and improve these predictions over time with repeated exposure to the data.

Natural Language Processing (NLP):

A component of AI that enables computers to understand and process unstructured text and extract meaning from it.

Analytical methods

The complexity and rise of data in healthcare means that artificial intelligence (AI) will increasingly be applied within the field. Several types of AI are already being employed by payers and providers of care, and life sciences companies. The key categories of applications involve diagnosis and treatment recommendations, patient engagement and adherence, and administrative activities.

The role of AI to analyse clinical data and real-world data was raised during our consultations. It has an increasing role in healthcare where its application can be useful in clinical practice as a tool for personalizing treatment, supporting decision making, and managing polypharmacy.

The development of AI for use in key stages of clinical trial development, from study preparation, recruitment of patients through to execution is another example of how AI could offer an unparalleled opportunity to deliver transformative benefits.

Incorporating into a RWE strategy, AI, ML and NLP innovations can lead to enhanced drug development and increase the likelihood of success, through defining optimal treatment pathways by:

- Understanding patient behaviour
- Automating identification of suitable participants
- Predicting (and providing for avoidance of) participant drop-out.

We have not explored analytical methods such as AI as a separate section within this report, although it is an area that will be of huge importance in the delivery of our data goals. The feasibility stage and development of a prostate cancer data platform will include the capability to incorporate AI to process and analyse clinical and patient-reported data.

www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181 VISIT SOURCE

Workshops

Four workshops elicited the views of key stakeholders regarding the current data landscape.

The view that the potential value of longitudinal health data is immense is not controversial. And unlocking the potential of UK health data has been a government priority for many years. However, the challenges inherent in delivering a comprehensive solution in line with the strategies and ambitions of the NHS mean that it will be some time before a comprehensive solution delivering defragmented and accessible data capable of enabling research is in place.

The complexity, heterogeneity and often long natural history of prostate cancer make it an ideal exemplar to test the premise that applying a structured approach to the curation and linkage of health data over time will unlock its full potential.

Summary of workshop findings

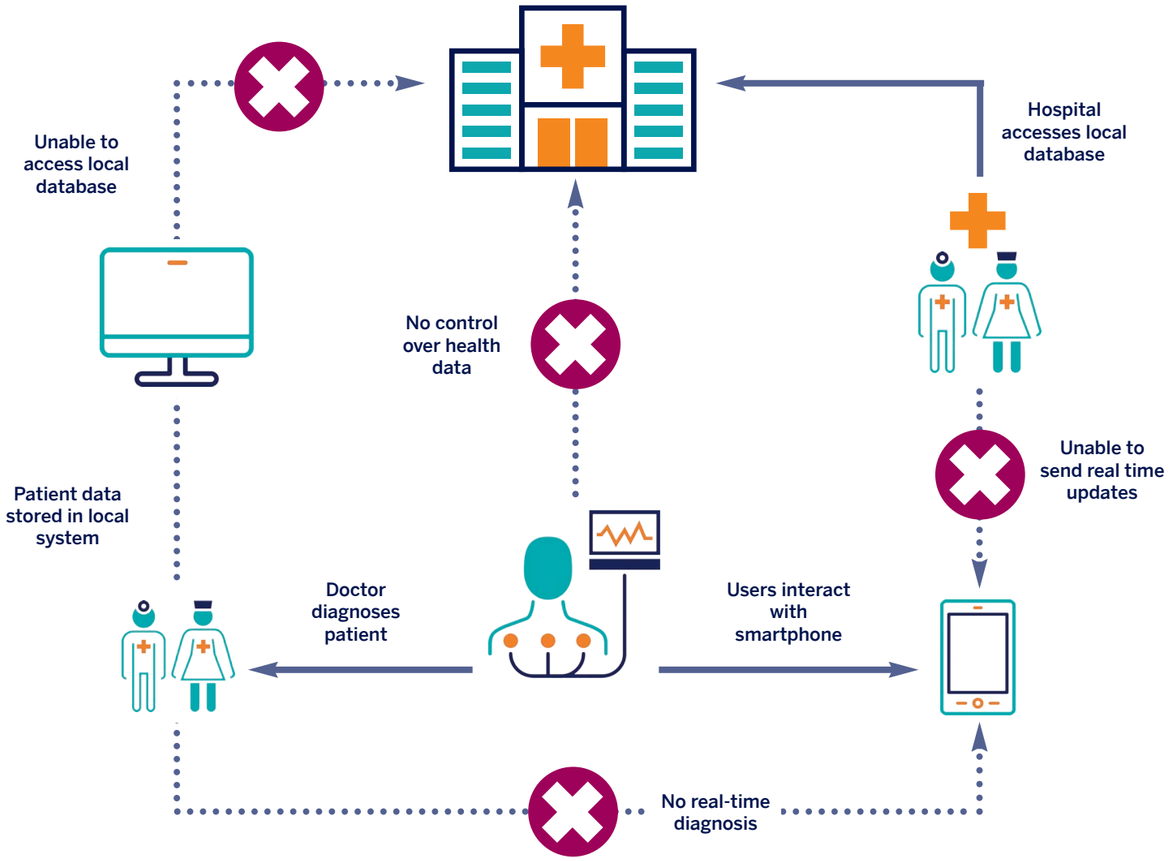
All workshop stakeholders agreed that there was huge value in the analysis of real-world data alongside patient-entered data to generate insights that deliver value to patients, researchers, health care providers, companies, and the wider economy. In our first three workshops, attendees were invited to discuss the challenges and opportunities in harnessing clinical data and real-world data and assessing its potential.

Challenge: Data silos

Fragmented data that sits in silos is a challenge. Siloed data creates barriers to information sharing and collaboration between researchers, health providers and patients. Due to inconsistencies in data that may overlap across silos, data quality often suffers. When data is siloed, it's also hard to get a holistic view of a patient's health in terms of their trajectory through the care pathway and other health information which could be critical to their response to diagnosis and treatment.

Inconsistencies in data capture, coding and storage, together with the time lag inherent in national data sets, drive many commercial and academic researchers to generate and retain their own data sets. The challenges inherent in sourcing data from across the devolved nations, where there are differing systems and practices are further compounded by differences within nations that have been set out by local data custodians. Whilst the creation of National and sub-national Secure Data Environments for access to NHS data will bring about substantial improvements, these stand to support the operations of disease specific initiatives, operating as TREs, rather than negate the need for their creation.

