Independent Review of CPRD Internal Data Governance Framework

1 PURPOSE, SCOPE AND METHODS

1.1 CONTEXT OF THE REVIEW

The Clinical Practice Research Datalink (CPRD) was established in 2012 by the Medicines and Healthcare products Agency (MHRA) and National Institute for Health Research (NIHR) as the UK-wide Government service supplying anonymised data and services for research to benefit public health. The origins of the CPRD database date back to 1988 when a group of local GP practices recognised the potential to study population health by sharing their electronic patient records. Many different organisations have managed the database over the years prior to its launch as CPRD.

From its inception 30 years ago when 57 GP practices encompassing 543,100 patients were involved, the dataset has been enhanced considerably by linkage to other datasets at individual patient level. The CPRD database now holds anonymised health records of 35 million patients. Over the past 3 decades, the external data governance environment has changed, and legal and regulatory frameworks have grown up around the database. CPRD has adapted to changes in data providers, computer technology, public attitudes and trust in governmental and public bureaucracies, and volumes of data that were probably hard to imagine at the beginning.

An Independent Scientific Advisory Committee (ISAC) was established in 2006 by the Secretary of State for Health, as a non-statutory advisory body to provide scientific advice to the MHRA on release of anonymised patient level data to support public health research. ISAC remains a central pillar of CPRD's internal data governance and its functions contribute to the assurance underpinning CPRD's regulatory approval from the Research Ethics Committee, for onwards sharing for research, of de-identified patient data collected from GP practices.

Since 2006, rising demand for CPRD data has driven the need for greater efficiency in the ISAC reviewing process, which has been balanced against a backdrop of increasing public scrutiny of data sharing and tighter external regulatory requirements. It is within this context that I was asked to review CPRD's internal data governance procedures and advise on processes that will enable CPRD to sustainably align with contemporary governance expectations and fulfil its important public health research mission.

1.2 TERMS OF REFERENCE

To undertake an independent external review of CPRD's internal data governance framework to ensure that it is fit for purpose in an evolving information governance environment by:
(i) assessing the requirements for proportionate oversight of CPRD data release for observational public health studies that will safeguard patient confidentiality and facilitate use of CPRD data for research; and
(ii) making recommendations for a fit for purpose CPRD data governance framework that will fulfil these requirements.

1.3 SCOPE

The scope of the review was to:

- Understand the context of the external governance landscape in which CPRD operates.
- Assess current approval processes for access to CPRD data encompassing the responsibilities and accountability of CPRD senior staff, the MHRA Executive and MHRA Board, CPRD Secretariat, CPRD researchers and the Independent Scientific Advisory Committee (ISAC).
- Seek feedback from CPRD staff and members of ISAC on current CPRD review processes.
- Gain insight from regular CPRD users to understand experiences, requirements and suggestions for possible improvements, including access to alternative research data sources.
- Consult with UK data custodians to understand governance processes for access to comparative research datasets.
- Evaluate stakeholder input, taking into account external regulatory requirements and public expectations.
- Produce a report for the MHRA Chief Executive of findings including recommendations for proportionate, fit for purpose governance structures and processes to enable data access

1.4 METHODS

- Desk review of relevant external governance requirements; documentation pertaining to application, review and approval processes to access CPRD data for observational research; and customer experience feedback from recent customer engagement exercises.
- Consultation with CPRD staff and interviews with or written submissions from 11 serving or past ISAC scientific and lay members, to gain insights into current processes and suggestions for future improvements.
- Consultation with 5 other data custodians who manage large datasets which are available to researchers, to understand comparative governance approaches and issues. Organisations consulted are listed in Annex 1.
- Information gathered was analysed and clustered into themes presented in this report. The recommendations in this report are based on my observations and issues recurrently raised by those consulted.

2 SUMMARY OF FINDINGS

2.1 ROLE OF ISAC IN RELATION TO MHRA RESPONSIBILITIES

Research proposals seeking access to data held by CPRD are subject to external expert review by ISAC, and prior to the establishment of ISAC in 2006, by its predecessor the MHRA Scientific Ethical Advisory Committee. ISAC's role is to review the scientific merit of proposals for research using CPRD data, including primary care data linked to other health-related datasets, by assessing scientific validity and feasibility of proposed studies, as well as potential ethical or confidentiality issues. ISAC terms of reference are found in Annex 2.

