

1. Overview

The China Kadoorie Biobank (CKB, formerly the Kadoorie Study of Chronic Disease in China) is a blood-based prospective study of 512,000 Chinese adults. The aims of the CKB are to study the relevance of lifestyle, environmental, biochemical and genetic factors for major chronic diseases (e.g. stroke, heart disease, cancer, diabetes) in adults, which will help to improve risk prediction, prevention and treatments of these diseases.

Participants were recruited between 2004 and 2008, they were interviewed, measured in various ways, gave 10 ml of blood (for long-term storage of DNA-containing buffy coat and plasma in Beijing and of plasma in Oxford), and gave permission for their subsequent health status to be monitored. Long-term follow-up, using each person's unique national ID number, continues via the government's nationwide Health Insurance System for any episodes of hospital admission, and via the Disease Surveillance Points (DSPs) system for cause-specific morbidity and mortality.

This large collaborative research project is being conducted jointly by the Nuffield Department of Population Health in the University of Oxford (NDPH), the Chinese Academy of Medical Sciences in Beijing (CAMS) and Peking University (PKU) in collaboration with the Chinese Academy of Medical Sciences in Beijing (CAMS) and Chinese government departments or institutes responsible for nationwide health insurance and the DSPs system. In 2001-02, when the study was originally being proposed by Oxford and discussed within the Chinese Ministry of Health and Ministry of Science and Technology (MOH and MOST). The Administrative Regulations of the People's Republic of China on Human Genetic Resources govern the use of collected data and samples since its implementation from July 2019.

Access to biological samples is limited by the small volume available and by the Chinese government's requirement for DNA to remain in China. Consideration will be given to collaborations that involve an extensive range of quality-controlled assays of all - or large numbers of - samples using high-throughput and cost-effective assay methods (eg, by low-cost genotyping or sequencing of DNA samples, by NMR for multiple analytes, or by methods for multiple antibodies). The CKB study group is actively seeking funding for assay strategies that will transform the available samples into accessible data for use by researchers in China, UK and elsewhere.

At baseline the participants were not asked specifically for consent to data sharing with outside bodies, consistent with standard practice at the time. Data sharing for research purposes with third parties was included more explicitly during the subsequent resurveys in 2008, 2013-14, and 2020-21 and the resurveyed participants raised no concerns. We understand that consent was sufficient to permit the supply of these data to *bona-fide* researchers of high scientific probity who have agreed to abide by the requirements described in this document and by any contractual arrangements with funders and external suppliers of the data relevant to the datasets.

Within the above constraints, CKB welcomes proposals for access to CKB data and stored samples from researchers in China and around the world, either for collaborative projects or for other forms of data and sample access to help achieve the study's aims. This document describes the CKB access policy and procedures. It has been developed in concordance with the general principles of data sharing promoted by various research organisations in the UK, China and elsewhere and the [Data Access and Sharing Policy of the Nuffield Department of Population Health](#), University of Oxford

2. Terminology

Data	Any CKB study dataset, including summary datasets, baseline survey data, re-survey data, follow-up information, blood assay results and genotyping data. Any additional datasets (eg meteorological, environmental) from project partners which CKB has stored in its central data repository.
Data Access System	The online tool which assists with the management of researcher registrations, applications for open access data and the delivery of approved datasets.
Data Access Agreement	Agreement covering the terms of data access to a Requestor of Open Access Data.
Collaboration Agreement	Agreement covering the terms of data access to a Requestor working with a member of the CKB study team in either China or the UK.
Open Access Data	Data being made available to external <i>bona fide</i> researcher through the Data Access System
Restricted Data	Data stored in the CKB data repository which has limitations placed on its use or wider distribution
Requestor	An individual or group of researchers seeking access to data and or samples from the CKB Study
Data User	An individual or group of researchers that has been granted access to data and or samples from the CKB Study

3. Principles of data sharing

As CKB has information on many different exposures and many different health outcomes over a period of many years, a wide range of investigators should be involved in determining which questions to address and how best to address them. As data custodian, the CKB study group must maintain the integrity of the database for future use and regulate data access. Data can be released outside the CKB research group only with appropriate security safeguards and approvals. The policy on data access is based on the need to:

- Protect participants and act within the scope of their signed consent;
- Ensure compliance with UK and China legal and regulatory requirements and prior conditions agreed with the Chinese Government (eg, the UK General Data Protection Regulation 2018 ; the UK Human Tissue Act, 2004; the China Data Protection Law, 2012; the Regulation on Administration of Chinese Human Genetic Resources, 2012 the Administrative Regulations of the People's Republic of China on Human Genetic Resources, 2019; the agreements with the Chinese Ministries of Health and of Science and Technology.)
- Ensure high quality research is fostered that will advance knowledge. Applications that include Chinese collaborators are particularly welcome since they would help to address the Chinese government's aim of developing and strengthening the research capacity of local investigators.
- Ensure that the data security is maintained.

