Minutes of the meeting of the Confidentiality Advisory Group

02 February 2023 via Zoom

Present:

Name	Role
Dr Tony Calland MBE	CAG Chair
Dr Patrick Coyle	CAG Vice Chair
Mr Umar Sabat	CAG Member
Dr Sandra Duggan	CAG Member
Professor Lorna Fraser	CAG Member
Dr Katie Harron	CAG Member, Present only for application 3a.
Mr Anthony Kane	CAG Member
Professor Sara Randall	CAG Member
Ms Diana Robbins	CAG Member

Also in attendance:

Name	Position (or reason for attending)

Ms Katy Cassidy	HRA Confidentiality Advisor
Ms Emma Marshall	HRA Confidentiality Specialist
Mr Dayheem Sedighi	HRA Approvals Administrator
Ms Caroline Watchurst	HRA Confidentiality Advisor
Mr Andrew Houston	Applicant - Research Data Scientist (attended for discussion of item 3a only)
Mr Charles Gutteridge	Applicant - Chief Clinical Information Officer (attended for discussion of item 3a only)

1. Introduction, apologies and declarations of interest

Apologies – Professor William Bernal - CAG alternate Vice Chair, Ms Rose Payne, & Dr Malcolm Booth, (CAG Members) gave their apologies for the meeting.

There were no conflicts of interest declared.

2. Support decisions

Secretary of State for Health & Social Care Decisions

The Department of Health & Social Care senior civil servant on behalf of the Secretary of State for Health & Social Care agreed with the advice provided by the CAG in relation to the 08 December 2022 meeting applications.

Health Research Authority (HRA) Decisions

The Health Research Authority agreed with the advice provided by the CAG in relation to the 08 December 2022 meeting applications.

Minutes:

The minutes of the following meetings have been ratified and published on the website: December 2022 Sub-Committee minutes

3. New Applications

a. 23/CAG/0017- Application of the Clinithink Natural Language Processing tool and Machine Learning methods to Clinical Notes for the Screening of Lung Cancer

Context

Purpose of application

This application from Barts Health NHS Trust set out the purpose of medical research which aims to examine the feasibility of Clinithink, a clinical Natural Language Processing (NLP) tool, to extract SNOMED codes from GP and hospital notes of patients referred for chest x-rays, and to use the extracted data to identify risk factors for lung cancer and to develop diagnostic models for GPs for early-stage lung cancer, and prognostic models for lung cancer outcomes. The study is a retrospective observational cohort study, and uses an NLP tool to anonymise free-text notes and extract relevant clinical information. 's251' support is requested because the identification of eligible patients and the running of free-text notes through the NLP software is undertaken by 2 researchers who are not considered direct care team, and in undertaking these activities these individuals will be viewing confidential patient information.

Around 48,500 new cases of lung cancer are diagnosed each year in the UK. Death from lung cancer is the most common of all cancers, with only 10% of patients surviving for ten or more years. Most lung cancer is diagnosed in the later stages of the disease, making treatment more difficult and putting strain on the NHS. However, when diagnosed sooner, treatment is much more effective, with survival rates increasing to 30-80% from 2-15%. Therefore, this research aims to develop tools to identify patients at high risk of lung cancer before they otherwise would be in the clinical pathway. At present, the process for referral is heavily reliant on the referrer's personal judgment and ability to recognise symptoms, potentially delaying a diagnosis or misdiagnosing the cancer as a less serious ailment. Medical notes are rich sources of information, however, due to their unstructured nature, extraction of relevant information from free-text notes is a time and labour intensive process. For this reason, the use of unstructured data for the screening of many diseases is largely unexplored. Clinithink is an NLP tool that extracts requested information from medical notes, formatting the output into a structured dataset (SNOMED-codes) which can be more easily analysed. In implementing a tool such as Clinithink, the limitations associated with unstructured data could be mitigated. The intention of this study is to investigate the viability of using the Clinithink NLP tool to extract information from the medical notes of a targeted demographic of patients and test the viability of then applying newly developed

predictive models to the extracted information to identify which patients are risk of lung cancer, so that lung cancer can be detected, and patients can be treated earlier and more effectively, and lives are saved.

Two researchers employed by Barts Health NHS Trust, who are not considered direct care team, will identify eligible patients by searching the Trust records via the inclusion criteria, recording the relevant NHS numbers. The National Data Opt Out will then be applied. Following identification of participants, a 'third party' who is a member of the direct care team within Barts Health NHS Trust, who is not a member of the study team, will link the NHS numbers to secondary care data, and send a request to the Data Discovery Service (held at North East London ICB), in order to gain access to primary care free-text notes. The third party will then remove the NHS number and replace it with a pseudonymous 'PERSON_ID', and disclose the free text data back to the applicants, alongside pseudonymous structured data extracted from medical records. Barts Health NHS Trust data warehouse team will hold the pseudonymisation key. 's251' support is still required at this stage, as although the structured data is pseudonymised, the researchers still have access to the free text documents provided, and it is expected that these may contain confidential patient information.