Although the role of ISAC hasn't changed, its structure and processes have varied considerably over time. This may account for some differences in understanding of the responsibilities of ISAC, CPRD staff and the MHRA Executive function. All other data custodians consulted have an approvals system in place, as well as systems of escalation from internal staff, through scientific advisors/ committee members/ and onward to various higher levels in the organisation, where it is clear that accountability for decisions rest.

My observations relating to the relationship between ISAC and the MHRA are:

- There appears to be a lack of clarity about responsibility and accountability for the consequences of a decision amongst some of those consulted.
- The confusion may be compounded by the words *Independent* and *Advisory* in ISAC's name, which suggests the Committee has more of a decision making function than a deliberative function.
- This ambiguity exists despite the fact MHRA is data custodian for CPRD data and manages the appeals process for applicants who disagree with an ISAC decision.
- Despite the lack of clarity on the distinction between the Committee and executive roles, the Chair, Committee and CPRD staff presently manage these tensions well to allow the process to fulfil its obligations.

2.2 PROTOCOL REVIEW AND APPROVALS PROCESS

In common with other data custodians, CPRD has struggled with the sheer volume of applications, efficiency demands, bottlenecks in the system and the desire for ISAC/approvals committees to review all research applications. Generally, a single committee jointly reviewing all applications has been found to be unworkable for most organisations, CPRD being no exception, so sub committees and delegated responsibilities have evolved.

The approach to systems for protocol review and approvals amongst data custodians varies, involving different degrees and types of screening, scientific appraisals and ethical assessment by internal staff, scientific advisors and committee members. The key question for most organisations is the degree to which the data being requested are proportionate and relevant to the research question being asked. Akin to other data custodians, CPRD staff carry out an initial screening and validation of research applications to ensure these have been correctly completed, which will result in iteration with the applicant if this is not the case. To ensure timely review of protocols, all reviews by CPRD staff and ISAC members are carried out 'virtually'.

Following the internal screening process, CPRD adopts a risk based approach to reviewing protocols. For the first stage, CPRD research staff conduct a risk

assessment to review feasibility, data governance and scientific rigor and then assign a low or high risk rating to the protocol. Risk ratings are then agreed by the ISAC Chair. Protocols judged to be low risk are reviewed internally by CPRD researchers and by the ISAC Chair. High risk protocols are reviewed by CPRD researchers, nominated ISAC members based on expertise and the ISAC Chair.

ISAC's role in peer review is to focus on the scientific question in the proposal to ensure that it will not damage the integrity of the data. In practice, a balance must be struck between harm and reputation on the one hand and facilitating research and encouraging innovation on the other. Internal CPRD staff reviewers and ISAC members comments are reviewed by the Chair before being sent to applicants. Any variation to an approved protocol, either minor or major, necessities protocol amendment approval, which varies in process according to the extent of the proposed change.

My observations on the current system are:

- A great deal of responsibility resides with the Chair. While this is not unusual in committee structures for authority to be vested in the chair, such reliance on the ISAC Chair to approve and edit feedback on all protocols represents a potential weakness in the system.
- The status of internal CPRD researchers engaging in the review process needs to be better defined. I was unclear whether the CPRD staff were considered as co-opted members of the Committee and therefore offering advice to CPRD, or if they were technical experts with special skills needed to help the Committee and the Chair reach a decision.
- CPRD research staff assign a risk rating at the start of the process which dictates how the protocol will be reviewed hence forth. However, the criteria for determining high and low risk do not appear to be clearly documented or universally understood.
- While the process of ISAC review is outlined in ISAC Annual reports, presently it doesn't appear elsewhere, so the process may not always be obvious to researchers.
- Case studies are discussed at ISAC face to face meetings which occur four times a year, however no regular calibration of internal and external review decisions takes place.

2.3 SCIENCE VERSUS GOVRNANCE AND THE JUDGEMENT PROCESS

Safeguarding patient confidentiality and considering the public benefit of the research is central to the decision making process. Making the data as easily available as possible to be used for research, is a priority and in many ways the *raison d'etre* for the existence of the dataset in the first place. To reach a judgement which considers both patient confidentiality and the need to facilitate public health research, involves scientific appraisal of the research application. For example, a judgement might need to take into account whether the research questions can actually be answered by the information in the dataset, and whether the proposed methods and techniques might compromise patient confidentiality, or the integrity of the dataset itself.

My observations on the scientific and governance judgement processes are:

- The current application form permits a degree of latitude in the level of methodological detail provided by applicants, which can make it difficult for reviewers to understand the kind of scientific appraisal that should be undertaken.