3.1. Key components of this data access policy:

- **Open Access Data Availability:** Before data is approved for any analysis relevant members the CKB team responsible for generating the data must first undertake required cleaning, processing, quality control, integration and imputation. Following this, a period of exclusive use for the CKB researchers and key collaborators involved in collecting and/or generating the relevant data is anticipated (1-2 years depending on the scale and complexity of the datasets). Where additional data is generated as a result of a specific research award sufficient exclusive access will be reserved for the co-investigators to meet the funders expected research outputs. A further short period (eg 3-6 months) of prioritised open access for other external researchers in mainland China and Hong Kong will take place to ensure that the general scientific community in China and Hong Kong can benefit from developing research and analysis expertise before the datasets are made available worldwide through the data access system. At this stage the presumption is that all reasonable requests for data from *bona fide* researchers will be granted. Details of the currently available data and a timeline for future dataset releases are available on the Data Overview page of the CKB study website.
- **Collaborations:** The CKB research group will actively seek and respond to requests for scientific collaborations on specific projects, especially when framed in ways that will help strengthen Chinese research capacity. From time-to-time calls for specific project proposals or collaborations in areas of strategic importance and/or major scientific interest will be published. This model of facilitated collaboration with external researchers will be adopted where it can increase the value and quality of the data. Such collaboration will typically involve the adoption of new methods or assays and investigations into new health outcomes; they will be governed by a separate Collaboration Agreement. Collaborative Agreements will: (i) identify dedicated project lead from within the CKB study group in either China or the UK; (ii) detail arrangements for co-authorship or papers; (iii) cover intellectual property issues; (iv) detail financial commitments where appropriate.
- **Independent Oversight of Access:** An Independent Access Committee has been established to provide advice and governance on data access and sharing procedures (see Annex A for details of current membership). The Committee will monitor data sharing requests and decisions made. The Access Committee will also review any requests for access that raise particular issues (such as those relating to the use of samples or with complex ethical considerations). An additional Independent Data Access Oversight Committee within the Oxford University's Nuffield Department of Population Health provides governance advice on data sharing for CKB and other department-wide projects, which will liaise, if necessary, with similar committees in PKU about general and specific issues related to CKB data access. A Requestor can appeal to such committees if their request is denied and they disagree with a decision.
- **Protecting the Identity of Participants:** Safeguards will be maintained to ensure the anonymity and confidentiality of participants' data. Researchers will enter a legal agreement not to make any attempt to identify participants, and the data provided to researchers will not contain any personally identifiable variables (i.e. every data set provided will be "anonymised" with uniquely encrypted participant identifiers [PIDs]).
- **Data Security:** All CKB data is held on secure servers in central data repositories (with coordinated storage in Oxford and China) that is compliant with internationally recognized information governance standards. A data management team acts as gatekeepers and ensure that any shared data is delivered through a secure data delivery system and that any usage of restricted data held in the repository is handled appropriately.
- **Sample Preservation and Access:** 10 ml of blood was taken at baseline from each participant, and this was divided into 1 buffy coat sample and 1 plasma sample that are stored in Beijing and 2 plasma samples that are stored in Oxford. Samples stored in Oxford are being reformatted into multiple smaller sub-aliquots to facilitate future large-scale or cohort-wide multi-omics assays of plasma samples (eg proteins, metabolites and antibodies), Given the range of assays that can be undertaken and specific expertise

required to design, plan and manage the work, we would welcome proposals from the wider research community, including industry, to support large-scale sample assays, particularly those that are relevant for a wide range of different conditions and to undertake subsequent data processing, QC and initial analyses. Such a strategy would maximise the information available to researchers while minimizing sample depletion, and would also facilitate different comparisons since the assay methodology and quality control would be consistent across the whole cohort. Suggestions for particular assays to be included in these multiple-assay schedules will be welcomed, and all assay values will become part of the dataset, widening access to the scientific community.