Free-text information will then be fed into Clinithink's NLP tool, Clix, by the researchers. The tool will extract an anonymised dataset including biopsychosocial variables from the text, and will not extract sensitive and personal information which could be used to identify a patient. In order to verify that Clinithink has performed this task correctly, the direct care team will audit the data extracted from the free-text to ensure it is free from confidential patient information. This will involve cross-referencing a random sample of patient notes against the information extracted by Clinithink. Should identifiable information be extracted, the NLP queries will be amended and the extraction process re-run. Following a satisfactory audit, the free-text information will be deleted. The retained data will then be analysed and used to develop a series of diagnostic and prognostic models for lung cancer. Upon conclusion of the study, the generated research datasets will be stored in the Trust Corporate Records Centre for 5 years, in accordance with the sponsors archiving SOP and the UK Policy Framework for Health and Social Care Research.

A recommendation for class 1, 5 and 6 support was requested to cover access to the relevant unconsented activities as described in the application.

Confidential patient information requested

The following sets out a summary of the specified cohort, listed data sources and key identifiers. Where applicable, full datasets and data flows are provided in the application form and relevant supporting documentation as this letter represents only a summary of the full detail.

Cohort	all patients aged ≥ 40 years old, who have received a chest x-ray at any of the hospitals within Barts Health NHS Trust between 01 Jan 2016 – 31 Dec 2019 AND 01 Jan 2022 – 31 Dec 2022, with at least one primary or secondary care data entry after the initial chest x-ray Approximately 250000 individuals (5000 patients with a positive diagnosis for lung cancer and 245000 with no diagnosis of lung cancer)
Data sources	1. Barts Health NHS Trust; a. Secondary care data; i. Electronic medical records ii. The referral letter for the chest x-ray iii. Report from the chest x-ray iv. Free-text medical records 1-year prior to the chest x-ray v. Free-text pathology records for the subsequent 3 years and 364 days after the chest x-ray 2. North East London ICB (data processor Voror Health Technologies Ltd); a. Discovery Data service - Primary care data; (primary care facilities across Northeast London) i. Free-text medical records 1-year prior to the chest x-ray ii. The subsequent free-text GP record following the scan
Identifiers required for identification of eligibility purposes, and running through NLP	1. NHS Number 2. Full Health Record (free text notes) 3. Date of death

Identifiers required for analysis purposes	Analysis will be on an effectively anonymised dataset – 1. sex 2. ethnicity 3. occupation 4. post code will be modified to Indices of social deprivation 5. Date of death will be modified to time from diagnosis to death

Confidentiality Advisory Group advice

The following sets out the Confidentiality Advisory Group advice which formed the basis of the decision by the Health Research Authority.

Public interest

The CAG noted that this activity fell within the definition of medical research and was therefore assured that the application described an appropriate medical purpose within the remit of the section 251 of the NHS Act 2006.

The CAG suggested to the applicant that social deprivation would also make a strong characteristic to explore to see if there was a bias present there. The applicant responded that they would consider that in future. The CAG was satisfied with the response.

Practicable alternatives

Members considered whether a practicable alternative to the disclosure of confidential patient information without consent existed in accordance with Section 251 (4) of the NHS Act 2006, taking into account the cost and technology available.

- **Minimising flows of identifiable information**

The CAG asked the applicant to explain how they estimated the sample size. The applicant explained that they were looking at patient results within 5 years excluding the Covid years. They were looking at the number of chest x-rays that took place in that range and from that they estimated 250,000 patients for the sample size.

The CAG asked the applicant to explain whether they needed that entire population to be able to achieve the aims of the study because this seemed a lot of individuals data that would be processed without consent. The applicant explained that the more data used in machine learning studies results in more accurate outputs, especially with regards to underrepresented groups. In order to create the most accurate diagnostic tool, and to prevent applying different screening criteria to different subgroups of patients, this amount of patient data is required in order to fulfil the study aims.

- **Feasibility of consent**

The applicants reason that the estimated number of patients that will meet the inclusion criteria is approximately 250,000, and therefore, it would not be feasible to contact each patient individually and would be a considerable time and cost burden to the Trust. Additionally, given lung cancer is the condition of interest, associated mortality rates suggest that many of the patients that were diagnosed with lung cancer, particularly those earlier in the observation cohort, may have died.

The CAG was content that consent was not a practicable alternative.

- **Use of anonymised/pseudonymised data**

Confidential patient information is required for the identification of the correct patient, and to undertake linkage between clinical data from both primary and secondary care. The applicants are requesting 's251' support in order to remove identifiable information. This is undertaken at the earliest opportunity. It would not be practicable to undertake these processes without the use of identifiers.

The CAG was content that using anonymous information was not a practicable alternative.

