QRESEARCH

A new ethical high quality general practice derived database for research

Protocol approved by Trent MREC 3rd April 2003

1

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
David Stables, EMIS computing

QRESEARCH
A new ethical high quality general practice derived database for research

Contents

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
      David Stables, EMIS computing

1 SUMMARY

When fully established, QRESEARCH will be a high quality clinical database derived from English general practices. The aim of QRESEARCH is to develop and maintain a high quality database of general practice derived data for use in ethical medical research.

The special features of QRESEARCH are:

- A nationally representative sample of 300-400 practices (2.5m to 3.5m patients) will contribute their data
- Census variables such as deprivation and rurality will be uploaded onto the practice database
- The data from practices will be assessed and its quality (in terms of completeness and accuracy) accredited.
- The data extracted will contain no strong patient identifiers
- Practices will be required to display a notice in their surgery waiting rooms informing patients that the practice contributes anonymised data for research. Patients will be able to opt if they would prefer their won anonymised data not to be included in the upload.
- Within the aggregated full database, a unique number assigned to each patient and to each practice to allow longitudinal tracking (pseudo-anonymised). Researchers will only have access to fully anonymised data (without the pseudo-anonymised identifying code).
- The database will be open to bone fide researchers who must be able to publish whatever the findings and who have a protocol with ethical committee approval
- Each use of the database must also be approved by the QRESEARCH scientific committee
- The researchers will be provided with patient or practice level data that is required to answer their research question (not the whole dataset)
- The researchers will only be 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 will be undertaken to demonstrate the accuracy and completeness of the data and will be made available for morbidity analyses
- Data quality checks will include comparison of birth rates, death rates and prescribing patterns with other aggregated sources (ONS, PACT data)

The key organisational features are:

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

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
        David Stables, EMIS computing

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 for use in ethical medical research. It will create a new, high quality, primary care derived database that contains descriptive data on the health needs, risks, care and outcomes for a large population. It will be one of the highest quality and largest such datasets in Europe and possibly the world. Research includes projects which generate or test hypotheses. It also includes pilot work to determine sample size calculations and feasibility of specific analyses.

In developing and validating the database, we intend to undertake a detailed survey of morbidity in general practice. The results of this will be available in tabular (summary) format as in previous publications such as the 4^h 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 (98%) and well used (over 30% of practices only use their computer for recording their clinical records), and they are often highly accurate and complete. The University of Nottingham holds a contract with the NHS Information Authority for Primary Care Information Services (PRIMIS) that builds on the success of IT in primary care by facilitating better use of systems and improvements in data quality.

Researchers have begun to exploit the potential of the clinical databases in general practice. In Nottingham, Julia Hippisley-Cox and Mike Pringle lead a team that has published widely using general practice data as their only source[1,2,3,4,5,6,7,8,9,10]. At present such research is done using high-recording practices in Trent, sometimes recruited through the Trent Focus' Collaborative Research Network[11], with data being extracted using a program called MIQUEST[12]. 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 needs to be revisited.

There are alternatives, GPRD (In practice systems) and IMS and DIN (Torex) extract GP data. Although useful for research, these datasets have limitations:

- There is limited recent information on data quality
- The access costs can be high particularly for GPRD
- They are predominantly used by the pharmaceutical industry for post marketing surveillance

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
David Stables, EMIS computing

EMIS is a general practice computer supplier with their system installed in 5,400 practices (over half of the practices in England). 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 will offer sufficient national standing to ensure ownership by the profession. We have already co-opted four leading national figures to a pilot advisory group. We have undertaken 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.

2.3   General "ways of working" for QRESEARCH

The following will apply to all uses to which the QRESEARCH database is applied:

The operation of QRESEARCH:

➢ QRESEARCH will operate as a not-for-profit joint venture between the University of Nottingham and EMIS.

➢ The accounts will 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 Group 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 or practice level analysis (a file that contains records at anonymised individual patient or practice level), only containing data that relates to their hypothesis. For others it will mean a tabular output containing no patient or practice 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

5

to use weak patient identifiers to attempt to identify any patient(s) or practices. All users will give signed assurance that:

- o the data provided will only be used for the purposes agreed with QRESEARCH
- o no data will be released to any third parties
- o if they wish to undertake analyses of the provided data beyond their agreement with QRESEARCH, then further ethical committee and written QRESEARCH Scientific Committee approval must first be obtained

2.4   RESEARCH GOVERNANCE AND CUSTODIANSHIP

The custodians of the data are Julia Hippisley-Cox and Mike Pringle who act as the Trusted Third Party. Both are General Practitioners and also hold academic posts. They are 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 Trent Multi-Centre Ethics Committee is responsible for ethical approval and monitoring adherence to protocol.

2.5   DATA SECURITY

This section describes the arrangements for the handling of the data, including the two servers at EMIS and Nottingham, which will service QRESEARCH

There will be two main computers (servers) involved in the QRESEARCH project.

(a) The data collection server at EMIS. This server will be linked to practices via the NSHnet in order to undertake the triggered upload ONLY after the practice has authorised the upload by activating the QRESEARCH module within its surgery system.
(b) 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 NHSnet or internal 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) are used solely for the purposes of QRESEARCH.