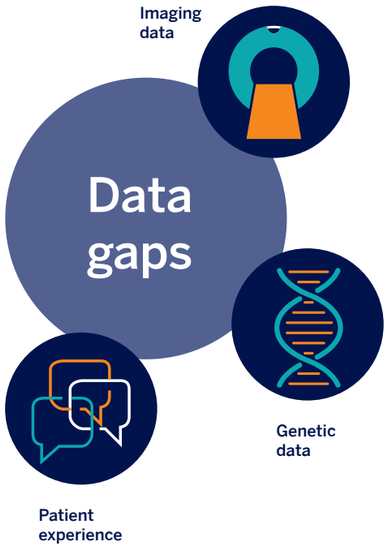
- Too many silos
- Different systems/system changes
- Different processes
- Different ethical procedures
- Formatting
- Linkage issues/ inter-operability



Challenge: Data quality

The following experiences and issues were highlighted:

- Errors at point of entry
- Bias and inequality of data. Reporting bias (also known as selective reporting) takes place when only a selection of results or outcomes are captured in a data set, which typically covers only a fraction of the entire real-world data. It is people's tendency to under-report all the information available
- Consent issues
- Poorly curated data
- Lack of validation
- Narrow focus with hypothesis-driven data sets do not support other uses.



Challenge: Data gaps

Stakeholders agree that there is a wealth of data that is either not routinely collected electronically, or not routinely captured.

- Patient-Reported Outcome Measures (PROMs) & Patient-Reported Experience Measures (PREMs) not being better able to identify patient needs and experiences
- Data gaps e.g. nutrition, social care, activity. Significant volumes of potentially important data do not make it into electronic medical records, ambulatory status for example
- Longitudinal data
- Imaging data, including "hidden data" such as metadata, time intervals, type of scan
- Genetic, Genomic and other omics data.

Challenge: Data access

Speed of access to data was viewed as highly problematic, with comparisons made between national systems and local systems and divergence of process and practice between the four nations of the UK. A single point of access for an integrated or linked data sets could help:

- Overcome issues associated with poor visibility/discoverability of data sets
- Multiplicity of access requests and therefore associated costs
- Ease of access to subsets of use-relevant data
- Override complexities inherent in national and local practices.

Opportunity: Defining and delivering a core data set

Identifying core data that could be collected consistently to serve the needs of multiple users was seen as a key opportunity to increase the value of data. A core data set would need to be developed as part of a focused feasibility study but would include existing data relating to:

- Primary diagnoses
- Test results including PSA
- Hormone sensitivity
- Referrals
- Prescriptions/treatments
- Mortality
- Symptoms.

Workshop participants were asked for their views on the types of additional data that should be included as core to any future prostate cancer data platform. The following were highlighted as having core value across multiple stakeholders:

- Quality of life
- Imaging data, with the Image Exchange Portals that provide for transfer between care providers as an enabler
- Data relating to important co-morbidities. Being purely cancer-focused would exclude data that is relevant to treatment possibilities (for example cardiac risk exacerbated by androgen deprivation)
- Genomics
- Histopathology
- Narrative data contained in hospital letters, with advances in Natural Language Processing as a key enabler.
- Data sets derived from analysis of samples and tissues
- Data derived from Apps and wearables
- Socio-economic data, including data from social care. The data that is linked would still need to be agreed and available but could include pad use, catheter care, ability for self-care.

Opportunity: Building a platform to support patient centricity in research

Most data repositories have been built with limited patient involvement in their design, oversight and operations, and as a result the data collected may not reflect patient needs and preferences. Given that capturing PROs and remote data is reliant upon patient participation, it is problematic that in the majority of registries that exist, patients cannot access their own data and nor can they add to it. The failure to actively engage and involve with patients in relation to data capture is lamentable.

"Linking the real-world health and PRO data of prostate cancer patients is the key to unleashing the potential of prostate cancer data."

"Capture of genetic markers whether for reimbursement to label of a particular genetic mutation, or as a prognostic indicator (determining treatments and determining outcomes). Patient outcomes could be improved if we had a really good understanding of the genetic subtype landscape." Workshop participant

Despite the rising importance of PROMs, they do not routinely form part of standard data collection in clinical settings. Systematic collection of PROMs via patient registries is an achievable goal. Capturing dynamic data in real time has the potential to deliver improvements in efficiency, and ensures patient centricity in research. The longitudinal nature of the data means that patients act as their own comparators over time, describing evolution of disease and treatment response. Capturing non-coded data, for example narrative data in hospital letters, and data sets derived from tissues and samples are also data that is effectively excluded from currently linkable sources.

In addition, remote symptom monitoring through data collection from smartphones, has potential value. In a pilot study of adult patients with cancer receiving chemotherapy, daily steps were monitored by a smartphone accelerometer to detect possible treatment-related toxicities. A decline in steps was demonstrably linked to treatment related toxicities requiring intervention[1].

The value of linking long-term outcome data

Long term outcome data, for example that which could be used to better understand the value of radical treatments would have significant value.

There are important questions yet to be answered, for example, those relating to the implications of treatment choice and the effects of this choice on the efficacy of later pharmacological treatments.

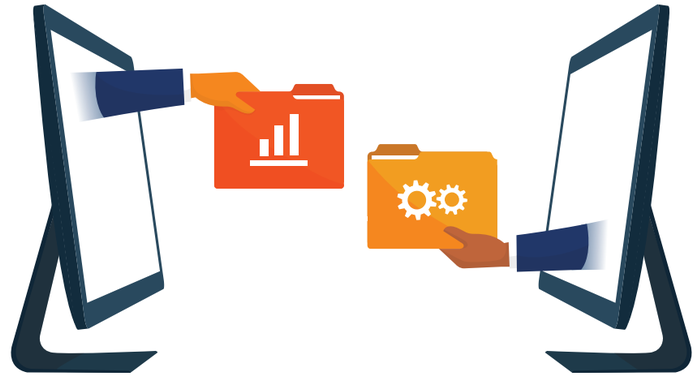
Linkage of longitudinal outcome imaging data would be of phenomenal benefit which could lead to improvements in diagnosis and enable early intervention of potential therapies.