- **Auditing the Tool**

The CAG noted that the direct care team would audit a proportion of the outputs of the NLP tool, to confirm it was appropriately anonymising the dataset, and that the SNOMED code outputs were correctly capturing the free text notes. It was not clear to CAG how many patient outputs would be audited in order to ensure this, and so asked the applicant to clarify.

The applicant responded that between 50 and 100 patient outputs would be audited by the direct care team, and that if there were any inaccuracies or confidential patient information, then the NLP tool would be adjusted and repeated, and a further selection of patients would be audited, until they were satisfied with the outcome. This is a technique that the applicant has used before in previous research. The CAG was satisfied with this response.

'Patient Notification' and mechanism for managing dissent

It is part of the CAG responsibility to support public confidence and transparency in the appropriate sharing and use of confidential patient information. Access to patient information without consent is a privilege and it is a general principle of support for reasonable measures to be taken to inform the relevant population of the activity and to provide a right to object and mechanism to respect that objection, where appropriate. This is known as 'patient notification'. This is separate to the local obligation to comply with the principles of the General Data Protection Regulation and Data Protection Act 2018.

A patient notification document has been provided, that will be displayed on the Barts Life Science website, and social media channels, and disseminated across all local health watches within the Northeast London ICB. The study will also be advertised via continued attendance and presentations to both the Patient and Carers Voice group, the over 50s Friday forum and attendance to the Hackney and City Neighbourhood Forums, for example the Hackney CVS Coffee and Networking Morning.

The National Data Opt Out will be respected, and a study specific opt out option is available.

The CAG asked the applicant to explain how they were going to manage the local opt out in the study. The applicant responded that they had provided an email address in the Notification document. The applicant explained that they specified a subject for that email called ‘opt out lung cancer’ where the research team would set a process where any specific emails that came in with that subject line would automatically be forwarded to the Barts Health NHS Trust data warehouse team. The Barts Health NHS Trust data warehouse team would be able to remove study identifiers from that patient within the health record and remove them from the study. The CAG was satisfied with the applicant’s response and requested that they also add a telephone number for local opt out.

The CAG noted that a link was provided for participants to opt out via the National Data Opt Out (NDOO), and advice was provided regarding opting out of all research. The CAG noted that whilst it was important for the study to respect the NDOO, and this needs to be made clear on the patient notification documents (to avoid patients opting out via the local mechanism if they have already registered an NDOO), it is not necessary to guide patients to opt out via the national data opt out for all research, if they only wish to opt out of this specific study. The CAG therefore requested that the study specific opt out option was made more prominent, and the paragraph on the National Data Opt Out is altered to merely state that the study will respect any registered NDOOs, in order to avoid patients inadvertently opting out of all research use of their data, whilst attempting to register an opt out from this activity only.

It was felt that the notification on websites alone were not enough, and that the applicant should develop a layered approach by creating in addition, a poster for relevant clinical areas, that had a link or QR code to the detailed website information.

The CAG noted that there was typographical error in the Notification document “no aspect of you care” and requested to be amended to ‘no aspect of your care’.

The CAG noted a slight error in the description of CAG and requested that the applicant amend the notification documents for accuracy with regards to removing the terminology ‘CAG approval’. As CAG is not the decision maker, it is more accurate to state that the application has been supported by the Health Research Authority (HRA) on advice from the Confidentiality Advisory Group (CAG).

Patient and Public Involvement and Engagement

Meaningful engagement with patients, service users and the public is considered to be an important factor for the CAG in terms of contributing to public interest considerations as to whether the unconsented activity should go ahead.

Patient and Public involvement groups were consulted throughout the project. This included;

1. A focus group with 15 individuals from the Northeast London Cancer Alliance Patient and Carers Voice group, comprised of past and present patients and carers of patients with cancer.
2. A presentation to 13 individuals from the Northeast London Age UK team at the Age UK Over 50s ‘Friday Forum’, comprising of members of the public around Bow, who were over 50.
3. 27 individuals from the Hackney CVS Networking Morning (a Neighbourhood Forum Event).

Study information was also disseminated to the local health watches of Hackney, Barking and Dagenham, and City, and two GP cancer leads for the NE London Cancer alliance. No response was received from the local health watches. A meeting was held to discuss the research plan with the GP cancer lead, which helped to shape aspects of the eligibility criteria and data sourcing.

The use of confidential patient information without consent was discussed. Further information can be sought from the advice form responses. Feedback was obtained following the presentations and discussion with the various groups, which was largely positive. Although patients raised the issue that the use of non-anonymised data was a concern, they were satisfied that there was the option to opt out from their data being used. Additionally, one member noted that “although using very in-depth information about a patient’s medical history is worrying, the potential benefit of diagnosing cancer outweighs that worry.” This appears to support the use of Confidential patient information without consent.