The data collection server will be located at EMIS in the short to medium term. In the long term, the server may be re-located if a satisfactory secure alternative can be found which the organisation an afford (see below).

6
Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
          David Stables, EMIS computing

EMIS will only transfer QRESEARCH data to one organisation namely, The University of Nottingham. The data transfer will be secure as the data will be encrypted.

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

➢ Before entry into QRESEARCH the practice database will be augmented with the deprivation and rurality codes allocated to each patient's postcode from the census data [the postcode is not subsequently extracted but the census variables are]

➢ The QRESEARCH database will consist 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 (histories etc) will be extracted.

➢ The upload will be undertaken by EMIS after the practices have given written informed consent (the consent form will be returned to The University of Nottingham) and activated the upload of QRESEARCH from within their own practice system.

➢ EMIS will patch a look up table to the practice which maps postcode to the census variables (such as deprivation scores, ethnicity and rurality) associated with the relevant electoral ward or enumeration district. These data will be 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. This procedure will enable researchers to identify possible inequalities in the uptake and provision of health services and this methodology will avoid the current need for a researcher who wishes to include social data to visit the practice to perform a local search that includes the postcode, to associate the ward data with each patient file, and to then strip off the postcode.

➢ No strong patient identifiers will be extracted from the general practices contributing to QRESEARCH. Each patient will be assigned a unique code (pseudo-anonymised) 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.

➢ Incremental uploads will be undertaken each day by EMIS although new versions of the database will be transferred to Nottingham at intervals as required to meet the needs of the users

➢ The data transmitted each day will be available for the practices to view on their system after it is transmitted if they choose in accordance with COSODE guidance

➢ Patients will be able to request that their data is not included in the anonymised upload from the practice. This will be 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 and also to delete any data already uploaded on the collection server at EMIS

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

(1) process of anonymisation – ie the measures which are taken to ensure complete confidentiality of patients and also of participating general practices

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
        David Stables, EMIS computing

(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   ANONYMISATION

Pseudo-anonymised data

No data is extracted from a 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 (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. As an additional protection, this GUID is further encrypted at the point of collection by the collection server using a hash key which the collection server maintenance personnel do not have access 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. This process of anonymisation is much stronger than the MIQUEST identifiers.

Similarly, each practice is also allocated a unique code but no practice identifiers are retained. In the national phase, when the data are fully pseudo-anonymised, no-one (including EMIS) will be able to track back to the practice of origin from the resulting database. [Note: in the pilot phase practices have agreed to be identified in order to check data validity against MIQUEST searches. This will not apply in the full QRESEARCH database]

All these data are referred to as pseudo-anonymised.

Anonymised data

Researchers, having gone through the process of approval, will be given, if appropriate, files that 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 studies etc, the results will not contain any records for individual patients. The outputs will be in tables, graphs etc and we refer to these as tabular analyses.

These data are referred to as anonymised.

Section 60 support

In order to clarify whether Section 60 support was necessary to cover the process of anonymisation/pseudo-anonymisation, 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

8

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
          David Stables, EMIS computing

consent or use anonymised/pseudonymised data. With the process of pseudo-anonymisation employed in QRESEARCH, no patient identifiable information will be shared with, or processed by, a third party (ie an individual or organisation not employed by the GP practice) and hence Section 60 support is not required for the QRESEARCH database.

2.7   SECURITY ARRANGEMENTS IN EMIS

Physical security

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

1. It hosts the mechanism for the additional pseudo-anonymisation over and above the initial pseudo-anonymisation performed by the practice system.
2. It generates 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.

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.

Practice and patient anonymisation
As described above

2.8   SECURITY ARRANGEMENTS IN NOTTINGHAM

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

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 will be in a metal cage which is locked and secured to the floor. The computer will not be linked to the external or internal networks (there will be NO Internet linkage) so there is no possibility that the data will be "hackable".

Ensuring authorised access

Named individuals will have access to the computer who will be bound by confidentiality clauses in their contracts. The number of individuals with access to the computer will depend on the scale of the QRESEARCH venture. Initially, it is anticipated that only Julia Hippisley-Cox and the research associate (to be appointed) will have direct access and they will control all accesses to the database on a daily basis. The advisory board will be notified each time a new member of staff is required to access the database. The data on the research computer will be encrypted (PGP or similar) 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.

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 returned signed consent to participate in QRESEARCH during recruitment. This list will 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 will be no practice or patient identifiers on the database because of the anonymisation process outlined above. In this way, patient and practice confidentiality will be completely secure.

There may be occasions where researchers wish to relate practice characteristics (for example practice size) to some clinical process or outcome. It will **not** be possible to link practices to external data sources (such as PACT) as the identity of the practice will not be known and there will not be any code which can link it to external sources of data. We will not therefore have ANY information on GP age or sex, achievement of target or contract payments.

However it might be possible to derive a limited number of practice level characteristics directly from the anonymised database held on the research computer in Nottingham for example, list size, average deprivation score and rurality. Only banded data (for example, the list size can be banded < 5,000; 5000-7999; 8000 +) will be provided to prevent any possibility of identifying the practice.

