QResearch

Protocol for an Ethical High Quality General Practice Database for Research

QResearch Protocol version 3.0, August 2015

Contents

1 Summary

QResearch is a high quality clinical database derived from UK general practices which started in 2002. The aim of QResearch is to develop and maintain a high quality database of general practice derived data linked to secondary care data for use in ethical medical research. The database is used for medical research into the causes of disease, history treatment and outcomes. In particular, it has been used for research into health inequalities; safety of medicines; the development of risk prediction tools to identify those at risk of a future disease, such as cardiovascular disease, (www.qrisk.org) or those at risk of a current diagnosis of a disease such as cancer (www.qcancer.org) .

This document is the protocol for QResearch. The QResearch website has additional information including publications arising from the QResearch database (www.qresearch.org) and application forms for researchers who wish to apply to use the data for research.

The special features of QResearch are:

- A nationally representative sample of approximately 1000 general practices contribute their data
- The data extracted contains no strong patient identifiers since the data are de-identified at source.
- Practices display a notice in their surgery waiting rooms informing patients that the practice contributes anonymised data for research.
- Patients are able to opt if they would prefer their pseudonymised data not to be included in the upload.
- The database is open to bone fide researchers employed by UK universities who must be able to publish whatever the findings.
- The researchers will be provided with samples of anonymised GP data that is required to answer their research question (not the whole dataset)
- The data from practices are assessed including its quality (in terms of completeness and accuracy).
- Following review by the QResearch Advisory Board, Trent Research Ethics Committee and the Ethics and Confidentiality Committee of the National Information Governance Board, the entire database has been linked to cause of death data, cancer and hospital data at individual patient level with linkages extending back as far as 1993. The linked data are only available at the University of Nottingham.
- Within the aggregated full database, a unique number assigned to each patient and to each practice to allow longitudinal tracking (pseudonymised). Researchers will only have access to fully anonymised data (without the pseudo-anonymised identifying code).
- Each use of the database is reviewed by the QResearch scientific committee
- The researchers are only allowed to use the data to answer their research question and must return to gain permission for use of the data for other purposes
- The researchers may not pass on patient data to anybody else

- The costs of use of the data will be carefully controlled to allow the scheme to be self-funding but to allow good access to bone fide researchers
- Tabular analyses are undertaken to demonstrate the accuracy and completeness of the data and made available for morbidity analyses
- Data quality checks will include comparison of birth rates, death rates and prescribing patterns with other aggregated sources

The key organisational features are:

- QResearch is be non-profit making
- Management decisions (pricing, use of funds) are be taken by a board representing the interests of EMIS and the University of Nottingham
- A QResearch Advisory Group sets policy and oversee the operating of the database
- A QResearch Scientific Board approves access to data

2 Details of the Proposal

2.1 Aim

The aim of QResearch is to develop and maintain a high quality database of general practice derived data linked to secondary care data for use in ethical medical research. It has created a new, high quality, primary care derived database that contains descriptive coded data on the health needs, risks, care and outcomes for a large population. It is one of the highest quality and largest such datasets in Europe. Research projects suitable for undertaking on the database are those which generate or test hypotheses and are intended for publication in peer reviewed academic journals. It also includes pilot work to determine sample size calculations and feasibility of specific analyses.

In developing and validating the database, we undertook a detailed survey of morbidity in general practice. The results of this were available in tabular (summary) format as in previous publications such as the 4th Morbidity Survey in General Practice or Key Health Statistics. Such information can be used to describe health need, care and outcomes.

2.2 Background:

While the introduction of IT into the clinical environment in hospitals has been slow and unsatisfactory, general practices have developed clinical systems that are widespread (99%) and well used (over 30% of practices only use their computer for recording their clinical records), and they are often highly accurate and complete. Researchers have begun to exploit the potential of the clinical databases in general practice. In Nottingham, Professor Julia Hippisley-Cox leads a team that has published widely using general practice data as their main source. Originally such research was done using high-recording practices in Trent, sometimes recruited through the Trent Focus' Collaborative Research Network, with data being extracted using a program called MIQUEST[1-7]. For research using MIQUEST report style queries this involves writing queries (a skilled and time consuming task), visiting the practices to run the queries and then integrating the data into a common file (which is again a skilled and time consuming task). If a search has not run correctly, the practice needed to be revisited.

There are alternatives, CPRD and THIN (In Practice Systems), THIN and ResearchOne (SystmOne) extract GP data. Although useful for research, these datasets have limitations:

- Geographical coverage is limited
- There is limited recent information on data quality
- The access costs can be high
- Some are mainly used by the pharmaceutical industry for post marketing surveillance

EMIS is a general practice computer supplier with their system installed in 5,400 practices (over half of the practices in the UK). EMIS have previously explored the creation of a large database with other partners but discussions have never come to fruition. The Nottingham team has the technical and research skills to create and run a large general practice database; it also has good relationships with general practices keen to participate in research.