Future planned activities include the continued attendance to both the Patient and Carers Voice group and the over 50s forum at landmark stages of the project to gain further feedback and to disseminate findings and attendance to the Hackney and City Neighbourhood Forums. The CAG was content with the amount of Patient and Public Involvement undertaken.

Exit strategy

Applicants anticipate requiring 2 months to identify the cohort and obtain primary care free-text data via the Data Discovery Service. Therefore, the applicant requires S251 to support the flow of data from the Data Discovery Service for 2 months following CAG support and REC Favourable Opinion being in place.

Applicants anticipate requiring 3 months of 's251' support to apply NLP to the free-text data obtained in this study. This process will commence following cohort identification and the obtaining of all data. Following extraction of the relevant information from the free-text and after a successful audit for correctness and anonymity, the raw free-text data will be deleted.

The study-specific person_ID/pseudo ID is generated by Barts Health NHS Trust data warehouse team and the key stored for the duration of the data extraction stage of the study, after which the key is deleted.

Therefore, applicants anticipate requiring 's251' support for a total of 5 months following CAG/REC applications being supported.

The CAG noted that they were unclear whether Barts Health NHS Trust data warehouse team was considered to be the direct care team and requested clarification on whether s251' support was required for the retention of the key by Barts Health NHS Trust data warehouse team.

Confidentiality Advisory Group advice conclusion

The CAG agreed that there was a public interest in this activity, were supportive in principle of this activity proceeding, and therefore recommended to the Health Research Authority that the activity be provisionally supported. However, further information and actions would be required prior to confirming that the minimum criteria and established principles of support have been adequately addressed.

In order to complete the processing of this application, please respond to the request for further information, and actions required to meet the specific conditions of support where indicated, within one month.

Request for further information

1. The patient information materials need to be revised as follows;

    a) Please add telephone contacts for local data opt-out.

b) Please make the study specific opt out option more prominent, and alter the paragraph on the National Data Opt Out to merely state that the study will respect any registered National Data Opt Outs.

c) Please amend the typographical error in the Participant Information Sheet 'no aspect of you care'.

d) Further methods of patient notification also need to be developed, adopting a layered approach, making information available in brief accessible leaflets as well as online. Please provide posters in clinical areas including QR codes or links leading to further information on the website.

e) Please change the wording where mentioned 'CAG Approval' as CAG is not the decision maker, it is more accurate to state that the application has been supported by the Health Research Authority (HRA) on advice from the Confidentiality Advisory Group (CAG).

2. The CAG noted that they were unclear whether Barts Health NHS Trust data warehouse team was considered to be direct clinical team. Please confirm whether CAG support is required for the retention of the key by Barts Health NHS Trust data warehouse team.

3. Please provide the Favourable opinion from the REC when available

Specific conditions of support (provisional)

The following sets out the provisional specific conditions of support. These may change in the final outcome letter depending on the responses to queries.

1. Favourable opinion from a Research Ethics Committee. **Pending**

2. Confirmation provided from the DSPT Team at NHS England to the CAG that the relevant Data Security and Protection Toolkit (DSPT) submission(s) has achieved the 'Standards Met' threshold. See section below titled 'security assurance requirements' for further information. **Confirmed**:

   The NHS England **21/22** DSPT review for **Barts Health NHS Trust, Voror Health Technologies Ltd** and **North East London ICB** were confirmed as '**Standards Met'** on the NHS England DSPT Tracker (03 February 2023)

b. 23/CAG/0019 - CLEOPATRAA Trial - Breast cancer molecular typing and grading trial

Context

Purpose of application

This application from Leeds Teaching Hospitals NHS Trust set out the purpose of medical research which aims to determine whether 4D Path's Technology Q-Plasia OncoReader Breast, that has been developed in the research setting, works robustly and reproducibly in the clinical diagnostic environment, to confirm its real-life clinical utility, in terms of breast carcinoma grading and molecular subtyping. Applicants will also establish the cost-effectiveness of the 4D Path platform. Applicants also wish to examine the long-term characteristics of the patient cohort, and the future potential applications of the technology.

Breast cancer is the most commonly diagnosed cancer worldwide, with approximately 55,900 new cases a year diagnosed in the UK. In order to determine the correct treatment for breast cancer patients, a biopsy is usually taken and reviewed by a pathologist, to obtain a definite diagnosis. The biopsy tumour tissue is also tested for oestrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) status – i.e. its 'molecular subtype'. The results from these tests inform oncologists in their choice of the most appropriate treatment for each patient. The results of these markers can also aid prognosis. These tests are usually interpreted by a pathologist. Carrying out of these tests requires extra work from pathology laboratories as well as the extra reporting time/expertise from the pathologist. With only 3% of the UK's current pathology departments fully staffed to meet current demand, there is a need for a tool to help provide accurate and efficient reporting of breast cancer biopsies.