"The ability to localise global data to the UK context would be ideal. A great deal of evidence is generated in the US and increasingly in Europe and Asia. The fundamental question is whether data collected in the US is adequate to meet the needs of UK patients. This could be addressed by a descriptive, real-world evidence program that could be provided for under the PCR model." Workshop participant

1. Purswani JM et al. 2019. Big Data from Small Devices: The Future of Smartphones in Oncology. Semin Radiat Oncol 29(4): 338-47
Soto-Perez-De-Celis E, et al. 2018. J Geriatr Oncol 9(2): 145-51

Opportunity: Fostering collaboration

There was strong agreement that collaborative working and agreement around a core data set were seen as the best route to delivering improvement in the data landscape, and ultimately in improving outcomes for patients.



Data collaboration between companies was flagged as relatively new territory. However, the potential for a prostate cancer data platform to collect data to speed the uptake of innovation was considered hugely valuable. In addition to the growing importance of real-world evidence in health technology assessments and the generation of information relating to clinical cost-effectiveness, we are moving rapidly towards a regulatory landscape where real-world evidence and randomised controlled trials are seen as complementary.

Examples of innovative areas of regulation included the MHRSA's Innovative Licensing and Access Pathway (ILAP). Providing access to real-world data and supporting the generation of real-world evidence would, therefore, be supportive of companies' efforts to develop innovative treatments and enhance patient care sooner.

- It is essential that patients are included in the steering groups and governance bodies of patient registries.
- Registries have value for research and for patients, enabling patients to benefit directly from the collection and use of their data, for example in terms of enabling them to have a better understanding of their disease and treatment.
- Patient organisations bring a patient-centric approach beyond functional outcomes to quality-of-life studies and can remove barriers to the collection of PRO data.
- It's imperative that any future prostate cancer data project is inclusive of different communities.
- Commercial organisations shared insights revealing differences in data needs dependent upon the nature of technology, the maturity and size of the company, commercial strategy, and development phase, commenting that it would be impossible to create a perfect universal solution.
- Augmenting data from NHS / CPRD / biobank with patient-reported outcome data would add a valuable dimension as this data is not available elsewhere.
- The multi-modal approach proposed by PCR would have value in terms of filling data gaps relating to patient journey, and in facilitating the delivery of personalised/ precision medicines.

Tissues and samples

We dedicated one of our workshops to discussions around access to tissues and samples and the value of linkage of samples to health data. The collection and storage of human samples is a significant and resource-intensive undertaking, fundamental to many aspects of effective research and innovation.

Our starting premise is that enabling access to tissues and samples linked to patient level clinical and molecular data would deliver benefits to innovators working to deliver targeted therapies for prostate cancer. Our starting observation is that there is a paucity of accessible sources of high quality and consistently processed and stored tissues/samples.

The practice of tissue banking is decentralised, subject to local governance, and from a technical point of view, different tissue banks employ different methods for the collection, processing, and storage of samples, and different rules and practices relating to their access and use. We have tested our working assumption that the potential to efficiently link real-world data (NHS and patient-reported) to tissues and samples would be a major advantage to the prostate cancer community, provided the challenges relating to quality, access, interoperability, and completeness can be addressed.

> **"If I have a new biomarker, and the patient has consented, I can go to all the different hospitals to request data but it takes a huge amount of time (months). It would be much quicker if there was a source from which I could upload the data to match the samples."**

Contributors to the tissue and samples workshop concluded that the current capacity of prostate cancer biobanks to link samples to clinical data is very limited. Those able to link digitally to clinical records and provide for updated data capture, are the exception rather than the rule, with the most prevalent approach being the de-novo and one-time capture of clinical data in the form of un-linkable spreadsheets.

A prostate cancer data platform might provide for consistency of data fields, formatting and storage of clinical data collected alongside tissues and samples. Tumour grade (how normal or abnormal tumour cells look under a microscope), PSA, BMI and dietary habits, lifestyle, ethnicity and socio-economic data were flagged as being important to capture.

Not all tissue and samples collections are easily discoverable. Some repositories are dedicated to specific prostate cancer studies and populations. Others are general repositories attached to pathology labs where prostate cancer samples represent a minority component of a larger cancer collection.

In neither instance is it possible to determine the nature, quality or extent of linked clinical data on the basis of desk research alone. The National Cancer Research Institute CM-Path estimates that around 50% of samples are non-discoverable and it is estimated that 85% of bio-banked samples are never used.

A consensus view was reached that the opportunity to provide for discoverability of collections, and to support linkage of longitudinal data to samples would be of value.

Recommendations

The overall recommendation was that there would be clear value in the establishment of a prostate cancer data platform. The following recommendations were made in relation to its design, delivery and function.

1. In order to derive the maximum benefit of investment, ensuring linkage capacity should be a priority, with an emphasis on collaboration rather than duplication. Working with the planned national NHS Secure Data Environment (SDE) and the developing NHS data ecosystem (and associated SDEs) should be prioritised, in order to focus resource to where challenges in case-use specific data access are likely to persist.
2. Patients should play a key role in governance and oversight of the platform, providing input into design and delivery.
3. A primary purpose of a prostate cancer data platform should be the enablement of translational and clinical research. In designing the platform, be mindful of future opportunities to support discoverability and access to tissues and samples linked to clinical and outcome data.
4. Taking into consideration the needs and preferences of stakeholders, a core minimum prostate cancer clinical data set should be developed, and an assessment made of existing data collections as source. Similarly, a core quality of life, socio-economic and lifestyle dataset for collection through a patient registry should be developed.