The only exception to this would be data which were entered by the practices themselves onto a template which was then included with the main download. If this

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
          David Stables, EMIS computing

were to be undertaken, then practices would need to give fully informed consent and enter the data themselves onto a template within their surgery system.

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 or practice level data (with patients always anonymous and practices anonymous unless practice permission sought and given) will need to fulfil the following criteria:

    o There will be a named principal investigator and named co-investigators
    o There will be a written protocol with a clear research question
    o There will be current research ethical committee approval for the work
    o There will be a research sponsor who will meet the requirements of research governance
    o They will guarantee their academic independence, including their ability to publish their findings regardless of the implications for any sponsor or funder
    o They will agree to acknowledge QRESEARCH as the source of data in any publication or report
    o 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

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

2.10   Users needing access to tabular output – not at patient or practice level

➢ 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:

    o There will be a named principal investigator and named co-investigators
    o 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
    o 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 will need to undertake morbidity analyses in order to establish the accuracy, completeness and functionality of the QRESEARCH database (details can be found later in this document).

➢ It is recognised that such analyses will be of general value to a range of individuals and organisations. It is intended therefore to make such analyses available on two levels – general analyses available freely to any person who wishes to access them through a QRESEARCH website; and more detailed analyses available to specified users.

➢ An initial set of detailed morbidity analyses will provide cross tabulation of, for example, incidence and prevalence of specific diseases and disease groups, tabulated by five-year age group, gender and deprivation. As users request other analyses these may be added to the database and will be available to all users approved to have access to the detailed morbidity analyses. It is anticipated that some analyses will be publicly available free of charge on the QRESEARCH website. Other analyses will be available on a website accessed via an annual license. We will publish a list of all organisations and individuals who have requested morbidity analyses. Where practical, this list will be liked to the specific analysis requested.

➢ Specified users of detailed morbidity analyses will meet the following criteria:

- They will have a specified reason for accessing the detailed morbidity analyses tables
- They will agree to acknowledge QRESEARCH as the source of data in any publication or report
- The QRESEARCH Scientific Committee will need to give approval within the guidelines set out by the QRESEARCH Advisory Board before access can be supplied to the user

It is possible that some organisations requesting tabular analyses will want these by Strategic Health Authority Area. These analyses will only be produced if we can guarantee that no practices can be identified. If there is a SHA with fewer than three practices, then the results will be combined with a neighbouring SHA.

2.11 Criteria for access to the QRESEARCH database

➢ The QRESEARCH Advisory Group will draw up the criteria for access that will be applied by the QRESEARCH Scientific Committee. However it is anticipated that the QRESEARCH Advisory Group will wish to build on the criteria set out in this document. The criteria will be freely available to the QRESEARCH practices and prospective users of QRESEARCH.

3 PRACTICE RECRUITMENT

We will recruit 400 practices geographically dispersed throughout England and Wales. We will randomly select these practices from volunteering practices, stratifying for partnership size, deprivation, and rurality. Practices will be given a choice as to whether they would like to contribute data to research analyses or whether they would like also to contribute data to morbidity analyses. Of the practices recruited, we expect all to participate in research and over 300 to participate in the morbidity survey. Practices that have participated in the QRESEARCH pilot will be invited to participate in the full project if they meet minimum standards for data quality.

The University of Nottingham, in collaboration with EMIS, will write to all EMIS practices in England to invite them to participate in the national scheme. We will write to all the PCTs of the volunteer practices to inform them of QRESEARCH in line with research governance standards.

4 QRESEARCH MORBIDITY AND HEALTH SERVICE UTILISATION STATISTICS

The QRESEARCH database will be used to produce morbidity statistics. Additional morbidity statistics will be produced on request if they meet the criteria and have Scientific Board approval.

4.1 Morbidity and health service utilisation statistics

We will calculate for each year, starting in 2001:
    (a) rates of morbidity (e.g. prevalence and incidence)
    (b) rates of uptakes of appropriate drugs (e.g. statins for IHD) or other disease management/recording (e.g. BP in diabetics etc)
    (c) rates of consultations by GP/nurses including location (surgery/home visit/telephone)
    (d) rates of referrals
    (e) patterns of recording of disease (i.e. patients with drugs but without the associated diagnostic Read codes)

Rates (including inter-practice rates) would be calculated by
    (a) mid-year population
    (b) per 10,000 person years at risk to make it comparable with MSGP4

We will report on the above for the following groups:
    (a) By patient age and gender
    (b) By socio-economic and geographical characteristics. This will be done by assigning patients to the electoral ward of their postcode and linking this to a

table of ward characteristics for 2003 e.g. Townsend score; ethnicity; rurality; housing tenure; unemployment; rurality.

We will compare trends over time as with previous morbidity surveys in general practice.