- There is a tendency for ISAC members to default to standard scientific and methodological peer review, which might be wasted effort as many applications have already been through a process of peer review from a funding body. Additionally, carrying out a full scientific peer-review, requires access to a different degree of methodological information than is currently captured in the application form.
- Presently it is assumed that everyone understands the concepts of science, methods and governance, but these have not been precisely defined within the context of the ISAC review process and in practice, there were different understandings of these terms.
- There was a general view that scientific appraisal of applications intrinsically involves consideration of governance issues. Whilst there is a clear link between scientific appraisal and data security, there is currently no separate distinction or deliberation of scientific method, feasibility and governance within the process.
- Currently the application form doesn't lend itself to focusing the scientific appraisal on issues which relate to governance, rather than just general methodological, statistical and epidemiological issues.
- All ISAC members undergo training at the outset of their tenure, however, there appears to be limited guidance or regular feedback available for external reviewers, with few opportunities for sharing collective wisdom to improve and standardise the quality of external Committee reviews.

2.4 COMMITTEE STRUCTURE

ISAC membership falls into two key categories; scientific and lay members. Scientific members provide advice on the medical, statistical/epidemiological and methodological aspects of protocols submitted to the Committee for review. Lay members provide advice on protocols seeking additional information from GPs and GP practices, and where there may be potential governance issues associated with a study. Studies requiring additional information from patients need separate ethics approval and are generally outside the scope of ISAC's remit. Lay members may be invited to comment on patient questionnaires which form part of a larger observational study within scope of ISAC's remit.

The structure of committee membership in other data custodian organisations varies. In some organisations there are specialist committees or groups dealing directly with patient benefit and who may not look at funding, study design or research questions. In several organisations data guardians figure strongly on committees and in approval processes, and various tests of public benefit are applied. The maintenance of personal anonymity is paramount for all the organisations. Some have special arrangements and place great emphasis on taking account of the legal frameworks surrounding data and data protection.

My observations on lay involvement in the Committee are:

- The lay members play a very positive role concerning issues relating to patient confidentiality and applying the patient benefit test.
- Lay members are treated as full and equal members of the Committee.
- Lay involvement in advising on public health benefit and public trust issues is important. Tasks such as reviewing whether lay summaries for all applications submitted meet Plain English standards, however, presents a significant workload and needs to be balanced against the practicality of meeting ISAC target Service Level Agreement response times to applicants.

Protocols have become increasingly complex in recent years especially those using Artificial Intelligence, Machine Learning and newer statistical techniques such as hypothesis generation. The emerging data lend themselves not only to hypothesis generation, but also to other forms of legitimate data exploration - to develop and test models for example. This is an important development. Its implication is that some features of the appraisal of applicants' protocols might require attention, particularly where hypotheses cannot be properly specified or identified *a priori.* Regardless of whether conventional or newer hypothesis generation approaches are applied, the role of the Committee is to ensure that the proposed methods used are the right ones to answer the research questions and whether the applicants are competent to execute the proposed method.

My observations on scientific expertise within the Committee are:

- Although the scientific expertise on the Committee has changed in recent times to align with the increasing complexity of protocols, the balance of expertise will remain an important consideration in the future as new techniques emerge and methodological approaches to data analysis change.
- The current form is geared towards conventional statistical approaches and may not capture the type of methodological detail needed for review of newer methodological approaches. This is especially pertinent when considering new methods and potential governance issues.

2.5 CONCLUSION

Across CPRD and ISAC and other large data custodians in the UK, there are features and problems which recur, which require solutions and which the organisations involved have dealt with in various ways. These issues are:

- maintaining the integrity of the data.
- how best to determine public interest and what will benefit public health.
- making the data available in a timely and straightforward way.
- how to reconcile conflicting demands for security and openness and scientific excellence and good governance.
- how to allay public anxieties about use and misuse of data and how to deal with mischievous and vexatious vested interests.
- legal frameworks within which internal governance take place.
- separation of powers between the deliberative, the executive and the advisory functions involved in the processes.

The working of ISAC and the systems underpinning it have adapted over the years due to changing customer demand, volumes of data, personnel involved and external governance circumstances. In the course of my review, I received a lot of very positive feedback about ISAC, about CPRD and the review process as a whole. Specific areas highlighted as working very well were the extensive interactions between CPRD staff and researchers, in particular providing help for new users; the improved website and published lay summaries; and the effort put into guidance for applicants on how to apply for data.