To review the notes from the workshops please visit **www.provenconnect.com** VISIT SITE

Next steps - A prostate cancer platform

Taking the key findings and recommendations contained in this report, we propose to develop and pilot a first-of-its-kind collaborative data platform. Core to this proposal is a patient registry that will:

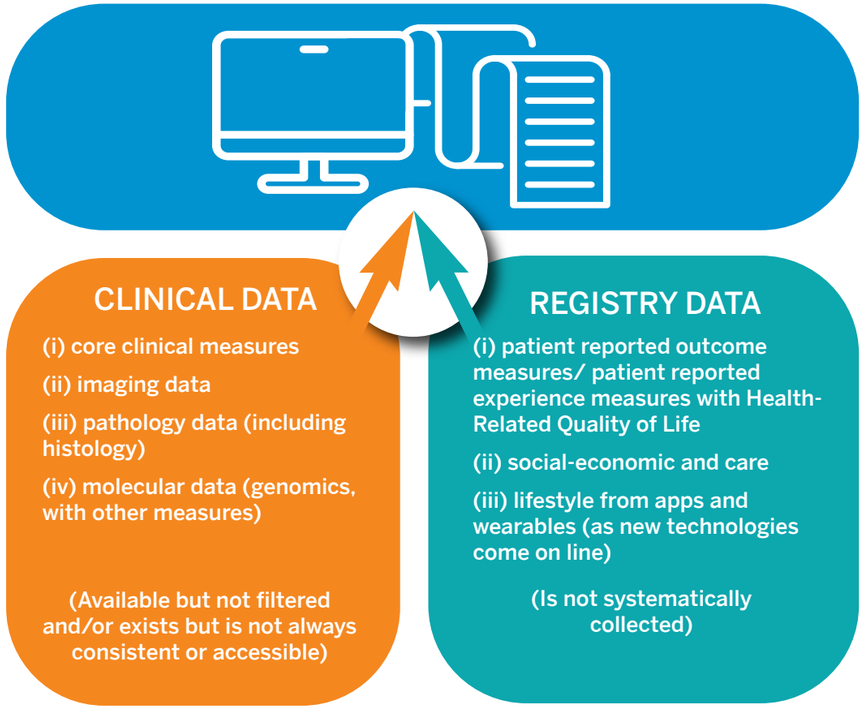
- Operate with direct involvement of persons with prostate cancer in all areas of design, content, and governance
- Deliver a detailed picture of patient perspectives on diagnosis, treatment, and wellbeing as determined to be of value to researchers and health providers
- Enable data donors to opt-in to recontact [in accordance with personal preferences]
- Enable patients to get a greater understanding of the impact of sharing their data over time, through feedback facilitated by PCR's patient engagement programme.

Patients that are signed up to the registry will be asked to provide broad consent to the access and use of their clinical data alongside data collected directly from them by means of questionnaires and quality of life instruments.

Linking patient reported data to clinical data will be achieved by:

- Working in collaboration with the NHS and in the context of the emerging structures for improving data collaboration
- Identifying the core minimum data set relevant to the proposed uses of the platform (principally translational and clinical research enablement).

Illustration shows the type of data a user would be able to access through the platform



Establishing a prostate cancer patient registry

Subject to the availability of resources, work to establish an interactive prostate cancer patient register could begin without delay to:

- Define the governance and oversight model for the registry
- Determine the nature of instruments and methodology for the collection of patient entered data
- Determine target conversion rate and onboard patient participants from PCR's Pledge Campaign (2,500 pledges)
- Commission and launch the platform.

Clinical data linkage - design and initiation phase study

Alongside establishment of the patient register we propose an 8 month design and initiation phase study that will include:

- Developing a detailed understanding of, and mapping, existing data sources, the data gaps and the mechanism for data linkage
  - Establishing the initial footprint of the platform in terms of clinical data linkage
- Identifying a core data set that will be valuable to multiple stakeholders
  - Assessing the long-term and wider data needs of stakeholders in order to ensure these can be built into the architecture of the platform
- Developing partner relations and identifying new potential in the following areas:
  - Health and care providers (critically NHSE)
  - Industry partners (some of which will be founding partners)
  - Data collaborators (including clinical trials units and active research teams)
  - Other partners (for example organisations with a focus on health data analytics)
- Identifying and engaging with key stakeholders to ensure the platform meets the needs of patients and end users
- Establishing a business model, including in relation to the development of intellectual property and service agreements in order to ensure the platform's sustainability whilst holding true to upholding the principals of privacy and equity
- Ensuring all protocols are in place including around governance and compliance
- Identify candidate hospital sites for initial pilot and confirm success criteria
- Develop a framework to gather future requirements and specifications for further developments, e.g., biobank linkage
- Engage with relevant steering committee members to document and agree key research topics / questions for the data platform
- Develop data specification for data platform which will include data element names and description, expected data recording format / coding standards and justification for requesting
- Coordinate with data platform development team to ensure agreed data specification is aligned with data platform technical capabilities.

Platform benefits

Short term

- Enable the delivery of patient centricity as a core principle of research, not only through the collection of PROMs and PREMs but also by facilitating engagement and involvement
- Provide value for patients in terms of capturing key information and monitoring their prostate cancer journey.
- Provision of some evidence in submissions for patient needs and in the context of health technology evaluations
- Enable industry to gain important patient perspectives and real-world experience from registry data and focus groups recruited via the registry.

Medium term

- Enable feasibility studies, and provide for enriched, speedier or more finely targeted recruitment of patients to clinical studies, and serve as a tool for improving diversity and inclusion in clinical research
- Support development of personalised treatment through improvements
- Provide a platform for nested socio-economic research studies
- Supporting decisions relating to reimbursement to label and as a prognostic indicator
- Capture significant volumes of potentially important data that are not yet recorded in electronic medical records, for example, ambulatory status, lifestyle, ethnicity, and socio-economic data.

Long term

- Fill data gaps relating to patient journey and facilite the delivery of personalised/ precision medicines
- Test global data, from examples generated in the US and increasingly in Europe and Asia, is generalisable to the UK context. Further, the platform could provide for comparisons and learning across different systems.

Final word

This report sets out a vision through which data could be harnessed to deliver faster, and patient-centric innovation.

Its purpose is to provide an overview of prostate cancer data and our findings are offered as a guide to future data initiatives in the prostate cancer space, to spark conversations and collaborations, and to build a community of interest.

Terminology

We have used the term ‘men’ throughout, but this should be considered to include not only men but also trans-women, non-binary people who were assigned male at birth, and some intersex people as they are also affected by prostate cancer.