4.2   Contents :

(a) Infections
(b) Neoplasms
(c) Endocrine and metabolic disorders
(d) Blood and blood forming organs
(e) Central nervous system organs
(f) Circulatory system
(g) Respiratory system
(h) Digestive system
(i) Genito-Urinary System
(j) Skin and subcutaneous tissue
(k) Musculoskeletal system
(l) Complications of pregnancy and childbirth
(m) Injuries and Poisonings

In particular we will produce morbidity statistics on specific conditions and their associated treatments, for example:

- Ischaemic Heart Disease
- Stroke
- Hypertension
- Hyperlipidaemia
- Hypothyroidism
- Psychoses
- Asthma/COAD
- Diabetes
- Osteoporosis
- Pagets Disease
- Peptic Ulcer Disease
- Malignant Neoplasms
- Parkinson's Disease
- Epilepsy
- Thrombo-embolism
- Sero-negative and sero-positive arthritis
- Inflammatory Bowel Disease (Crohns and Ulcerative Colitis)

There will be analyses of selected performance indicators. These will be undertaken to provide feedback to the practices so that they can compare their performance against the whole group of 400 practices. Such tabular analyses will be available for benchmarking and service development. All such analyses will be within guidelines set by the Advisory Board and individually checked by the Scientific Committee.

5 CONFIDENTIALITY AGREEMENT FOR PRINCIPLE INVESTIGATORS

Name …..……………………………………………………

Address...

....

....

I acknowledge the following conditions under which the information is supplied by QRESEARCH:
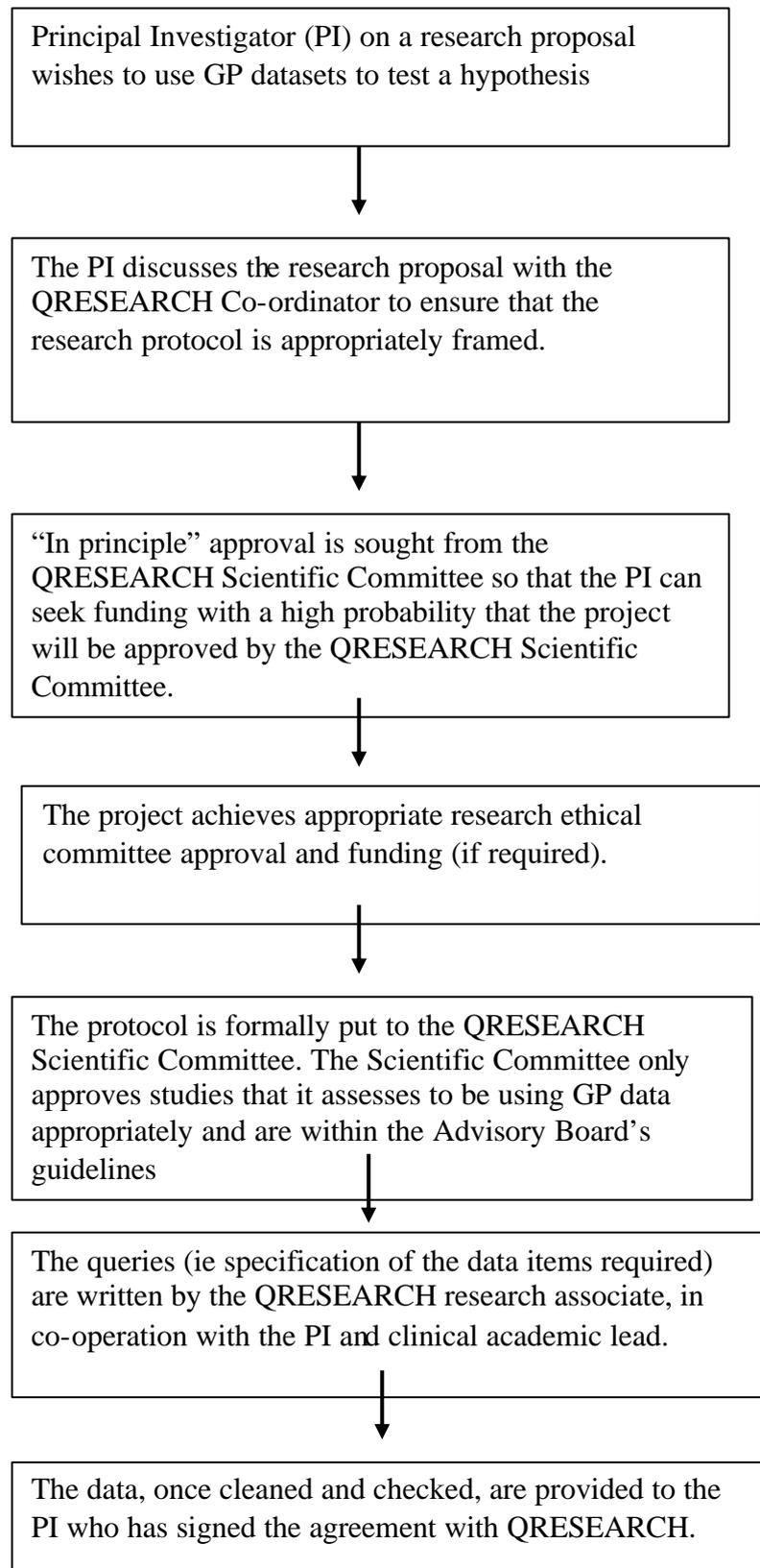
1. The data supplied will be used solely for the research purposes stated in the protocol which has been approved by an appropriate ethical committee.

2. I will hold all information in accordance with the principles and conditions set out in the Data Protection Act 1998 and with proper safeguards to ensure confidentiality.