The combination of EMIS and the Nottingham team offered sufficient national standing to ensure ownership by the profession. In 2002, we co-opted leading national figures to a pilot advisory group. We undertook a successful pilot project involving 22 practices which are participating in other studies approved by Trent MREC (MREC/01/04/012 and MREC 02/4/052) and Nottingham LREC (P2100201). These studies have compared data extracted by the QResearch methodology with that obtained from MIQUEST and found no important discrepancies. In 2003, following the successful conclusion of the pilot project, the national project was begun and has continued since then with favourable annual reviews by the ethics committee and national QResearch advisory board.

2.3 General “ways of working” for QResearch

The following apply to all uses of the QResearch database:

- QResearch operates as a not-for-profit collaboration between the University of Nottingham and EMIS (the leading supplier for computer systems in primary care).

- The accounts are be transparent to the two parent organisations and the QResearch Advisory Group; access fees will be agreed in order to ensure fair and reasonable access to researchers and other users, while ensuring the efficient operation of the database.

- The extraction of data from practices for QResearch and the general methodology used will be covered by research ethical committee approval.

- The QResearch Advisory Board will oversee the operation of QResearch, including setting the criteria for access by prospective users.

- The QResearch Scientific Committee will be required to give prospective consent for access to the QResearch database for all users

- All users will be provided with data that is appropriate to their requirement. For researchers that will normally mean a patient level analysis (a file that contains records at anonymised individual patient data), only containing variables that relates to their hypothesis and a sample size sufficient to answer the research question. For others it will mean a tabular output containing no patient level data. For further details, see below.

- Requests for the provision of data will be risk-assessed by the QResearch Scientific Committee to ensure the highest protection for patient confidentiality. Users of the QResearch database will be required to sign that they will not try to attempt to identify any patient(s) or practices. All users will give signed assurance to the following questions:

  - ➢ Did you have the original idea for this research project?
  - ➢ To your knowledge is this work original and capable of publication as original research in a peer- reviewed journal?
  - ➢ Are you free to undertake this study and publish its findings without needing to clear it with the funding source or any other organisation?
  - ➢ Do you agree to acknowledge the source of QResearch data in any publication, paper, report or software/tool?
  - ➢ Do you agree NOT to attempt to identify patient(s) or practice(s)?
  - ➢ Do you undertake to provide a copy of the final report of the project and copies of any publications within one year of the project completion?
  - ➢ Do you agree NOT to release the data to any third party including the funder, sponsor or other such body?
  - ➢ Do you agree not to use the data for any other project except that which is expressly described in your protocol
  - ➢ Do you undertake to check the data you are given within a month of receipt and report back any problems within that time?
  - ➢ Do you have a statistician on the project team who has contributed to the design of the study and will advise on the analysis?
  - ➢ Do you agree to have a project summary on the QResearch website once the project starts?

2.4 Research Governance and custodianship

- The custodian of the data is Professor Julia Hippisley-Cox - Professor of General Practice and Epidemiology at the University of Nottingham and an NHS General Practitioner. She is indemnified by the Medical Defence Union and by the University of Nottingham.

- The University of Nottingham is the research sponsor for QResearch and responsible for ensuring research governance.

- The East Midlands Research Ethics Committee is responsible for ethical approval and monitoring adherence to protocol. Along with the QResearch Advisory Board, the ethics committee reviews QResearch on an annual basis.

2.5 Data Security

This section describes the arrangements for the handling of the data, including the two servers at EMIS and Nottingham, which service QResearch. Technical details are given in the enclosed document entitled “Systems Levels Security Policy QResearch Data Linkage Project 2015” which is reviewed annually. The next section provides a lay summary.

There are two main computers (servers) involved in the QResearch project.

- The data collection server at EMIS. This server will be linked to practices via the NHSnet in order to undertake the triggered upload ONLY after the practice has authorised the upload by activating the QResearch module within its surgery system.
- The research server, which will house the resulting aggregated database, and which will be located at The University of Nottingham. The research computer is a stand-alone computer (i.e. it will not be linked to the NHS net or external networks). This computer will be the single point of access to the data collected by QResearch.

Each of the two servers (at EMIS and at Nottingham) is used solely for the purposes of QResearch.

EMIS only transfer QResearch to one organisation namely, the University of Nottingham. The data transfer is secure as the data is de-identified prior to extraction and the data are also encrypted.

EMIS and the University of Nottingham are contractually bound not to use the data collected by QResearch for any other purpose that that stated within this protocol.

a. The QResearch database consists of a triggered upload of all coded data from patients registered with the participating practices. By coded data, we mean all computer entries which have been coded using the Read or other similar code classifications. No clinical free text will be extracted.
b. EMIS patches a look up table to the practice which maps postcode to the census variables (such as deprivation scores and rurality) associated with the relevant

electoral ward or enumeration district. These data are uploaded into the patients' record within the practice system and hence they can be extracted without the need for any postcode information to ever leave the practice.

c. The uploads are undertaken by EMIS after the practices have given informed consent and activated the upload of QResearch from within their own practice system.
d. No strong patient identifiers are extracted from the general practices contributing to QResearch. Each patient will be assigned a unique code (pseudonymised) in order to maintain the chronological integrity of the database and to allow follow up of individuals and cohorts. Full details of anonymisation are given below.
e. Since 2011, QResearch has been linked to three additional data sources Hospital Episodes Statistics database, cancer registries and ONS mortality registrations. The data linkage is undertaken using a pseudonym on the QResearch database without any strong patient identifiers being supplied from any of the source systems (see further details of the data linkage methodology on page 22).
f. Incremental uploads are undertaken by EMIS and new versions of the database are transferred to Nottingham at intervals as required to meet the needs of the research.
g. Patients will be able to request that their data is not included in the anonymised upload from the practice. This is implemented by the use of a specific Read code provided by EMIS which allows the data from individual patients to be filtered out of the data collection.