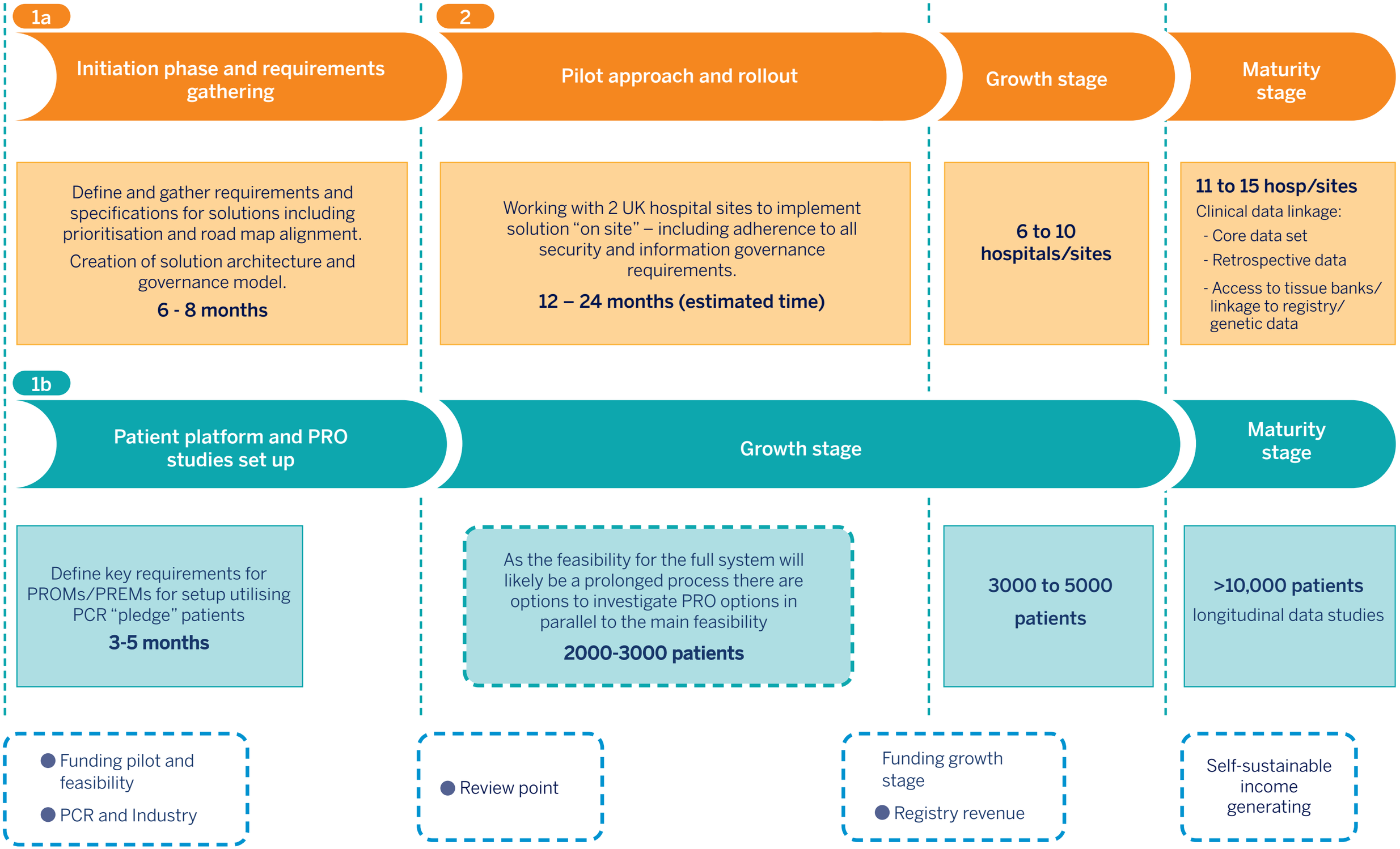
We use commercial researchers as being all researchers working outside of academic institutions. Service providers – clinicians. Patients – all people who have or have had prostate cancer.

- **Annex 1:** Proposed plan for the development of a prostate cancer platform
- **Annex 2:** NHS data silos
- **Annex 3:** Prostate specific data collections
- **Annex 4:** Discoverable prostate tissue and sample collections in the UK

To join the conversation please contact: **Sonja.Lawrence@provenconnect.com**

EMAIL

Annex 1 - Proposed plan for the development of a prostate cancer platform



Annex 2 – NHS health data silos

Type of Data	Collection / Set	Nature of Routine Clinical Data	Coverage	Access
Primary Care	**SUMMARY CARE RECORD (SCR)**	Basic information about allergies and medication prescribed by GP and elsewhere. May contain additional information on long-term health conditions subject to patient consent.	98% of GP practices in England.	Can be viewed by health and care staff and community pharmacists.
	GPDPR (General Practice Data for Planning and Research)	The GPDPR dataset contains coded data on symptoms, diagnoses, test results, allergies, referrals, vaccination and prescriptions. It does not contain any written notes (free-text) or any letters or scans.	Proposed by NHS Digital to cover England. Predicated on SCR. This service is not yet operational.	Patients can request access to their own SCR. Intended to support and enable research in addition to service planning and development of policy. This initiative has been delayed. Data when available will only be permissible through TREs.
	CPRD (Clinical Practice Research Datalink)	Rich source of data that includes data relating to primary diagnosis, hormone sensitivity, staging and all-cause mortality. It is possible to derive from CPRD the longitudinal data on clinical staging alongside data entered in a primary care setting in near real time. Data relating to imaging (though not the images) can be accessed via CPRD Gold.	Large electronic dataset of 60 million, (active for 16m patients) primary care medical records that is jointly sponsored by NIHR and MHRA (DHSC is the data controller).	CRPD operates on a cost recovery model (£339K multi-study license fee per annum).
	SPIRE (Scottish Primary Care Information Resource)	SPIRE provides a platform for practices to see information about their patients, through reports on topics such as: • Practice activity • Vaccination uptake • Multimorbidity	Information not available. Incomplete. Oversight and delivery of SPIRE sits with Public Health Scotland.	Currently being rolled out to GP practices across Scotland.
	WLGP (Welsh Longitudinal General Practice	Individual-level health data including read codes. attendance and clinical information for all general practice interactions: includes patients' symptoms, investigations, diagnoses, prescribed medication and referrals to tertiary care. Test results are electronically transferred from secondary care systems. Each individual clinician can record information in their own way. The majority use Read Code Terminology.	This dataset covers 83% of the population of Wales and 80% of GP practices in Wales. It is linkable with anonymised fields for individuals and GPs to other datasets, including bespoke project specific cohorts.	Data provision is free from SAIL. Project costing depends on the number of people that require access to the SAIL Gateway, the activities that SAIL needs to complete data refreshes, analytical work required, disclosure control process, and special case technological requirements.