3. I am free to publish any research findings, regardless of the nature of the findings, without hindrance or interference.

4. Any protocol changes will be submitted for ethical approval and when obtained will be submitted to QRESEARCH Scientific Committee for approval before analyses are undertaken

5. Data will not be released to any third party, including the funder, sponsor or other such body.

6. Anyone replacing me and becoming responsible for this data must complete a further QRESEARCH Confidentiality Declaration before receiving access to the data.

7. All research relating to the data supplied will have a research sponsor and meet the requirements of research governance

8. No attempt will be made to use the data to infer the identity of patient(s) or practice(s).

9. Any reports, papers etc to be published will acknowledge the source of the data and be submitted to QRESEARCH.

10. A final report of the project and copies of any publications will be submitted to QRESEARCH within one year of the completion of the research project.

11. Any contravention of the above conditions may result in media supplied being recalled by QRESEARCH and/or restrictions on future issue.

Signed..Date.....................................

Print Name…………………………………………

6   PROCESS FOR ACCESSING DATA:

We intend to process requests for access to patient or practice level data as follows:

Principal Investigator (PI) on a research proposal wishes to use GP datasets to test a hypothesis

The PI discusses the research proposal with the QRESEARCH Co-ordinator to ensure that the research protocol is appropriately framed.

"In principle" approval is sought from the QRESEARCH Scientific Committee so that the PI can seek funding with a high probability that the project will be approved by the QRESEARCH Scientific Committee.

The project achieves appropriate research ethical committee approval and funding (if required).

The protocol is formally put to the QRESEARCH Scientific Committee. The Scientific Committee only approves studies that it assesses to be using GP data appropriately and are within the Advisory Board's guidelines

The queries (ie specification of the data items required) are written by the QRESEARCH research associate, in co-operation with the PI and clinical academic lead.

The data, once cleaned and checked, are provided to the PI who has signed the agreement with QRESEARCH.

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
David Stables, EMIS computing

7   MANAGEMENT STRUCTURE

QRESEARCH is housed within the Division of Primary Care, University of Nottingham. The QRESEARCH team consists of a strategic director, academic director and research staff with the following roles and responsibilities:

A. Strategic Director - Professor Mike Pringle

Responsible for
- Overseeing whole project.especially use of data for morbidity statistics
- Liaison with local and national organisations and professional bodies
- Guarantor and custodian of QRESEARCH
- Acting as a Trusted Third Party in holding the QRESEARCH database

B. Academic director- Dr Julia Hippisley-Cox

Responsible for
- Overseeing whole project especially use of data for research
- Line management, supervision and training of the research associate(s).
- Recruiting practices to QRESEARCH
- Ensuring data security, specifying the systems and data requirements necessary for making data useful for research, quality assurance
- Specifying the data analysis necessary to report on data quality standards and working in conjunction with the research associate to undertake the analyses
- Guarantor and custodian of QRESEARCH
- Acting as a Trusted Third Party in holding the QRESEARCH database

C. Research Fellow

Responsible for:
- Day-to-day handling of QRESEARCH data, maintaining data security systems in conjunction with the Academic Director
- Liasing with researchers
- Maintaining log of details of research projects using QRESEARCH
- Producing reports
- Ddeveloping and implementing routines to
    (1) to check internal and external consistency of data
    (2) to identify samples for research
    (3) To prepare, clean, check and supply data for analysis.

8   PROFESSIONAL AND ACADEMIC REGULATION

There will be three board/committees

(1) QRESEARCH Management Board
(2) QRESEARCH Advisory Board – the Public and Professional Body
(3) QRESEARCH Scientific Committee

8.1 QRESEARCH Management Board

The management board will "own" QRESEARCH and run it as a non-profit making venture. This board represents the key interests in the setting up and running of QRESEARCH – EMIS and the team in the University of Nottingham. It decides the business (including financial) issues. It consists of two representatives from EMIS (including David Stables) and one from the EMIS National User group; and three from the University of Nottingham including Julia Hippisley-Cox, Mike Pringle and a representative of University of Nottingham.

8.2 QRESEARCH Advisory Board

8.2.1 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 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

8.2.2 Membership of Advisory Board

The board will need to 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 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
➢ Patient representation
➢ The Research Ethics Committee
➢ University of Nottingham
➢ Chair of the Scientific Committee

In order to carry out the development and piloting of QRESEARCH we have appointed a pilot Advisory Board. Its membership, in addition to Julia Hippisley-Cox, Mike Pringle and David Stables, is Dr John Williams, Dr John Ford, Dr Dai Evans and Professor Azeem Majeed.

8.3 QRESEARCH Scientific Committee

8.3.1 Terms of reference/remit:

➢ To follow the criteria and principles set out by the QRESEARCH Advisory Board in assessing and approving, if appropriate, applications to access the QRESEARCH database
➢ To assess the risks involved in each application for use of the QRESEARCH database and to refer difficult areas for decision to the QRESEARCH Advisory Group
➢ To advise the QRESEARCH team on technical issues

Although the Scientific Committee will need to meet from time to time, its decisions will usually be undertaken by e-mail, with dissent requiring a meeting or other form of collective decision making process.