There are three main issues regarding the security of the data on the servers. These are:

(1) process of de-identification – i.e. the measures which are taken to ensure complete confidentiality of patients and also of participating general practices
(2) the physical security of the server - measures to restrict physical access and prevent theft
(3) electronic security of the server - measures to prevent unauthorised access and monitor authorised access

2.6 De-identification

2.6.1 Pseudo-anonymised data

No data is extracted from a general practice database that contains any strong patient identifier, such as name, address, full postcode, date of birth etc. The practice computer allocates a unique number to each patient (known as a GUID). This GUID is used by the practice system to allocate later data to the same patient file. The collection server cannot identify which patient the GUID refers to. This additional protection prevents the potential for the GUID from the research database being taken back to the practice, the database being illegally accessed and the GUID cross referenced back to the patient. All these data are referred to as pseudonymised.

2.6.2 Anonymised data

Researchers, having gone through the process of approval, will be given, if appropriate, access to the GP records which contain records for individual patients. However these records will not contain a GUID and are therefore truly anonymised.

When the database is interrogated for information for morbidity analyses, the results will not contain any records for individual patients. The outputs will be in tables or graphs and we refer to these as tabular analyses. These data are anonymised.

2.6.3 Section 60/section 251 support

In order to determine whether Section 60 support was necessary to cover the process of anonymisation/pseudonymisation in 2003, we contacted Sean Kirwan from the Department of Health with a copy of the protocol and details of the processes to be used. He advised us that Section 60 support was necessary only when patient identifiable information is required and it is not practicable to either obtain patient consent or use anonymised/pseudonymised data. With the process of pseudonymisation employed in QResearch, no patient identifiable information is shared with, or processed by, a third party (i.e. an individual or organisation not employed by the GP practice) and hence Section 60 support is not required for the QResearch database.

Advice regarding the need for section 251 support (which replaced section 60 approval in 2008) was sought in 2011 from the Ethics and Confidentiality Committee (ECC) of the National Information Governance Board. This was in order to link general practice data from the HES, Cancer Registry and mortality data at individual patient level. The linked cancer, mortality and hospital data enables researchers based at the University of Nottingham to analyse additional information on patient characteristics, treatment and outcomes which will improve the epidemiological analyses of studies since the data will be more complete (the current QResearch database does not capture all this information reliably). The linked also contains additional detail to allow research into the causes and outcomes for major diseases including the development and validation[8] of tools designed to assess risk of cardiovascular disease[9-11], osteoporotic fracture[12], risk of current cancer[13-20] and future cancer[21]. Without the data linkage, for example, research may under-estimate the incidence of cancer[21], or the risk associated with interventions such as prescribed medicines[22-27].

Following a detailed review, the ECC concluded that, given that (a) no strong identifiers were extracted from the system and (b) there was an irreversible pseudonymisation of the NHS number prior to disclosure from the source system (c) the strong IG controls in place at the University of Nottingham, then the extraction did not constitute identifiable data and hence section 251 support was not required. See enclosed letter entitled “NIGB advice QResearch Sept 2011.pdf”. For further details of the method for data linkage see page 22.

2.7 Security arrangements

2.7.1 Physical security

The data collection server at EMIS has two critical security related roles.

1. It collects the pseudonymised data generated by the practice system
2. It assembles the research database to be used by all subsequent users of the data for secure onward transmission

It also is the single point of failure in the collection process and thus must have resilient connections to the NHSNet and resilient hardware

Therefore it is a security requirement that the server is hosted in a secure data centre with full NHS security clearance, personnel access restrictions and physical access obstacles. This includes steel doors, ID cards, close circuit television etc. The managers of the server must also be subject to full security reviews and the management organisation should ideally also be ISO accredited for security. EMIS meets these requirements.

The QResearch server is hosted in the EMIS secure data centre. Physical access to the server is restricted to specifically named engineering staff who use strong authentication to gain access for the purpose of hardware repair.

2.7.2 Ensuring authorise access

Up to three named support personnel have access to the dedicated QResearch server. This includes the EMIS software support manager and up to two support staff. These personnel have a dual role in software support, providing online support to EMIS customers as a whole. All databases are password protected, and in addition hardware access authentication is used. Logs of all access to the server will be maintained.

2.8 Security arrangements in Nottingham

The main issues for data security in Nottingham are guaranteeing physical security and preventing unauthorised access.

2.8.1 Physical security

The stand-alone computer will be in a locked room with restricted access (named key holders). There is CCTV in the building. The computer itself is in a metal cage which is locked and secured to the floor. The computer is not linked to the external networks.