Annex 2 – NHS health data silos

Type of Data	Collection / Set	Nature of Routine Clinical Data	Coverage	Access
Secondary Care	**Northern Ireland GP Data** **Hospital Episodes Statistics Data (HES)**	Nil available Relevant to prostate cancer: for prostate cancer records (ICD-10 C61) finished consultant episodes (FCEs), hospital admissions, bed days, length of stay and lists procedures, linking FCEs to procurement of drugs. HES data contains detailed procedure codes (for example robotic vs non-robotic prostatectomy).	Nil available England – for around 20 million patients admitted to hospital each year.	Policy in Northern Ireland precludes access. Via CPRD for linked data. Organisations and individuals wanting to use certain kinds of data need to show they meet strict data governance standards by completing NHS Digital's DARS application process.
Cancer Registration - collection, maintenance and management of data on every new diagnosis of cancer occurring in a population	**NCRAS**	Anonymysed data available to bone fide individuals and organisations who wish to use it for research purposes, providing patient confidentiality is not compromised and appropriate research ethics approvals have been granted where required. Options such as a secure safe haven space and a cancer data model containing synthetic or perturbed data are planned over the next year.	England	Unknown
	WCISU	Anonymysed data available to bone fide individuals and organisations who wish to use it for acceptable research, providing patient confidentiality is not compromised and appropriate research ethics approvals have been granted where required.	Wales	Unknown
	NICR	The Northern Ireland Cancer Registry is run under agreement between the Public Health Agency and Queen's University of Belfast and collects information on around 14,000 cases of cancer each year.	Northern Ireland	Unknown
	DataLoch	DataLoch in Scotland records around 55,000 cancer registrations annually. The data controller is the Scottish regional councils and SE SHBs.	Scotland. Collecting since 1958 and holding in excess of 1.8m	Whilst there are plans to provide for commercial access, at time of writing data access is not possible.

Annex 2 – NHS health data silos

Type of Data	Collection / Set	Nature of Routine Clinical Data	Coverage	Access
Cohort Data Sets	DATA-CAN	Oncology population and cohort datasets. Partnership of NHS organisations, patients, charities, academia and industry. Partners include Genomics England and IQVIA. Data controllers are hospitals and universities.	Working across all four nations.	Hosted by UCL Partners. Data is not commercially accessible, and any research predicated on the data must be delivered by NHS or university staff.
	The National Prostate Cancer Audit	Collects anonymised information from hospital records on treatments and outcomes. This data is analysed and compared to see if hospitals are following national clinical standards and identify where improvements can be made. Several important data sets, such as MRI, biopsy, and genomic data are not included in the audit.	England and Wales	Do they share? Or only report? They also have PROM/PREMs so worth asking?
	UK BioBank	Large-scale biomedical database and research resource, containing in-depth genetic and health information from half a million UK participants (age 40-69 years and including 229,000 men). The Biobank is a prospective cohort study and has measured a panel of biomarkers in blood samples from its entire cohort, including biomarkers that have not previously been studied in prostate cancer. Tumour characteristics, such as tumour state and Gleason grade are not currently available in the UK Biobank for participants who have gone on to a diagnosis for prostate cancer. Biobank includes WGS data for 200,000 participants.	Four nations	Biobank is globally accessible to approved academic and industry researchers.
	Genomics England	The 100K Genomes Project combines whole genome sequencing data with medical records from around 85,000 inidividuals. The data is not linked to routinely collected clinical data nor to data gathered in research. GEL can provide for patient recruitment via treating clinicians. The NHS Genomic Medicine Service (GMS) has provision considering applications for data access in the form of collaborative research proposals, but currently holds limited relevant data pending rollout of the service.	England	Available for commercial purposes, provided the company concerned is a member of the GEL Discovery Forum.
	Our Future Health	Aim is to recruit up to 5 million adult volunteers. People taking part will be asked to provide information about their health and lifestyles and a small sample of their blood. The information and samples will be linked to health records.	UK	Supported by UK Research and Innovation (UKRI). Our Future Health will accept access requests from academic and commercial researchers.

Annex 3 - Prostate specific data collections

The prostate specific collections that we have identified provide rich sources of specific data relating to prostate cancer. The only register to link patient entered data and clinical data is in Sweden. This provides an excellent example of the value of data linkage.