8.3.2 Membership

The Scientific Committee will be chaired by Professor Azeem Majeed.
➢ One EMIS National Users Group representative
➢ One practices representative (different from above Advisory Board member)
➢ Two reputable academics from external institutions
➢ One representative nominated by the Joint Computing Group of the RCGP/GPC
➢ One Trent Focus representative
➢ One PRIMIS representative

Mike Pringle, Julia Hippisley-Cox and the Research Associate(s) will be ex-officio on the Scientific Committee and on the advisory board.

Julia Hippisley-Cox
Mike Pringle
March 2003

9   PRACTICE INFORMATION SHEET

Background

The University of Nottingham has been working in collaboration with EMIS to develop a new method, known as QRESEARCH, to make data from GP clinical computers available for research. This involves computer searches that are run centrally by EMIS. Data are anonymised within the practice and then extracted and aggregated to produce a database suitable for research.

Currently researchers wanting access to General Practice datasets for their research can undertake data collection using MIQUEST computer software. However, there are drawbacks because MIQUEST 'report style' searches needed for research may need to be repeated several times.

We have completed a pilot of QRESEARCH with 38 practices in the Trent Region and are now recruiting a national sample of 300-400 practices to be involved in a long-term national scheme. Full approval has been obtained from the Trent Multi-Centre Research Ethics Committee and the EMIS National User Group.

Why has my practice been invited to participate?

We are inviting all practices using EMIS Version 5 or above 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 the QRESEARCH module 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 (called the Carstairs rurality score) 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 enumeration district in which the patient lives. The scores are used in research to measure or monitor uptake and provision of health services.

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
         David Stables, EMIS computing

The practice system will then generate an upload of the anonymised data and transmit this data via the NHSNet to a dedicated server within EMIS. The data are updated on a daily basis and this update does not interfere with the running of the practice or any back-ups. Practices can view the data extracted each day if they choose in accordance with guidance from the Committee of Standards of Data Extraction (COSODE).

EMIS will undertake a secure transmission of the aggregated data to the University of Nottingham which is the single point of access to the data. To ensure maximum security, the data will be encrypted and will be stored on a stand-alone PC in the University of Nottingham. The University of Nottingham will be the only point of access to the data.

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 included in future research projects. If a practice wishes to withdraw, it just selects the 'deactivation' option from the QRESEARCH 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 will be able to run searches on their own system and compare their results with the anonymous aggregated searches obtained from the 300-400 QRESEARCH practices. Practices will be able to contribute to the decision about what feedback will be available by indicating preferences on our website (www.qresearch.org). We anticipate that our feedback will cover conditions NOT included in the New Contract as there are many other important diseases where comparative analysis may help benefit patient care. 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, our pilot work shows that the amount of time involved in activating the upload is minimal.

We will provide voluntary guidelines on data recording for practices which will be consistent with advice given by other organisations, including PRIMIS and the NHS Information Authority.

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 uses of the data, will 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.

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

21

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
          David Stables, EMIS computing

informing patients. The notice is attached to this information sheet. 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 which will cancel any future data collection from that patients' record. All data already collected from that patients record will then be deleted form the database.

What is the Advisory Board and what is its role?

The role of 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
➢ Patient representation
➢ The Research Ethics Committee
➢ University of Nottingham
➢ Chair of the Scientific Committee

In order to carry out the development and piloting of QRESEARCH we have appointed a pilot advisory board. Its membership, in addition to Julia Hippisley-Cox, Mike Pringle and David Stables, is Dr John Williams, Dr John Ford, Dr Dai Evans and Professor Azeem Majeed.

What are the requirements for the projects undertaken using QRESEARCH?

There will be 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 academic principal investigator and named co-investigators
(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 will made available to pharmaceutical companies for pharmacosurveillance.

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?

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 and this will be available to the practices on request.

Will all practices who volunteer be included?

If we have more than 400 practices included, then we will choose a geographically representative sample.

Who is organising the research and who are the guarantors?

The project is organised and guaranteed by Dr Julia Hippisley-Cox (Senior Lecturer) and Professor Mike Pringle, Division of Primary Care, 13th Floor, Tower Building, University Park, Nottingham, NG7 2RD.

Telephone: **0115 8466923 or 0115 8466901**
E-mail: **julia.hippisley-cox@nottingham.ac.uk** or **mike.pringle@nottingham.ac.uk**

QRESEARCH has been developed in conjunction with Dr David Stables, Medical Director of EMIS. Trent Multi-Centre Research Ethics Committee has approved the protocol. Anyone replacing these individuals will complete a QRESEARCH Confidentiality Declaration and notify Trent Multi-Centre Research Ethics Committee before receiving access to or involvement with QRESEARCH data.

Further queries

If you have any queries about QRESEARCH, then please contact either Dr Julia Hippisley-Cox or Professor Mike Pringle (as above) by email or phone. Further details about QRESEARCH including the full protocol and documentation 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 either of:

Professor Tony Avery
Head of Division of Primary Care, C floor, Medical School, Queen's Medical Centre, NG7 2UH.
Telephone: 0115 9709 387
E-mail: tony.avery@nottingham.ac.uk

Dr Dai Evans
EMIS National User Group, 2nd Floor, 2 Eslington Terrace, Jesmond, Newcastle upon Tyne NE2 4RJ.
Email: devans@emisnug.org

EXAMPLE NOTICE FOR PATIENTS

'SURGERY NAME' IS A RESEARCH PRACTICE.