- **Fees for data access:** Data is freely available to academic applicants from mainland China and Hong Kong. Researchers based elsewhere in the world will incur an Access Charge for each approved data request (currently £2,500 GBP). This will contribute to the administrative costs incurred in managing and reviewing the application, and in preparing the individual datasets. Collaborating researchers may also be required to cover the costs of administering the data sharing (including legal fees if applicable), retrieving, processing and sending the data or samples or costs associated with computational needs and data storage in support of analyses. Estimated costs for a particular request will be provided during the development of the project proposal.

4. Data access process

Potential Collaborators and data Requestors should first contact CKB investigators or review the CKB study website (www.ckbiobank.org) including the data access pages to gain an understanding of the available study data and projects that have previously been completed and are currently being undertaken.

4.1 Registration / Eligibility

All Requestors are required to register an account via the CKB website and complete a researcher registration form. Requestors should be employees of a recognized academic institution or health service organization with experience in medical research. They should be able to clearly demonstrate, through their peer reviewed publications in the area of interest, their ability to conduct independent research.

4.2 Collaboration requests

Approved researchers who are interested in collaborating on projects with CKB study researchers in China or the UK are encouraged to approach the CKB study group informally in the first instance by email ckbaccess@ndph.ox.ac.uk or to contact relevant CKB investigators to discuss research ideas and feasibility. Formal enquiries should include a project title and brief outline of the research project and the relevant data of interest. Proposed collaborations will typically involve the adoption of new methods or assays and investigations into new health outcomes or research fields. Each project requires a co-investigator from within the CKB study group who has a common interest in the project and relevant or complementary research expertise. Once identified, the Requestor/collaborator and the co-investigator will co-develop a research proposal which will then be reviewed by the CKB Steering and/or Research Management Committees.

4.3 Open Access Data requests:

Submission of a Data Request

Once the researcher registration is approved, data Requestors are able to login to the system and access the Data Requests section and the CKB Data Request Form. Details of data currently available on the open access platform and the information required for are provided on the form. Required information includes: project title and abstract; scientific rationale / methodology; anticipated outputs and project timeline. Additional questions cover ethical issues; collaborators / research team; funding support and data security.

Review of a data request

Open Access Data requests will be initially assessed by the CKB study team. Each application will be considered on individual merit. If necessary, independent peer review will be sought; disputed applications will be referred to the independent CKB Access Committee, which will oversee the process. The Access Committee will also review any requests for access that raise particular issues (such as those relating to the use of samples or with complex ethical considerations). Approved projects will: (i) have clearly defined objectives; include a sound methodology that is likely to generate meaningful results; (ii) be based on an appropriate and available selection of data; (iii) have clearly defined timelines and outputs (e.g. 1-2 papers in peer-reviewed journals).

To avoid duplication of effort, where there is substantial overlap between separate proposals submitted at the same time we may suggest that researchers collaborate on a project (after seeking appropriate permissions). The CKB will not insist on collaboration; if proposals meet the criteria for approval the same data may be shared with different institutions at the same time. Projects that overlap significantly with approved and completed projects as listed on the CKB website, may be rejected.

The CKB Team aim to review and respond to Data Requests within 4-6 weeks. A Requestor can appeal to Access Committee if their request is denied and they disagree with a decision.

5. Terms of data access

Once proposals are approved the following conditions and undertakings are required as conditions of access:

- **Data Access Agreement / Collaboration Agreement.** Before any data is transferred a signed transfer agreement must be in place between the Requestor's institution and either the University of Oxford or PKU. The choice on institution will depend on the country of origin of the Requestor and the institutional affiliation of the principal CKB study team co-investigator. A Template Data Access Agreement for the University of Oxford is included on the CKB Website. The agreement will include a copy of the approved project proposal as a Schedule.
- **Signing Authority.** Requestors should be acting as members of a recognised academic institution, research organisation or health organisation. Their request should come from a recognised email domain (eg, .ac.uk, .edu.cn). Their organisation should have formal policies and procedures to comply with any legal, ethical or data protection constraints and to ensure that the dataset is stored securely and used responsibly.
- **Ethics and Research Governance Approval.** Where applicable Ethics Committee approval for the research is the responsibility of the Requestor. The Requestor, in conjunction with study investigators, may also need to obtain approval from the Research Ethics Committees responsible for the CKB study. Local Research Governance approval and R&D approvals, if required, are the responsibility of the Requestor. Approval will need to be in place before any data is transferred.
- **Limitations on Use.** The data will be used for the purposes of medical research only and within the constraints of the consent under which the data were originally gathered, and of any contractual agreements between the CKB study and its funders or external data sources. Data supplied may only be transferred to Requestors named at the time of the original application or in subsequent applications and specified in the Access Agreement or later amendments. Data from the collection cannot be transferred to individuals outside the Requestor's research group without formal approval by the CKB Access Committee, and cannot be used for any direct commercial purposes.
- **Identifying Data.** The data provided to researchers will not contain any personally identifiable variables. Data sets will be "anonymised" with uniquely encrypted