Collection	Data	Population coverage
The National Prostate Cancer Audit VISIT SITE	Anonymised information derived from hospital records on treatments and outcomes. This data is analysed and compared to see if hospitals are following national clinical standards and identify where improvements can be made. Several important data sets, such as MRI, biopsy, and genomic data are not included in the audit. **www.npca.org.uk**	England and Wales
Janssen Prostate Cancer Registry VISIT SITE	RWD patients with metastatic castration resistant prostate cancer. Prospective study of patients with metastatic castration resistant prostate cancer (mCRPC) in Europe. Its purpose is to provide real world data to help improve the quality of care for men with mCRPC. **bit.ly/3EAlM8o**	More than 2500 patients have been recruited from 192 centres in 16 European countries with follow up for 3 years.
IRONMAN Registry VISIT SITE	Patient medical history, treatment information and banks a blood sample. Eligibility: Metastatic Hormone Sensitive and CRPC. **ironmanregistry.org**	16 countries, 5,000 individuals. Australia, Brazil, Canada, Ireland, South Africa, Spain, Switzerland, Sweden, United Kingdom and U.S.A. Registrations 2017-2022
HEAT Registry VISIT SITE	Prospective registry of men treated with High Intensity Focused Ultrasound (for all UK patients since 2004). Data includes pre-treatment and post-treatment follow-up, each with c20 data points including PSA, age, surgeon, MRI, biopsy info, histopathology & Gleeson score. **bit.ly/3Mr2yEe**	Globally accessible, web-based platform for storing clinical data on patients (2000+) treated with HIFU, managed by Imperial College London.
PIONEER VISIT SITE	UK data onboarded to date is derived from CPRD and Diamond (control database) with Scottish HER onboarding scheduled for inclusion from July 2022. Demographics, Epidemiology, diagnostic, imaging, treatment, outcomes, quality of life, genomics. **prostate-pioneer.eu**	The consortium, which operates in 9 countries, was established in January 2018 and will run to December 2023. Again – will this be an accessible for all? Ask the team?
National Prostate Cancer Register NPCR VISIT SITE	Registry inclusive of socio-economic data, cancer characteristics, diagnostic work-up, and primary treatment for patients diagnosed with PC. The PCBaSe is a platform for clinical research established in 2008 by linking NPCR to other health-care registers and databases. **npcr.se**	Sweden Comprises more than 180,000 PC cases with data inclusive of socio-economic data. NPCR registers comprehensive data on cancer characteristics, diagnostic work-up, and primary treatment for patients diagnosed with PC in Sweden. Since 1998, NPCR includes information for 98% of all incident PC cases registered in the Swedish Cancer Registry, to which reporting is compulsory and mandated by law. A limitation of the Swedish data arises from the homogeneity of the population (i.e. predominantly Caucasian). An annual report presenting information at a department level is publicly available at www.npcr.se.
Prostate Cancer Data Base Sweden (PCBaSe) VISIT SITE	Platform for clinical research established in 2008 by linking NPCR to a number of other population-based health-care registers and demographic databases. The linkage was performed with the Swedish Cancer Registry, the Cause of Death Register, the Prescribed Drug Registry, the National Patient Registry, the Longitudinal Integration Database for Health Insurance and Labor Market Studies (LISA) and the Multi-Generation Registry. PCBaSe includes five prostate cancer-free control men for each PC case. The inclusion of controls allows for case-control studies as well as for cohort comparisons following cases and controls after date of diagnosis. PCBaSE does not current provide for collection, linkage and analysis of Patient Reported Outcomes Measures (PROMs). Van Hemelrijck M et al. Cohort profile: the national prostate cancer register of Sweden and prostate cancer data base Sweden 2.0. Int J Epidemiol. (2013) 42:956–67. doi: 10.1093/ije/dys068). **bit.ly/3essVgr**	Sweden. 80,000 cases.
MPower Prostate Cancer Registry VISIT SITE	The project collects data on trends in patient experiences by asking patients questions about their prostate cancer diagnosis, treatment and quality of life. The aim of the registry is to improve understanding of the experiences of men diagnosed with prostate cancer. The registry does provide for analysis of clinical data. **bit.ly/3TxfACD**	USA (Washington State) registry, hosted by the Fred Hutch Cancer Care Alliance.

Annex 4 - Discoverable prostate tissue and sample collections in the UK

UK Prostate Sample Collection Database | prostatedatabase.org.uk VISIT SITE

A database jointly established between ProMPT and the Southern Collaborative, both of which have built up extensive collections. UK researchers who would like their collections to be included are invited to contact the collaborative.

Tissue Directory & Coordinating Centre (TDCC-UKCRC) | biobankinguk.org VISIT SITE

The TDCC is a searchable data base that covers multiple diseases.

There is little overlap between the two initiatives and those collections which appear in neither are more difficult to discover. Of those collections registered with either initiative, the collection size and nature of the tissue within these repositories varies greatly.

Protocols for collection and storage of samples and associated clinical data, are repository specific.

Of those repositories sitting within the TDCC data base, those which accept tissues requests and who have declared them, the numbers of requests received and granted have been small (tens of requests).

	Collection (N° of patients)	Prostate database.org.uk	UKCRC Tissue Directory	Sample Types
1	ProTect (77223)	+	-	Blood, biopsy, TMA Core, TURP, urine, Whole Prostate- FFPE
2	UK Genetic Prostate Cancer Study (10589)	+	-	Blood
3	ProMPT (3174)	+	-	Blood, biopsy, TMA Core, TURP, Urine, Whole Prostate - Fresh, Whole Prostate - FFPE
4	Trans-CHHip (less than 3000)	-	+	Biopsy
5	REQUITE (less than 3000)	-	+	Whole blood, DNA
6	STAMPEDE (less than 3000)			Tissue (paraffin)
7	University of Bradford (less than 1000)	-	+	Plasma, tumour, urine
8	RAPPER (less than 1000)	-	+	DNA, whole blood
9	TransAtlantic Prostate Group (TAPG) series (800)	+		TMA Core
10	Wales Cancer Bank (662)	+	+	Biopsy, Blood, TURP, Whole Prostate - Fresh, Whole Prostate - FFPE, DNA, RNA, Plasma

Annex 4 - Discoverable prostate tissue and sample collections in the UK

	Collection (N° of patients)	Prostate database.org.uk	UKCRC Tissue Directory	Sample Types
11	Exeter Tissue Bank (569)	+	-	TURP
12	University of Southampton (less than 500)	-	+	Blood products, serum, tissue (paraffin)
13	Robert Lane Tissue Bank (less than 5000)	-	+	Tissue (frozen/paraffin)
14	LBIH (less than 500)	-	+	Tumour, blood, urine
15	Active Surveillance Trial (292)	+	-	Biopsy, TMA Core
16	NCRI Prostate Cancer Map Project (221)	+	-	Whole prostate - Fresh, Whole Prostate - FFPE
17	MRC RT01 radiotherapy trial (205)	+	-	TMA Core
18	TARRY Series (173)	+	-	TMA Core
9	ICGC Prostate Sequencing Project (155)	+	-	Whole Prostate - Fresh

CASE STUDY: The Wales Cancer Biobank

WALES CANCER BIOBANK
BIOFANC CANSER CYMRU
VISIT SITE

The Wales Cancer Biobank approaches patients in Wales with known or suspected cancer to ask them to consent to donate bio samples (including and beyond those usually collected as part of clinical care) and data for research. In addition to their large archive of samples and data, WCB offers a range of services including:

- Establishment of bespoke collections
- Storage of sample for external projects, such as clinical trials collections
- Tissue MicroArrays that enable hundreds of tissue samples to be arrange into a single block so that simultaneous analysis of molecular targets can be undertaken on a single glass slide (including an existing TMA for prostate cancer).

Whilst establishing a prostate cancer tissue bank to meet the needs of diverse research initiatives would be an extremely expensive and long-term undertaking, our preliminary discussions with stakeholders highlighted the potential for a prostate cancer data platform to deliver efficiencies and create opportunities through collaboration.