We ask you for information about yourself so that you can receive proper care and treatment. We keep this information, together with the details of your care, because it may be needed if we see you again.

We may use some of this information for other reasons, for example, to help us protect the health of the public generally, to plan for the future, to train staff and to carry our medical and other health research for the benefit of everyone.

We are currently involved in some research studies that require anonymous information from patients' notes. You can not be identified from this information.

If anything to do with the research would involve you personally, you will be contacted to see if you are willing to take part, you will not be identified in any published results without your agreement.

You have a right of access to your health records. If at any time you would like to know more, or have any concerns about how we use your information you can speak to: 'Person nominated by practice' tel: number

EVERYONE WORKING FOR THE NHS HAS A LEGAL DUTY TO KEEP INFORMATION ABOUT YOU CONFIDENTIAL

QRESEARCH CONSENT FORM

I have read and understood the Practice Information Sheet on QRESEARCH and give consent for the practice to be a part of QRESEARCH. I understand that the upload by EMIS will only occur after the practice has activated the QRESEARCH facility on its system.

Signed... Date..........…………………

Name in capitals……………………………………..

On behalf of **«Organisation_Name»**
«EMIS_ref»
«Address_Line_1»
«Address_Line_2»
«Address_Line_3»
«Address_Line_4»
«Postcode»

We undertake to honour the principles of good medical research and the details of QRESEARCH as set out in the Practice Information Sheet. We will ensure that the data are used responsibly and protect patient and practice anonymity.

Signed... Date….19th May 2003

Dr Julia Hippisley-Cox, Division of Primary Care, University of Nottingham

Signed... Date...19th May 2003

Professor Mike Pringle, Division of Primary Care, University of Nottingham

PLEASE RETURN TO

Dr Julia Hippisley-Cox
Division of Primary Care
13th Floor
Tower Building
University Park
FREEPOST MID 17779
Nottingham NG7 2RD

11 INSTRUCTIONS ON ACTIVATING QRESEARCH UPLOADS

This is an Information Sheet Provided by EMIS

Activating QRESEARCH Uploads

General Information

- Activating the upload does not start transferring data to EMIS – it merely queues up data ready for EMIS to collect. The whole process of activation takes only a few moments and can be done at any time.
- EMIS will collect the data from this queue at a pre-agreed time. At present, data collection will take place only outside surgery hours so that system speed is not affected.
- Deactivation deletes the contents of the queue completely, rather than pauses transmissions. It should only be used if you wish to no longer participate in the QRESEARCH study.

If you have any questions or problems at any time during this process, please contact either John McLusky or Eamon Harvey at EMIS Project Support (0870 122 1133)
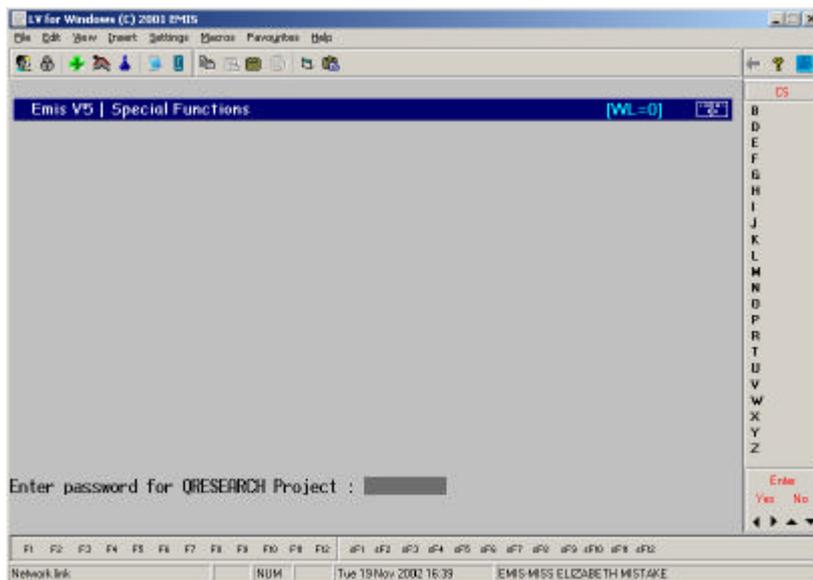
1. Check if your system is up to date.