participant identifiers (PIDs). The Access Agreement will contain confidentiality undertakings to further safeguard participants' privacy. Recipients must agree not to link the anonymised data provided with any other data set without permission. Recipients must not attempt to identify any individual from the data provided. Should recipients believe that they have inadvertently identified any individual, they must not record this, share the identification with any other person or attempt to contact the individual.

- **Intellectual Property.** All Intellectual Property Rights in the Data are and shall remain at all times the property of the University of Oxford and PKU. All Arising Intellectual Property shall vest in and be owned by the Requestors. The Requestors shall promptly disclose any such Arising IP in writing to the University or PKU. The University or PKU will be granted rights to use all Arising Intellectual Property for academic and research purposes, including research involving projects funded by third parties provided that those parties gain or claim no rights to such Arising Intellectual Property. This right shall be sub-licensable by the University to PKU and by PKU to the University.
- **Payment of Access Charges.** Data Requestors from institutions outside of mainland China and Hong Kong are expected to pay Access Charges to contribute to the administrative cost to the study of reviewing the application and preparing data for sharing, etc. Where these are applied, no Data will be provided to the Data Requestor until or unless the Access Charges are received in full.
- **Data Release and Delivery.** Once the proposal is approved and the Access Agreement signed, the data and its documentation will be generated in CSV (or any other pre-specified) format, encrypted and released in a secure manner via the Data Access System or by using encrypted physical media.
- **Publicity and Dissemination.** The CKB study team reserves the right to publish the title, the names(s) and affiliations(s) of the Chief Investigator(s), a lay summary and a scientific abstract of each piece of collaborative research for which access to the resource has been granted, before identification or publication of results. Requestors who do not wish details of their study to be openly available need to state this in their data request and give the reason. The Requestor shall not use the name or any trademark or logo of the University of Oxford or PKU in any press release or product advertising, or for any other commercial purpose, without prior written consent.
- **Authorship and Approvals.** Collaboration Agreements and Access Agreements will specify expectations regarding authorship and acknowledgements on research outputs. Collaborations require at least one co-author from the CKB study group. For Open Access Agreements no authorship from CKB team is required. The CKB should be acknowledged in accordance with the Access Agreement. Requestors are asked to submit proposed publications to either The University of Oxford or PKU for review not less than 30 days in advance of the submission for publication. CKB study staff may be willing to assist with drafting papers with appropriate acknowledgement. Approval from the CKB study team is not required prior to submission for publication.
- **Publications and Open Access.** All publications of the Results in a peer-reviewed journal, or as a scholarly monograph or book chapter, should be made available from PubMed Central and Europe PubMed Central by the official publication date and published under a Creative Commons attribution licence (CC-BY). All journal requirements for data release and deposition that are attached to publication should be complied with in full.
- **Integration of the Data.** After completion of work using released CKB data, the original dataset as well as any derived dataset and/or variables generated during the research must be returned to the CKB central data repository for archiving and/or merging with the main database for future use. If considered appropriate, the CKB staff may carry out independent checks and/or validation of the data and results to ensure the continued data integrity and reliability of the study findings.

- **Monitoring and Accountability.** The Data User shall be required to submit annual reports and any other information reasonably requested to evidence the work undertaken by the Data User in connection with the proposed project. If there is substantial delay or difficulty in completing the planned research, the CKB study team will have the right, after consultation with the Access Committee, to terminate the work if in its view there is little chance that the problem will be rectified. If there is substantial deviation or change in the planned use of the data, further approval will be needed.