2.8.2 Ensuring authorised access

Named individuals will have access to the computer are bound by confidentiality clauses in their contracts. Only Julia Hippisley-Cox and the software developer John Croasdale have they will control all accesses to the database on a daily basis as described in the accompanying systems level security policy (SLSP). The advisory board are notified if a new member of staff is required to access the database directly. The data on the research computer are be encrypted and that passwords will be required to access the data. As with the EMIS server, all accesses to the data will be logged (time, user) using electronic tracking software.

2.8.3 Practice or patient identification

One named member of staff in Nottingham and one in EMIS will have a list of the practices which have given and given informed consent to participate in QResearch. This list is be kept on a separate computer from the EMIS file server or the research server in Nottingham; and will be encrypted. The list of participating practices will not be released to other individuals or organisations by EMIS or Nottingham. There are no patient identifiers on the database because of the anonymisation process outlined above. In this way, patient confidentiality is completely secure.

Members of the Advisory Board will undertake, as a minimum, annual site visits in Leeds and Nottingham to check the adequacy of the security measures. The Advisory Board will decide the frequency of the site visits. Logs of all access to the computer will be made available to the Advisory Board on request.

2.9 Users needing access to patient or practice level data

Users requiring access to patient level GP data will need to fulfil the following criteria:

- Did you have the original idea for this research project?
- To your knowledge is this work original and capable of publication as original research in a peer- reviewed journal?
- Are you free to undertake this study and publish its findings without needing to clear it with the funding source or any other organisation?
- Do you agree to acknowledge the source of QResearch data in any publication, paper, report or software/tool?
- Do you agree NOT to attempt to identify patient(s) or practice(s)?
- Do you undertake to provide a copy of the final report of the project and copies of any publications within one year of the project completion?
- Do you agree NOT to release the data to any third party including the funder, sponsor or other such body?
- Do you agree not to use the data for any other project except that which is expressly described in your protocol

- Do you undertake to check the data you are given within a month of receipt and report back any problems within that time?
- Do you have a statistician on the project team who has contributed to the design of the study and will advise on the analysis?
- Do you agree to have a project summary on the QResearch website once the project starts?

Users of output from the QResearch database must provide QResearch with copies of publications or reports. These will be made available to the QResearch Advisory Group and, unless confidential, to the QResearch practices on request

The linked data for mortality, cancer registrations and hospital episode statistics is only available on the secure QResearch server located at the University of Nottingham. The linked data are not made available outside of the University of Nottingham

2.10 Users needing access to tabular output

Most analyses resulting in tabular output will be either using it for establishing evidence to use in research applications including pilot studies and sample size calculations (referred to here as "pilot studies") or for describing care need, care or outcomes in general practice (referred to here as "morbidity analyses")

- Those wanting to undertake research studies must fulfil the following criteria:

    - There will be a named principal investigator and named co-investigators
    - There will be a written protocol with a clear statement of the intended research question, the pilot data required and an intention to develop a full research protocol
    - They will agree to acknowledge QResearch as the source of the pilot data in any application, publication or report
    - The QResearch Scientific Committee will need to give approval within the guidelines set out by the QResearch Advisory Board before analysis can occur and data can be supplied to the user

- The QResearch team in Nottingham undertake morbidity analyses in order to establish the accuracy, completeness and functionality of the QResearch database.

- It is recognised that such analyses may be of general value to a range of individuals and organisations. It is intended therefore to make such analyses, where they have been undertaken, would be made publically available through the related publications on the QResearch website.

2.11 Criteria for access to the QResearch database

The QResearch Advisory Board will draw up the criteria for access that will be applied by the QResearch Scientific Committee. It is anticipated that the QResearch Advisory Board may wish to refine on the criteria set out in this document over time at the annual review. The up to date criteria are published in the QResearch website (www.qresearch.org).

2.12 Practice recruitment

The University of Nottingham, in collaboration with EMIS, has invited all EMIS practices to invite them to participate in the national scheme. See page 17 for a copy of the practice information sheet which is also available on the practice website. We have recruited approximately 1000 practices geographically dispersed throughout the UK. If more practices volunteer than are required, we will randomly select practices, stratifying for partnership size, deprivation.

3 Process for accessing data:

We process requests for access to data as follows:

Principal Investigator (PI) for a research proposal wishes to use QResearch to test or generate a hypothesis

The PI discusses the research proposal with the QResearch team to ensure that the research protocol is appropriately framed and feasible.

“In principle” approval is given for suitable projects so that the PI can seek funding with a good probability that the project will be approved by the QResearch Scientific Committee.

The project achieves appropriate research approval/funding.

Once funding is obtained, the protocol is reviewed by the research committee according to the standard criteria

The data specification for the data required to undertake the research project is prepared and the extraction is scheduled.

The data, once cleaned and checked, are provided to the PI once the agreement has been signed and the fees paid.

4 Professional and academic oversight

There are three board/committees

(1) QResearch Management Board
(2) QResearch Advisory Board
(3) QResearch Scientific Committee

4.1 QResearch Management Board

- The management board represents the key interests in the setting up and running of QResearch - EMIS and the team at the University of Nottingham.
- It owns QResearch and runs it as a non-profit making venture.
- It is responsible for strategy and development.
- It consists of one director from EMIS (Dr Shaun O'Hanlon) and one from the University of Nottingham (Julia Hippisley-Cox). Shaun O'Hanlon is the Clinical and Development Director for EMIS. Julia Hippisley-Cox is Professor of Clinical Epidemiology and General Practice at the University of Nottingham.