Acknowledgements

This project has been delivered with pro-bono support from IQVIA, a global provider of advanced analytics, technology solutions and clinical research services to the life sciences industry, with which Prostate Cancer Research have agreed a collaboration to enhance the UK life sciences sector's understanding of prostate cancer through the use of clinical and patient-reported data to enable accelerated medicines development and improved health outcomes for people with prostate cancer. The two organisations are seeking opportunities to bring patient voice, need and experience into clinical research to support patient directed end to end drug delivery.

With special thanks to all the community members who took part in our focus groups and so willingly gave their time, wisdom and expertise, with extra special thanks to Mark, Rory, Aidan, Martin, Dafydd and Henry for going above and beyond in their support for this project.

Name and Title		Position	Organisation
Barbara	Arone	VP Medical Affairs Real World Solutions	IQVIA
Sally	Baker	Strategic Partnership	Astellas
Rachael	Barber	Head of Strategic Partnerships	Cancer Research UK
Alexander	Bates	Senior Health Economics & Market Access Mgr.	Janssen
Fuen	Bellvis	VP Clinical Biomarkers and CDx	Cambridge Oncometrix
Prof Charlotte	Bevan	Professor of Cancer Biology	Imperial College
Andrea	Biondo	Director, Clinical & Translational Research	Astex Pharmaceuticals
Dariusz	Brenski	Chief Science Officer	Novartis
Dan	Brewer	Professor in Medical Bioinformatics & Genomics	University of East Anglia
Guenther	Brueggenwerth	Head of Clinical Services Imaging	IQVIA
James	Charnock	Integrated Real World Evidence & Solutions	IQVIA
Colin	Cooper	Professor of Cancer Genetics	University of East Anglia
Rebecca	Cosgriff	Head of Strategy, Value of Data	Centre for Improving Data Collaboration, NHSE
Clare	Delmar	Founder	Focal Therapy Clinic
Sergio	Diaz	Patient Advocacy Manager Europe	IQVIA
James	Donaldson-Briggs	Senior Strategic Partnerships Manager	Cancer Research UK
Tim	Dudderidge	Consultant urologist – prostate cancer specialist	NHS
Mike	Fisher	Head of Business Development	Oncimmune
Dr Jim	Fisher	Chief Executive	Physiomics
Anthony	Fix	Founder and CEO	Lucida Medical
Colette	Goldrick	Executive Director Strategy & Partnerships	ABPI
Joanna	Gorniak	Technical Development Lead	APIS Assay
Rosa	Greasley	Medical Science Liaison	Novartis
Sarah	Greenhalgh	Operations Director	GlycoScore Dx
Joanne	Hacket	Head of genomic and precision medicine	IQVIA
Prof Rakesh	Heer	Professor of Urology	Newcastle University
Sinead	Holmes	Oncology Medical Affairs	Pfizer
Andrew	Hughes	Chief Development Advisor	CellCentric
Bas	Hulsken	Chief Technology Officer	Quibim
Gerald	Hysenaj	Research Manager	G;yco

Acknowledgements

Name and Title		Position	Organisation
Sarah	Johnson	Head of Patient Advocacy	IQVIA
Varsah	Khodiyar	Data & Connectivity Project Manager	HDR-UK
Christopher	King	Category Lead Real World Solutions	IQVIA
Jurek	Kozyra	CEO, Co-founder	Nanovery
Jessica	Lee	Head of Patient Focused Partnerships	Medicines Development, Catapult
Tom	Lee	Senior Cancer Intelligence Manager	Cancer Research UK
Melissa	Lewis-Brown	Head of Research Data Strategy	Cancer Research UK
Madhu	Madhusudhan	Charity Sector Business Manager	LifeArc
Prof Malcolm	Mason	Founder	Wales Cancer Bank
Prof Iain	McNeish	Chair of Oncology	Imperial College
Grace	Melvin	Senior Policy Officer	AMRC
Rod	Middleton	Project Manager	Swansea University SAIL
Lynn	Morrice	Director of Operations, BHF Data Science Centre	HDR-UK
Ian	Newington	Head of Special Projects	NIHR
Andrew	Nolan	Head of Local Study Unit	IQVIA
Dr Alison	Parry-Jones	Operations Director	Wales Cancer Bank
James	Peach	CEO & Co-founder	Precision Strategy
Julian	Porter	Commercial Manager UK	AstraZeneca
Shonit	Punwani	Prof of Magnetic Resonance & Cancer Imaging & Radiologist	UCLH
Tim	Rebbeck	Associate Director for Cancer Equity & Engagement	Harvard University
Jon	Reed	Co-founder	Future Perfect
Slaney	Ronan	Business Analytics & Insights UK	Pfizer
Emily	Ruben	CEO & Co-Founder	Duchenne UK
Jelena	Sassmann	Medical Manager- Oncology	Ipsen
Imran	Sid	Medical Advisor	Bayer
Alexandra	Smith	Sr Consultant genomic and precision medicine	IQVIA
Elliot	Smith	CEO & Founder	Maxwell Plus
Dmitry	Soloviev	Clinical Lead	Cambridge Oncometrix
Robin	Stern	Co-founder	Future Perfect
Daniel	Stevens	Vice President Medical & Clinical Development	Blue Earth Diagnostics
Andrew	Thompson	Strategic Development Lead	Jiva ai
Nenad	Tomašev	Staff Research Engineer	Google DeepMind
David	Tuch	Founder and Executive Chair	Lightpoint Medical
Priya	Viswanathan	Research Collaboration Manager	NIHR
Ian	Weatherhead	Head of Corporate Affairs	Ipsen
Tom	White	Chief Operating Officer	Maxwell Plus
Tina	Woods	Director	The All Party Parliamentary Group (APPG) for Longevity
Ashley	Woolmore	Initiative Lead, Collaboration for Oncology Data in Europe (CODE)	IQVIA
Report author	Dr Jayne Spink & Sonja Lawrence.		Design by DIGOnline.co.uk

Prostate Cancer Research

pcr.org.uk

Prostate Cancer Research
Suite 2, 23-24 Great James Street, London. WC1 3ES
Registered Charity No: 1156027

CALL 02037355444
EMAIL info@pcr.org.uk
VISIT SITE pcr.org.uk

prostatecancerresearchnews
PCR_News
prostatecancerresearch