It is hoped that this tool could;

- improve earlier diagnosis of breast cancer, including molecular subtype results, to provide treatment-critical information to oncologists sooner,
- reduce cost to NHS laboratories by negating the need to perform additional tests when digital images of breast cancer biopsies can be used instead,
- eliminate the time a pathologist needs to spend interpreting these additional tests, freeing up time for more complex diagnostic work,
- freeing up laboratory staff who would otherwise be performing these extra tests, allowing allocation of their time to other critical tasks,
- avoid destruction of biopsy tissue which is otherwise needed to carry out these additional tests, leaving tissue for further potential research or testing,

- improve the diagnostic accuracy of grade on biopsy specimens (which is incorrect relative to the resection/excision specimen in about 20% of cases): this is clinically relevant in instances where patient receive neoadjuvant therapy.

The trial will run alongside the normal pathology workflow at St James's University Hospital (SJUH), (part of Leeds Teaching Hospitals NHS Trust), and make use of existing digitised pathology workflows, since slides for all cases that are reported by the department are digitally scanned for routine diagnostics. Eligible cases will be identified by the direct care team within the pathology laboratory system CoPath. At this stage, the National data Opt Out will be applied, by the Trusts Data Access Committee. 's251' support is required as the research team, who are not considered direct care team, will extract the digital images, clinical and pathological data required from the Trust digital pathology server, Sectra, and electronic patient records, which will require the research team to view confidential patient information. The research team are required to undertake these processes in order to keep the direct care team blinded. Digitised images of malignant breast cancer biopsies will be effectively anonymised prior to disclosing to 4D Path. Once the 4D Path algorithm has produced a result, this will then be compared to the result produced by the reporting pathologist. The unblinding and comparison of the case diagnostic output comparison will be performed by the research team to avoid bias, which will also involve access to confidential patient information, and require 's251' support. The baseline clinical and pathological data will create a study database. This database will be retained by the direct care team only, so no 's251' support will be required for retention of the database. The database will contain LH number, (but no other identifying information), which the direct care team will use at 5 and 10 years to search the Trust clinical records for outcome data. Again, this will be outside the scope of 's251' support.

A recommendation for class 1 and 6 support was requested to cover access to the relevant unconsented activities as described in the application.

Confidential patient information requested

The following sets out a summary of the specified cohort, listed data sources and key identifiers. Where applicable, full datasets and data flows are provided in the application form and relevant supporting documentation as this letter represents only a summary of the full detail.

Cohort	All patients with a breast biopsy specimen processed at the pathology laboratory at St James's University Hospital, Leeds Teaching Hospitals Trust approximately 700-800 patients over 12 months. This will start once all appropriate approvals are in place.

Data sources	1. Leeds Teaching Hospitals NHS Trust: a. Sectra imaging software - imaging b. LTHT laboratory information management system CoPath – specimen reports c. electronic health record (PPM+)
Identifiers required for purposes of creating a pseudonymised data set for analysis	4. NHS number 5. LH specimen number 6. Electronic patient record will be viewed
Identifiers required for analysis purposes	1. Age 2. Gender
Additional information	Follow up at 5 and 10 years will be undertaken by the direct care team only, and therefore no ‘s251’ support required for this.

Confidentiality Advisory Group advice

The following sets out the Confidentiality Advisory Group advice which formed the basis of the decision by the Health Research Authority.

Public interest

The CAG noted that this activity fell within the definition of medical research and was therefore assured that the application described an appropriate medical purpose within the remit of the section 251 of the NHS Act 2006 and was in the public interest.

Scope

The Scope of the application was clarified as a response to CAT queries, however as this is quite different to what is in the application documents, the CAG requested that it is re-stated in this outcome letter, for clarity. Support is requested to allow the disclosure of confidential patient information to researchers who are not considered direct care team, at Leeds Teaching Hospitals NHS Trust, during the process of extracting and pseudonymising image data for disclosure to 4D Path Inc, and extracting a clinical/pathological dataset for analysis which will be retained by the direct care team, and also during the unblinding and comparison of the case diagnostic output.

No ’s251’ support would be required at all if the direct care team undertook the extraction and anonymisation of the image data. The applicant has confirmed this is not a practicable alternative as the study is meant to be blinded - as such, the members of the primary care team involved in image retrieval, eligibility checks and clinical metadata collection have to be distinct from those verifying the algorithm-based diagnoses returned by 4D Path. Although the process is simple and quick, making use of independent people (who are not part of the direct care team) is essential to prevent bias and not influence the clinical team. Whilst the researchers who will undertake this processing are clinical scientists with honorary NHS contracts, they are not direct members of the primary care team – hence the request for ‘s251’ support.

There are elements that do not require ‘s251’ support, and are out of scope for this application;

- The direct care team will undertake the identification of eligible patients, and therefore the identification does not require ‘s251’ support.

- Images shared with 4D Path are effectively anonymised and therefore this flow does not require ‘s251’ support.