4.2 QResearch Advisory Board

4.2.1 Terms of reference/remit

- To oversee the general working of QResearch including the handling of the data, the type of analyses undertaken and access to the database.
- To oversee communication with and benefits back to contributing patients and practices
- To agree and update the criteria and principles for access to the QResearch database and oversee their application.
- To review any changes to the context of the data extracted for QResearch or changes to the terms of its use
- To advise on policy for accessing data (i.e. how quickly should researchers get data; how much should it cost)
- To offer general advice on professional issues to the QResearch management team

4.2.2 Membership

The board represents the key stakeholders in QResearch in order to gain and retain the respect of the public, the NHS, practices and the research community. Its membership includes representatives of:

- The Royal College of General Practitioners
- The British Medical Association itself and its General Practitioners Committee
- Society for Academic Primary Care

- EMIS National User Group
- EMIS
- Patient representation including National Association for Patient Participation
- University of Nottingham
- Chair of the Scientific Committee

In order to oversee the ongoing development of QRESEARCH we have appointed an Advisory Board as follows:

- Mr Jon Ford
- Dr Caroline Mitchell
- Mr Antony Chuter
- Dr Patricia Wilkie
- Professor Azeem Majeed
- Dr Jonathan Meadows
- Dr Joanne Reeve
- Dr Geoff Schrecker
- Professor Julia Hippisley-Cox

Further details of the advisory board are published here
http://www.qresearch.org/SitePages/The%20Advisory%20Board%20In%20Detail.aspx

4.2.3 Minutes of advisory board meeting

The minutes of the advisory board are published here
http://www.qresearch.org/SitePages/Advisory%20Board.aspx

5 Practice information sheet

This is published at the following location:

http://www.QResearch.org/SitePages/Information%20for%20practices.aspx

Background

In 2003, the University of Nottingham and EMIS established a new high quality ethical database for research known as QResearch. This database, which is now the largest of its kind worldwide, consists of the pseudonymised records of over 18 million patients who have been registered with over 1000 practices spread throughout the UK. Results of research undertaken using QResearch are all published and can be found at http://www.QResearch.org .

We have many examples of projects which have helped inform health care policy or enrich understanding of the causes and outcomes from diseases and their treatments.
Click here to see an overview of QResearch and related projects.

Why is QResearch wanting to expand?

- We want to expand the size of QResearch in order to
- get better coverage of some parts of the UK (we need at least three practices in a geographical area in order to be able to report at that level)
- increase to increase the number of practices with longitudinal data tracking back over a decade to enable us to develop new risk prediction algorithms such as QRISK (new CVD risk score developed - see http:\\www.qrisk.org) which can be integrated back into the clinical system
- enhance our ability to undertake research on rare diseases
- to enable the rapid evaluation of the safety and effectiveness of newly introduced medication.

Why has my practice been invited to participate?

We are inviting all practices using EMIS Web to take part.

Does the practice have to take part?

It is entirely up to the practice as to whether they wish to take part. Practices who agree to take part will activate QResearch within their surgery system. The practice will only be included in the project once this activation has been done.

What will happen if the practice takes part in QResearch?

GP computer systems are currently connected to EMIS via the NHSNet for support and patching of system upgrades. EMIS will use this link to first upload key census related variables such as the deprivation score (called the Townsend Score) and a marker of rurality into the patient’s electronic record. These data items will be stored in the patients’ record but will not be available on the patient’s registration screen. The scores are derived from the national census and are linked to the electoral ward or output area in which the patient lives. The scores are used in research to measure or monitor uptake and provision of health services.

The practice system will then generate an upload of the pseudonymised data and transmit this data via the N3 network to a dedicated server within EMIS. The data are pseudonymised with a 'one way' hashing algorithm which cannot be reversed thereby protecting patient confidentiality. The data are updated on a daily basis and this update does not interfere with the running of the practice or any back-ups.

EMIS undertakes a secure transmission of the aggregated data to the University of Nottingham. To ensure maximum security, the data are encrypted and will be stored on a very secure server in the University of Nottingham. Physical access to the database is very restricted. The University of Nottingham is the only point of access to the full database.

What data is uploaded?

QResearch only uploads pseudonymised coded data: year of birth, sex, registration dates, date of death, deprivation scores, ethnicity, referrals, consultations, prescriptions, clinical values (e.g. blood pressure), test results (e.g. full blood count) and clinical problems/events such as diagnoses. Users are assigned a pseudonymised code. Patient dates of birth are rounded to year of birth. No free text or strong patient identifiers are uploaded. Patients are excluded from the uploads if they have a read code indicating they have opted out of QResearch (EMISNQOP15) or the Summary Care Record. Any items marked as confidential by the GP are also excluded from the uploads.

How does the data linkage work?