The CPRD database is a national treasure, with extremely rich data which attracts a range of users. It is run well by CPRD and ISAC is well organised with lots of CPRD staff support. My overall impression is that the current CPRD data governance process is working well. However, change is inevitable and there is a need to

consider redesigning and rebuilding some processes to address the observations that I have made in this review. Accordingly, I have made recommendations which I believe will meet the needs of the contemporary environment and will create a more resilient system of CPRD internal data governance for the future.

3 PROPOSED RECOMMENDATIONS

3.1 ROLE OF AN EXTERNAL COMMITTEE IN RELATION TO THE MHRA

3.1.1 The distinction between the advisory role of an external committee and the executive data governance responsibilities of the MHRA needs to be clarified. Accountability for all decisions relating to data release and governance resides with the MHRA as the legally recognised data controller. As such, the role of any committee comprising external experts can only ever be advisory. However, there should be an understanding that the MHRA will take on board the advice of an advisory committee, where it is not in conflict with any overarching data governance requirements.

3.1.2 In its capacity as data controller, the executive functions of the MHRA should include for instance:

- Deciding the process and providing guidance on criteria that should be used to triage and assess research applications as well as proformas to guide reviews.
- Determining matters of process like establishing committee terms of reference (TOR), appeals processes, auditing and monitoring of applications.
- Providing the infrastructure and staffing to support the delivery of the research governance function.
- Reporting on the research governance activities including metrics, publication of protocol summaries and maintaining data release registers.

3.1.3 The role of any external expert committee should be to provide expert peer review and advise on the merit of individual research applications.

3.2 PROTOCOL REVIEW AND APPROVALS PROCESS

3.2.1 The process of protocol review and approval should be redesigned and rebuilt. It is clearly for the MHRA to determine the final architecture, but these are my suggestions based on apparent pinch points, ambiguities and risks in the current process. The rebuilt system could consist of the stages below.

3.2.2 An optional *pre-application advisory review service* for applicants whereby CPRD staff advise on feasibility and any other issues that may need to be addressed to meet research governance requirements. These activities already take place, but without being formally acknowledged as an advisory service.

3.2.3 *Initial validation checks* by CPRD staff to ensure the application is complete, as is current practice, but also that questions relating to broader data governance, for example data security, are appropriately answered before the application is progressed. There should be a suitable Plain English summary outlining the proposed research aim and public health benefit that can be understood by a lay reviewer. Ensuring lay summaries are written in Plain English could be part of the initial validation process.

3.2.4 *Triage of protocols*: The current distinction between high risk and low risk protocols needs to be revisited. A more transparent and explicit triage process should be considered. The concern should not be so much about risk *per se*, rather it should be about the extent to which what is proposed is routine and "normal science" or something which is moving beyond the routine. The scope of what is 'routine' would need to be revisited at regular intervals to keep up with advances in scientific methods. Triage should be a separate stage in the process and should be conducted internally by CPRD staff, with oversight by a central advisory committee as part of the calibration process (see section 4).

3.2.5 The question of public benefit and patient confidentiality should be dealt with as a distinct part of the triage. If in doubt, the default position should be to err on the side of caution and require a referral for external review. The triage criteria should be revisited at regular intervals by the MHRA, in consultation with relevant experts as required.

3.2.6 *Protocol review*: All protocols should be reviewed by internal CPRD staff to assess feasibility of the proposed study using CPRD data, alignment with overarching research governance requirements and risks to patient confidentiality.

3.2.7 Protocols judged to be routine in the triage stage could receive a lighter touch review, for example by internal CPRD staff reviewers only.

3.2.8 Protocols that are highlighted as not routine in the triage process should undergo the following:

- Where the science is not considered routine, protocols should also be reviewed by external scientific experts with suitable topic expertise to assess the scientific merit in the context of public health benefit.
- Where there maybe potential confidentiality risks, protocols should be referred for specialist information governance review.
- Where there may be ambiguous public health benefit, protocols should be referred for review by a lay patient representative.

3.2.9 The process of protocol review will be overseen by MHRA appointed Chairs who will be external to the organisation. The Chairs will make a final decision on non-routine applications reviewed by external reviewers after considering all relevant reviews.

3.2.10 *Protocol amendments*: The definitions of minor and major amendments should be further clarified. In general, amendments should still be within the scope of the overarching aim as stated in the originally approved protocols and could be assessed and decided upon by internal CPRD staff reviewers.