- The retention of the study database does not require ‘s251’ support, as the CI confirmed that this was only accessible by the direct care team only.

- 5 and 10 year follow-up data for each patient will also be collected, using the LH number to access the medical record. No ’s251’ support required for this, as the CI confirmed that this was undertaken by the direct care team only.

The CAG requested an amended data flow diagram, which clearly outlines all the links and flows between data sources, clearly stating the common law legal basis for each flow so it is clear what ‘s251’ support is requested.

Practicable alternatives

Members considered whether a practicable alternative to the disclosure of confidential patient information without consent existed in accordance with Section 251 (4) of the NHS Act 2006, taking into account the cost and technology available.

- **Feasibility of consent**

The applicants reason that consent is not practicable for a number of reasons. Firstly there will be no intervention or change in a person's management brought on by this study, and therefore no opportunity for consent to be taken. In addition, the primary care team in this trial is based in histopathology, and it therefore does not have a patient-facing clinical role. As such, it has no reach across the range of environments in which biopsies are collected, both locally and regionally, and could not readily integrate within the clinical infrastructure/space available for the purposes of consenting the eligible cohort. It is also possible that attempting to gain consent from patients may erroneously bring confusion and concern that a computer-aided system may be influencing their treatment, or accessing their identifiable data when this is not the case. This feedback was provided after discussions with the PPI group, and first-hand experience from an ongoing study. The CAG was content that consent was not a practicable alternative.

- **Use of anonymised/pseudonymised data**

Confidential patient information is required in order to extract pseudonymised images and a pseudonymised dataset for analysis. The data is minimised at all possible timepoints. The CAG was content that using anonymous information was not a practicable alternative.

'Patient Notification' and mechanism for managing dissent

It is part of the CAG responsibility to support public confidence and transparency in the appropriate sharing and use of confidential patient information. Access to patient information without consent is a privilege and it is a general principle of support for reasonable measures to be taken to inform the relevant population of the activity and to provide a right to object and mechanism to respect that objection, where appropriate. This is known as 'patient notification'. This is separate to the local obligation to comply with the principles of the General Data Protection Regulation and Data Protection Act 2018.

There is a patient notification webpage document provided, which will be posted onto the LTHT research website. A local opt out is offered.

The CAG stated that the notification was too long and complex, making it challenging for patients to understand. The CAG therefore suggested that a layered approach was implemented, where simplified, easy-read versions of the material were available which then provided a link or QR code to a longer more detailed document, if people wished to read on.

The CAG noted that not all of the target population would have access to a computer and/or the internet and requested whether in clinical areas posters could be used to promote the study as well as online information.

The CAG noted a slight error in the description of CAG and requested that the applicant amend the notification documents for accuracy with regards to removing the terminology ‘CAG approval’. As CAG is not the decision maker, it is more accurate to state that the application has been supported by the Health Research Authority (HRA) on advice from the Confidentiality Advisory Group (CAG).

The CAG discussed the potential use of video or infographics as another method of communicating the study to patients, and this is provided as a comment only to the applicant, to consider if this might be possible.

Patient and Public Involvement and Engagement

Meaningful engagement with patients, service users and the public is considered to be an important factor for the CAG in terms of contributing to public interest considerations as to whether the unconsented activity should go ahead.

Leeds Teaching Hospital’s Cancer Patient and Public Involvement and Engagement (PPIE) have been involved in the development of this AI tool since conception and have been supportive throughout. The PPIE group were presented this specific project outline on 23rd March 2022, via an interactive session and PowerPoint presentation. The group was composed of 5 members with either personal or family experiences of cancer and are representative of the cohort. The trial design was discussed, specifically the use of data without consent. The group were very supportive and acknowledge the benefits that this technology could offer both patients and healthcare services and provided support for the use of data without consent.

4D Path also has ongoing involvement with the PPIE group, typically providing quarterly or six monthly updates, and also has ongoing patient advocacy support from two independent advocates on its advisory panel, one of which has reviewed the application and provided a letter of support.

The CAG noted that sufficient patient and public involvement to support the application has been undertaken with cancer patients in general. However the CAG requested that further ongoing patient and public involvement was undertaken specifically with breast cancer patient to ensure that patients who represent the specific cohort of the application are supportive of this use of confidential patient information without consent.

The CAG also requested that the wording of the online information and any other patient notification materials, such as posters, is reviewed by a relevant Patient and

Public Involvement group, and feedback from this is to be provided to the CAG within 3 months from the date of this letter.

Exit strategy

No ‘s251’ support will be required after the research team have undertaken the unblinding and comparison, as the CI confirmed database retained by direct care team. This will happen twice in the baseline section of the study: at the mid-point review, and at the end. Unblinding and comparison does not happen at 5/10 year follow up, only at baseline, and so ‘s251’ support will be required until 2024.

The CAG was content with the exit strategy.