The QResearch database is be linked to other sources of data such as cancer registries, deaths data and hospital episode statistics. This information is linked to QResearch using the same pseudonymisation process which is applied to the NHS number in the secondary data source before the data is released to the University of Nottingham. The data is then linked on the pseudonym without the need for disclosure of any strong identifiers to QResearch.

How do I activate QResearch in EMIS Web?

Please click here for instructions on how to activate QResearch in EMIS Web.

Can the practice withdraw from QResearch at any time?

Practices can withdraw from QResearch at any time and without giving a reason. If a practice does decide to withdraw, no further data from the practice will be extracted for future research projects. If a practice wishes to withdraw, it just selects the ‘deactivation’ option from the QResearch agreement in the ‘Information Upload Services” module within their surgery computer system.

What are the benefits of taking part?

Practices will be contributing to the improvement of patient care through a good quality ethical research programme, carried out on large volumes of data. Practices are able to access tools and utilities within their practice system based on research findings from QResearch such as the QRISK2, QDiabetes, QFracture and QCancer risk prediction tools. In addition, we will try to facilitate any of our practices who wish to undertake their own research project by discussing the project with them and providing advice on how to get the submission to the Scientific Committee ready.

QResearch has been set up on a not-for-profit basis and there is no funding available to cover practice time. However, the amount of time involved in activating the upload is minimal.

Are there any risks to taking part?

We do not believe there are any significant risks to taking part in QResearch. The main issue is security and confidentiality of the data collected and we have put systems in place to ensure maximum security of data. We have approval from Trent Multi-Centre Ethics Committee and have consulted widely with professionals with expertise in this area. All information will be held in accordance with the principles and conditions set out in the Data Protection Act 1998 and with proper safeguards to ensure confidentiality. No patient identifiable information will be extracted from your database and therefore patient identifiers will never be released by QResearch. The identity of practices in QResearch will be kept confidential. Although the practice identifiers will be known to the EMIS and one member of the QResearch team, no practice identifiable data will be released by QResearch to a third party. All categories of uses of the data require the approval of the National Advisory Board. Under no circumstances will a copy of the entire database be made available to any individual or organisation.

Practice notice

Trent Multi-centre Ethics Committee have advised us that all practices contributing to collection of anonymised data for research need to display a notice in their waiting room informing patients. The notice can be found here. Patients who do not wish their data to be included in the upload are able to opt out by informing their General Practitioner who then assigns a specific EMIS Read code (EMISNQOP15) which will cancel any future data collection from that patients’ record.

What is the Advisory Board and what is its role?

The role of the Advisory Board is:

- To oversee the general working of QResearch including the handling of the data, the type of analyses undertaken and access to the database.

- To agree and update the criteria and principles for access to the QResearch database and oversee their application.
- To advise on policy for accessing data (i.e. how quickly should researchers get data; how much should it cost)
- To offer general advice on professional issues to the QResearch management team
- The full board will represent the key stakeholders in QResearch and to gain and retain the respect of the public, the NHS, practices and the research community. Its membership will include representatives of the following organisations:
- The Royal College of General Practitioners
- The British Medical Association itself and its General Practitioners Committee
- EMIS National Users Group
- A representative of the practices contributing data to QResearch
- EMIS
- Two patient representatives
- University of Nottingham
- Chair of the Scientific Committee

What are the requirements for the projects undertaken using QResearch?

There are two types of projects that will use QResearch. The first type will be research projects and many of these will require a subset of QResearch to analyse. Research projects undertaken using the QResearch data will fulfil the following criteria

(a) They have a named principal investigator and named co-investigators (at least one co-investigator will have an academic track record)
(b) They will have a written protocol with a clear research question
(c) They will have current ethical committee approval for the project
(d) They will have a research sponsor and will meet the requirements of research governance
(e) They will agree to acknowledge the source of data in any publication or report (which will never identify any individual practices).
(f) They will guarantee their academic independence, including their ability to publish their findings regardless of the implications for any sponsor or funder
(g) The QResearch Scientific Committee will need to give approval within the guidelines set out by the QResearch Advisory Board before any analysis can occur and before the data can be supplied to the user.

The second type of project will require tables of data. These morbidity surveys will not require the release of a subset containing patient level data, but just analyses. The results of these analyses will be placed on the QResearch website (www.QResearch.org). Applicants for tabular data will need to be approved by the Advisory Board.

Under no circumstances will QResearch data made available to pharmaceutical companies for post marketing surveillance. There may be circumstances in which is it may be in the public interest to undertake analyses of safety for new and commonly used medications and in these special circumstances it may be necessary for QResearch to have contact with the pharmaceutical industry via a third party such as the MHRA or the MRC/Wellcome. For example, in a flu pandemic situation we may need to rapidly assess whether a newly developed flu vaccine is likely to be safe and effective. As with any research project however, will be undertaken by a researcher who can guarantee academic independence in design and analysis as well as freedom to publish in according with our governance framework. All studies will be approved by the Scientific Board and notified to the Advisory Board. The protocol for the studies will be publically available as will any information on funding to support this activity.

Practices who wish to be in QResearch but not allow their data to be used for drug safety studies will be able to indicate this and their data will not be used for this purpose.

Will the identity of practices taking part in QResearch be kept confidential?