3.2.11 The protocol review process should be published and transparent to external researchers and the public, including the role of internal CPRD staff and the appeals process.

3.2.12 External and internal review decisions should undergo regular calibration checks to ensure consistency.

3.3 SCIENCE VERSUS GOVERNANCE AND THE JUDGMENT PROCESS

3.3.1 There should be a separate and distinct appraisal of protocols for public benefit, scientific methods in relation to feasibility, and data security including issues germane to data governance.

3.3.2 Clear definitions of scientific method, feasibility and governance as applied to the review process need to be developed. Guidance should be available for reviewers to enable them to understand the type of scientific review required and its relationship to data governance, to distinguish the scientific assessment from standard scientific peer review, as for journal submissions or grant applications.

3.3.3 The research application form and the reviewer assessment form should both be redesigned to focus the proposal and scientific appraisal on issues related to governance, rather than standard methodological, statistical and epidemiological issues. The level of methodology detailed required in the application form should be reviewed to ensure reviewers have sufficient information on proposed methods, in particular to enable assessment of any arising governance issues.

3.3.4 Reviewer feedback and knowledge sharing should be built into the process to improve quality and consistency of reviews.

3.4 COMMITTEE STRUCTURE AND OPERATIONS

3.4.1 The committee structure should be reviewed, and the following options could be considered, taking into account requirements around rigour, public accountability, timeliness of response, scalability and resource.

3.4.2 A number of small ‘virtual’ topic specific expert review committees comprising relevant clinical and scientific members with data governance expertise, should be established to review non-routine applications for data access. Each committee would have its own external Chair and be supported by an internal CPRD staff reviewer to advise on feasibility of proposals and data governance aspects as described in section 2. The Chair will be responsible for the final recommendation based on the committee members’ reviews. Moderation of expert committee members decisions via a virtual meeting may be required where there are differences of opinion.

3.4.3 Where proposals are judged to be particularly unusual, innovative, or where public health benefit is equivocal, or governance and security of the data potentially put at risk, the expert review committees should have the option to draft in additional expertise on an *ad hoc* basis to help them.

3.4.4 There should be a pool of lay patient representatives who can be called upon to provide a patient/public perspective on applications where the public benefit is ambiguous. They should also be consulted on observational research protocols which require separate ethics approvals, such as patient questionnaires as part of the study. Similarly, there should be a pool of expert reviewers who can advise on newer methodologies like machine

learning or specialist clinical topic areas, who can be called upon by the expert review committees as required.

3.4.5 Oversight of the outputs of the expert review committees should be provided by a central advisory committee comprising the Chairs of the expert review committees, along with lay patient representatives. Additional scientific or governance experts may be included as necessary. The central advisory committee will convene at least annually for calibration and validation of decisions and to assist CPRD staff in developing feedback to expert committee members. The central advisory committee may meet on an *ad hoc* basis to adjudicate on difficult applications, as referred by CPRD staff or an expert review committee.

3.4.6 All reviewers, whether external or internal or lay patient representative must undertake training before they are assigned any reviews. Reviewers should also be part of a regular calibration exercise to ensure consistency of decisions and feedback across the various expert review committees.

Professor Michael P Kelly,

Primary Care Unit,

Department of Public Health and Primary Care,

Institute of Public Health,

University of Cambridge.

21st July 2019

ANNEX 1 DATA CUSTODIANS CONSULTED

SAIL databank

ISD Scotland

UK Biobank

NHS Digital

Public Health England

ANNEX 2 INDEPENDENT SCIENTIFIC ADVISORY COMMITTEE (ISAC) TERMS OF REFERENCE (TOR)

- Consider and provide advice to the MHRA on the feasibility, quality and public health value of research studies proposing use of anonymised patient level data from the CPRD.
- Provide timely and high-quality peer reviews on the scientific (medical, epidemiological, methodological) merit of research protocols proposing access and use of CPRD data.
- Highlight important ethical or confidentiality issues that may arise during access and/or use of CPRD data in research studies, taking into consideration input from the Confidentiality Advisory Group or research ethics committees.
- Advise on, and contribute to, the scientific content of guidance relating to the development of research protocols proposing access and use of data from CPRD.
- Review internal workings of the Committee to ensure consistency, efficiency and high standards of peer-review are maintained.
- Advise on other specific issues as requested.