Confidentiality Advisory Group advice conclusion
The CAG agreed that there was a public interest in this activity, were supportive in principle of this activity proceeding, and therefore recommended to the Health Research Authority that the activity be provisionally supported. However, further information and actions would be required prior to confirming that the minimum criteria and established principles of support have been adequately addressed.

In order to complete the processing of this application, please respond back to all of the request for further information, and actions required to meet the specific conditions of support where indicated, within one month.

Request for further information

1. The CAG requested an amended data flow diagram, which clearly outlines all the links and flows between data sources, clearly stating the common law legal basis for each flow so it is clear what ‘s251’ support is requested.

2. The patient notification materials need to be revised as follows:

    a) please ensure the wording and language of the patient notification documents is simple, and easy to read for a lay person.

    b) Please provide posters in clinical areas, including QR codes or links leading to further information on the website. Please ensure the opt-out options and contact details for local opt-out are clear.

    c) Please also add contact details next to information about local opt-out on the patient notification.

    d) Please change the wording where mentioned ‘CAG Approval’ as CAG is not the decision maker, it is more accurate to state that the

application has been supported by the Health Research Authority (HRA) on advice from the Confidentiality Advisory Group (CAG).

e) Please emphasise that the researchers will be looking at patient's documents and notes.

3. Please provide Favourable opinion from the REC when it is available

Specific conditions of support (provisional)

The following sets out the provisional specific conditions of support. These may change in the final outcome letter depending on the responses to queries.

1. Ongoing patient and public involvement needs to be undertaken specifically with breast cancer patients, to ensure public interest in the activity. Please also ensure that the wording of the online information and any other patient notification materials, such as posters, is reviewed by a relevant Patient and Public Involvement group, and feedback from this is to be provided to the CAG within 3 months from the date of this letter.

2. Favourable opinion from a Research Ethics Committee. **Pending**

3. Confirmation provided from the DSPT Team at NHS England to the CAG that the relevant Data Security and Protection Toolkit (DSPT) submission(s) has achieved the 'Standards Met' threshold. See section below titled 'security assurance requirements' for further information. **Confirmed:**

   The NHS England **21/22** DSPT review for **Leeds Teaching Hospitals NHS Trust** was confirmed as '**Standards Met'** on the NHS England DSPT Tracker (03 February 2023)

4. Annual Reviews

a. 21/CAG/0008 Clinical Practice Research Datalink Service (CPRD)

The Clinical Practice Research Datalink (CPRD) has been operating under Regulation 5 support since 2012 for NHS Digital to act as a trusted third party in order to link data from numerous sources. The CPRD is a function of the MHRA. The application sets out the activity to process a broad range of specified datasets by NHS England, and to enable de-identified disclosures to research applicants by the CPRD, following review through the Research Data Governance (RDG) process. Due to its national nature, the annual review is considered at full CAG meetings.

The CAG noted that, as of 01 February 2023, NHS Digital was merged into NHS England. All references to “NHS Digital” have now been revised to NHS England.

CPRD was previously supported under reference ECC 5-05 (a)/2012. In June 2020, the applicants were asked by the CAG to submit an updated application due to changes in the regulatory and legal landscape since 2012, some press interest, the age of the enterprise and the increase in its scale.

At a high level, support under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 has been provided for the following aspects:

- NHS England to receive identifiers, undertake linkages and provide the CPRD a de-identified dataset. While CPRD is not receiving identifiers directly, it is important to recognise that for the purpose of transparency, NHS England is processing identifiers on behalf of CPRD under this legal support.
- GP practices and specified others (according to the approved ‘Master Dataset’ list) to transfer confidential patient information to NHS England.
- NHS England is operating under the direction of the MHRA (via CPRD). The applicant for the purposes of this application is the CPRD who are responsible for the actions of NHS England (who in turn are operating under instruction of the CPRD).
- The CPRD do not receive identifiable data from NHS England or others under the terms of this support. Any processing by CPRD of confidential patient information must rely upon another legal basis.

Confidentiality Advisory Group Advice

Steps taken to anonymise the information or obtain consent from individuals:

Minimum use of identifiers to perform linkage

The applicants noted that NHS England acts as CPRD’s Trusted Third Party (TTP), to receive and process a defined and minimum number of personal identifiers, patients NHS number, date of birth, postcode and gender. These identifiers are securely and directly provided to NHS England by participating GP system providers and relevant third-party controllers and used by NHS England to link to the required datasets. NHS England undertake the linkage so that CPRD do not need to receive any identifiers. The identifiers used are the minimum required to ensure accurate linkage. The updated Master DataSet list was provided for review. The CAG raised no queries around the minimisation of identifiers.

Practicality of CPRD gaining consent from participants

Due to the large numbers of patents involved, CPRD could not feasibly contact all GP patients individually to obtain consent. The CAG raised no queries around the feasibility of seeking consent.