Yes, the participation of the practice in QResearch will be confidential. The Advisory Board, let alone any users of the QResearch data will not know the identity of the practice. The practice will not be named in any report or publication.

What happens to the results of the research projects?

All projects undertaken using QResearch are to be made publically available without charge. We expect researchers who use QResearch data for their research to publish the findings in peer reviewed journals. We will keep a list of all publications which is available on http://www.QResearch.org

Who is organising the research and who are the guarantors?

QResearch is a joint venture between the University of Nottingham and EMIS. The project is organised and guaranteed by Professor Julia Hippisley-Cox, GP and professor of General Practice (University of Nottingham) and Dr Shaun O'Hanlon (Clinical and Development Director, EMIS).

Contact details:
E-mail: julia.hippisley-cox@nottingham.ac.uk

Further queries

If you have any queries about QResearch, then please contact Professor Julia Hippisley-Cox email. Further details about QResearch can be found on the QResearch website (www.QResearch.org)
If you have any complaints about the way that QResearch is being conducted, please contact:
Dr Geoff Schrecker geoff.schrecker@gmail.com
EMIS National User Group, Suite 15, Enterprise House, Kingsway North, Team Valley, Gateshead, Tyne & Wear. NE11 0SR

6 QResearch Information for Patients

The information below is available from the QResearch website. It was last updated by the QResearch advisory board in July 2015 and is reviewed at least annually.

http://www.qresearch.org/SitePages/Information%20for%20patients.aspx

Q: What is QRESEARCH?

A: QResearch is a not profit making collaboration between the University of Nottingham and EMIS (company which provides computer systems to doctors surgeries). With the permission of the GP surgery, QResearch takes data from GPs computer system, sorts it, links it to other medical information centrally held by the NHS, then makes the information available to medical researchers.

Q: What sort of information is taken from your GPs computer system?

A: Information is taken from the GPs system about the illnesses people have had and what treatment the doctor gave them. Before the information is taken, it is 'de-identified' which means it has had all the information which might identify a patient stripped off including the name, address, postcode, date of birth etc. Only coded medical information (such as blood pressure measurements, diagnoses, prescriptions) is extracted. The free text notes which a GP makes are not included. You cannot be identified from this information.

Q: What is this information used for?

A: The information is used to study patterns of diseases and compare different treatments used. It is also used to discover if, for example, having one type of illness makes people more likely to get another type of illness. It is also used to compare safety of different types of medicines. Recent examples include a study which has looked different types of the oral contraceptive pill to see which ones have the highest risk of side effects such as blood clots and which have the lowest risk. QResearch has also been used to develop algorithms, such as QRISK, which calculates a person's risk of to identify having a heart attack or stroke so that they can lower their risk. The QCancer tool helps doctors identify patients at risk of cancer early. The information is never used for insurance purposes. QResearch solely for research purposes where there is a clear research question.

Q: Who supplies this information?

A: The information is supplied from GPs who use the Emis computer system. A sample of approximately1000 GP practices is used the EMIS computer system (there are approximately 9,800 practices in the UK overall). The database covers a population of over 18 million over the last 20 years.

Q: Exactly what information is sent to QRESEARCH?

A: Only coded medical information is sent. No personal details about you are ever sent. You cannot ever be recognised from the information sent to QRESEARCH. In fact, it would be against the law for personal or recognisable information to be sent. The GPs computer system allocates your information a unique number but the link between you and that number never leaves your GPs practice.

Q: How is the information given to researchers?

A: The information is given to researchers, who are approved by the QResearch team based at the University of Nottingham, in the form of tables. Approved researchers are employed by UK universities and the team always includes at least one medically qualified doctor. No personal information (your name, address etc.) can be given to the researchers because QRESEARCH does not have this information.

Q: can I opt out of QRESEARCH?

A: yes you can opt out by asking you GP to flag your records with a special code which will prevent the uploads. The code is EMISNQOP15. You can find out if your GP surgery is taking part by looking at the surgery website or leaflet. Practices may also display a notice in the waiting room.

Q: Who runs QRESEARCH?

A: QRESEARCH is managed by the chief medical officer of EMIS – the company who supplies your GP with their computers – and a Professor of General Practice at Nottingham University. QRESEARCH has an Advisory Board made up of medical professionals and patient representatives. The advisory Board ensure that the system is run ethically and is accountable.

Q: Does anyone profit financially from the use of my information?

A: No. Apart from the computer technicians who are paid by Nottingham University, no one is paid for their role in QResearch. QResearch reimburses reasonable expenses for members of its advisory board, covering cost of travel and attendance at meetings.

Further, more detailed information on QRESEARCH and the results of research projects can be obtained from our website:www.qresearch.org.

Originally drafted by Terence Wiseman, Patient Representative, QResearch Advisory Board and last updated August 2015.

7 Data linkage methodology

QResearch adopts a 'pseudonymisation-at-source' approach in which the data are irreversibly pseudonymisation before the data leaves each of the four source systems (GP, ONS mortality, ONS cancer registration, Hospital Episode Statistics).