High-risk scenarios – when to refer studies to CAG

As described in the applicant's response letter in their 2021 resubmission and, following advice from the Confidentiality Advice Team in 2018, should a study protocol involving the use of linked data presents a potential 'high-risk' scenario CPRD will seek to mitigate the risk to acceptable levels first, before considering referral to CAG as an alternative. The applicants provided details on the internal escalation process within CPRD, which were unchanged from the details provided in the previous year's Annual Review submission.

The applicants advised that, for the period of February 2022 to December 2022, there were no 'high risk' cases where mitigations have not been possible and that would have needed referral to CAG under the CPRD S251 exemption.

The CAG noted the information on the CPRD escalation process and the information on the criteria for determining whether a study is high-risk, which had been provided in response to a condition of support of the previous annual review. Members agreed that the criteria were not well-defined. The CAG asked that the criteria were re-examined, and more precise criteria produced. The CAG asked that these criteria included a list of the conditions considered to be rare and an estimation of the prevalence of the condition (i.e. how many patients in a thousand have the condition). The characteristics of sensitive conditions needed to be provided. How "vulnerable demographic characteristics" are defined also needed to be provided.

The CAG noted that, when responding to the conditions placed on last year's Annual Review, CPRD had provided examples of studies that were potentially high risk. These included PEARL (which operates with support under CAG reference 16/CAG/0053) and a proposal for mental health related research. For the latter application, CPRD had suggested an approach that was lower risk, which had been accepted. The CAG found this information helpful and members queried whether any studies applying to use CPRD had gone through the CPRD escalation process, without advice being sought from the CAG, since this response was provided in March 2022. If yes, could an overview of the studies that had been flagged as potentially high-risk be provided.

The CAG noted the importance of lay representation during the escalation process. Members requested that details of any lay involvement in the process were provided. If no lay input is sought, members asked that this was considered.

Ongoing patient and public involvement

During 2022, the CPRD Primary Care Recruitment Team Primary Care Engagement Team have presented at practice PPG (Public Participating Group) meetings in Southwest Peninsula, Northwest Coast, North Thames and Wessex Clinical Research

Network. In November 2022, CPRD presented at a large research conference in Yorkshire and Humber. This was comprised of both medical professionals and patient representative groups, charities and other lay organisations and was attended by over 400 people. CPRD also conducted its annual PPIE workshop in June 2022. This was attended by 21 individuals representing a range of patients and patient user organisations including Alzheimer's research UK and cureparkinsons.org. The CPRD Research Data Governance Central Advisory Committee lay members were also in attendance.

The CPRD research bulletin, circulated to all CPRD contributing practices, will again remind practices to contact CPRD should they wish CPRD to present at their public engagement meetings. An MHRA wide public engagement event is also currently being discussed for 2023. It was not clear to the CAG the extent to which all these presentations were just presentations or whether they also elicited feedback from participants. It was also unclear whether specific issues of processing confidential patient information without consent were addressed.

The CAG noted that 21 people had been consulted at the PPIE workshop in June 2022. Information on around 40-50% of the population would potentially be included in CPRD. Members agreed that patient and public involvement needed to be undertaken with a much larger group. Feedback from this patient and public involvement would need to include the questions asked of participants and the responses given. A breakdown of the demographic characteristics of those involved also needed to be provided, to demonstrate that those consulted were sufficiently representative of the population included in CPRD. The CAG agreed that the applicants needed to provide a detailed plan on the further patient and public involvement would be conducted within 3 months of the issuing of this outcome letter. A report of the further activity conducted needed to be provided at the next annual review.

Members also noted that it was not clear how CPRD would ensure that those applying to use their data had undertaken sufficient patient and public involvement for their specific project. The CAG queried whether CPRD asked that applicants seeking to use CPRD data conducted patient and public involvement before approving requests to use the data.

Report against standard and specific conditions of support:

1. The following conditions need to be reported on within 3 months of the issuing of this outcome letter:

    a. The criteria for determining whether a study is high-risk need to be re-examined and more precise criteria produced and provided to the CAG.

b. Examples of studies that were potentially high-risk and went through the CPRD escalation process, without advice being sought from the CAG, were provided, so the CAG could see the type of activity that was being flagged as potentially high-risk.

c. Provide details of any lay involvement in the decisions making and escalation processes within CPRD. If no lay input is currently sought, this needs to be considered in future.

d. A detailed plan on the further patient and public involvement to be conducted needs to be provided.

e. Clarify whether CPRD request that applicants seeking to use CPRD data conduct patient and public involvement before granting access to CPRD data.

2. A report of the further patient and public involvement activity conducted needed to be provided at the next annual review.

5. Any other business

No other business was raised.

The Chair thanked Members for their attendance and the meeting was closed.

Signed – Chair		Date
Dr Tony Calland MBE, CAG Chair		*16 February 2023*
Signed – Confidentiality Advice Team		Date
Dayheem Sedighi, HRA Approvals Administrator		*06 February 2023*