The HSCIC (which is the source for the hospital episode statistics database) pseudonymises the NHS number on the full HES database before it leaves the HSCIC. EMIS (which supplies the GP data) and the Office of National Statistics (which supplies the cancer and morality data) apply the exactly the same pseudonymisation technique within the source system before the data are extracted. This means that the same ID can be generated and can used to link the records from each data sources together.

The steps are as followed:

a. The HSCIC, EMIS and ONS download the batch processing software (known as OpenPseudonymiser) which is used to pseudonymise the NHS number within the source data. This procedure is undertaken by the staff at the relevant organisation in a consistent fashion. The software is downloaded from the software pages on www.openpseudonymiser.org
b. The software applies a project specific 'salt code' (which is a random text word similar to a password) which is combined with the hashing algorithm to ensure that the resulting pseudonymised ID is unique to QResearch project. The 'salt code' and software is be kept confidential by the source systems and not used for any other project.
c. The OpenPseudonymiser software is then applied to a comma separate file of source data using a secure one-way hashing algorithm to the NHS number which generate a unique pseudonymised ID.
d. The OpenPseudonymiser software then removes the NHS number from the data file so that resulting data file has pseudonymised ID but no NHS number. The batch processor will also replace the full date of birth with the year of birth and remove other strong identifiers such as postcode if present on the data file.
e. The fully de-identified pseudonymised data file is then encrypted by the source supplier before it is securely transferred to Professor Hippisley-Cox at QResearch at the University of Nottingham at a pre-arranged time, who will telephone to confirm safe receipt.
f. On receipt of the pseudonymised data at the University of Nottingham, the pseudonymised ID is used to link the data files together. The pseudonymised ID is further transformed to a second ID for use within the linked databases.
g. No other stronger identifiers are extracted from the any source system (i.e. no name, NHS number or postcode or full date of birth).
h. There is no disclosure by QResearch of any patient or practice information which could directly or indirectly lead to the identification of patients or practices contributing to the QResearch database.

8 References

1. Hippisley-Cox J, Pringle M. Are spouses of patients with hypertension at increased risk of hypertension? A population based case-control study. British Journal of General Practice 1998;**46**:1580-84.
2. Hippisley-Cox J, Pringle M. Depression and risk of ischaemic heart disease in men: authors' reply. BMJ 1998;**317**:1450.
3. Hippisley-Cox J, Pringle M, Crown N, et al. Sex inequalities in ischaemic heart disease in general practice: cross sectional survey. BMJ 2001;**322**(7290):832.
4. Hippisley-Cox J, Pringle M, Hammersley V, et al. Antidepressants as risk factor for ischaemic heart disease: case-control study in primary care. BMJ 2001;**323**(7314):666-9.
5. Hippisley-Cox J, Coupland C, Pringle M, et al. Married couples risk of same disease: cross sectional study. BMJ 2002;**325**:636-38.
6. Hippisley-Cox J, Cater R, Pringle M, et al. A cross-sectional survey of the effectiveness of lipid lowering drugs in lowering serum cholesterol in 17 general practices:how well do they work? BMJ 2003;**326**:689-94.
7. Hippisley-Cox J, Pringle M, Crown N, et al. Does hormone replacement therapy protect against ischaemic heart disease? Evidence from a general practice based case control study. BJGP 2003;**53**:191-96.
8. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open 2014;**4**(8):e005809.
9. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008:bmj.39609.449676.25.
10. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007:bmj.39261.471806.55.
11. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.
12. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ 2012;**344**(may22 1):e3427-e27.
13. Hippisley-Cox J, Coupland C. Identifying patients with suspected gastro-oesophageal cancer in primary care: derivation and validation of an algorithm. British Journal of General Practice 2011;**61**(592):e707-14.
14. Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. British Journal of General Practice 2011;**61**(592):e715-23.
15. Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. BMJ 2012;**344**.
16. Hippisley-Cox J, Coupland C. Identifying patients with suspected colorectal cancer in primary care: derivation and validation of an algorithm. British Journal of General Practice 2012;**62**(594):e29-e37.

17. Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm. British Journal of General Practice 2012;**62**(594):e38-e45.
18. Hippisley-Cox J, Coupland C. Identifying patients with suspected renal tract cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract 2012;**62**(597):e251-60.
19. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract 2013;**63**(606):11-21.
20. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify men with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract 2013;**63**(606):1-10.
21. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. BMJ Open 2015;**5**(3).
22. Coupland C, Hill T, P B, et al. Thiazolidinedione drugs and the risk of cancer in patients with diabetes: nested case-control studies using a primary care database. SAPC conference 2015 2015.
23. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of common non-gastrointestinal cancers: series of nested case-control studies using two primary-care databases. Br J Cancer 2013.
24. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. BMJ 2013;**346**:f114.
25. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to statins and risk of common cancers: a series of nested case-control studies. BMC Cancer 2011;**11**:409.
26. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to cyclooxygenase-2 inhibitors and risk of cancer: nested case-control studies. Br J Cancer 2011;**105**(3):452-9.
27. Logan RF, Vinogradova Y, Coupland CA, et al. Is Low-Dose Aspirin Use Associated With a Reduced Risk of Colorectal Cancer? A QResearch Primary Care Database Analysis. Gastroenterology 2011;**140**(5):S-402.