   To do this, please choose option CS then SF from the main EMIS menu. You should see a menu similar to the following:
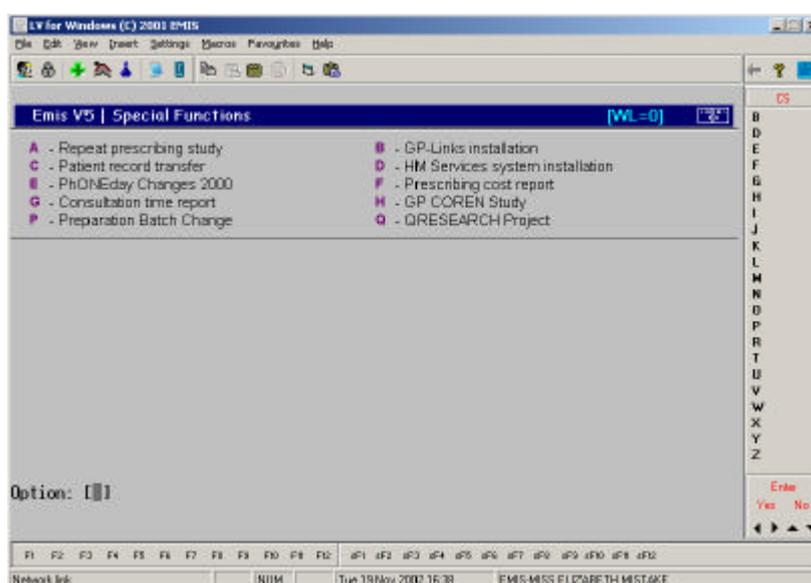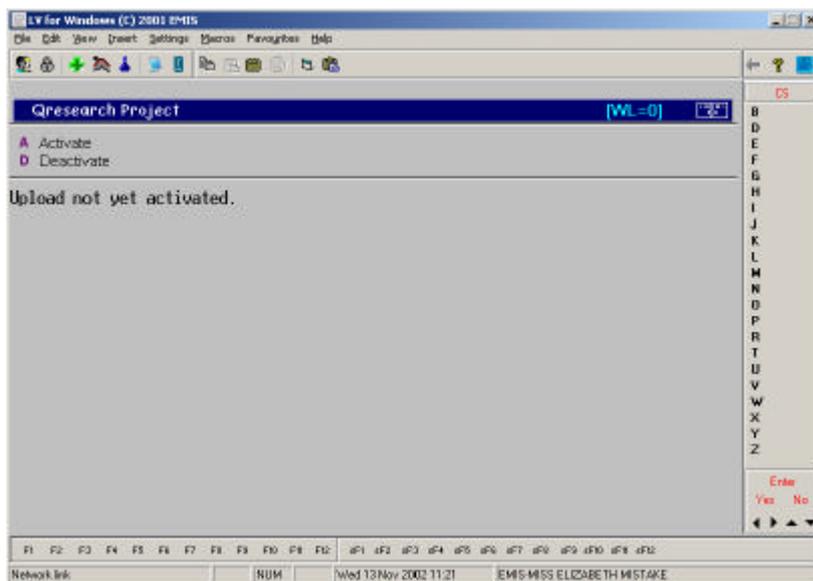

   If the 'Q – QRESEARCH Project' option is not available, you will need to contact Project Support at EMIS before you can continue.

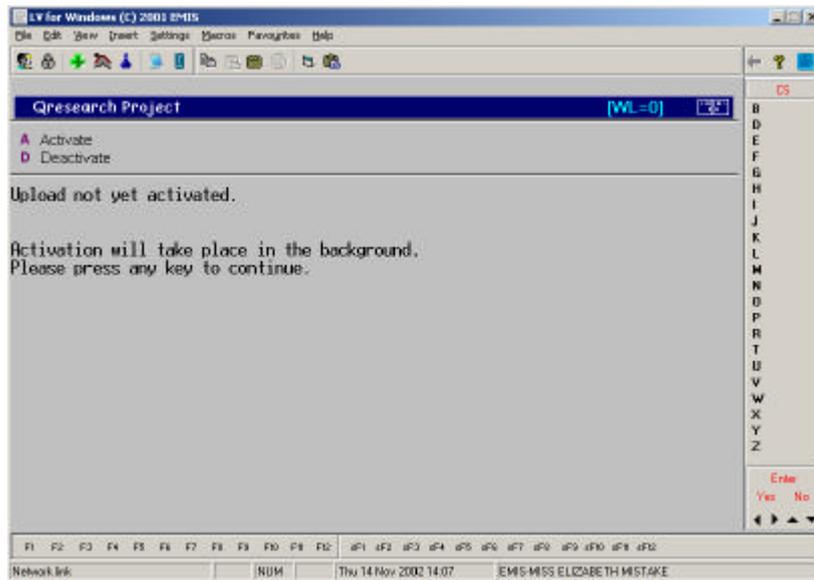2. Press 'Q'. You should now be prompted for a password:

At this stage, enter 'QRESEARCH' in upper case and press the 'enter' key.
3. You should now see a screen similar to that below.

To activate the upload, press the 'A' key.

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
           David Stables, EMIS computing

4. A message will now appear stating that activation will take place in the background.

5. After a few seconds, the 'Upload not yet activated' message will be replaced by a message similar to 'Activating: 34/10000', denoting that 34 patients from the practice population have been queued for transfer. At this point, you can press 'F1' to return to the Special Functions menu. You can return to the QRESEARCH menu at any time to check the status of activation.

Once activated, you can deactivate the upload at any time you choose. To deactivate the upload, please choose the 'D' option from the QResearch Project menu. **You should only do this if you no longer wish to participate in the study.**

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
          David Stables, EMIS computing

12 PRINCIPLES OF DATA EXTRACTION FOR GP SYSTEMS

Professor Mike Pringle wrote this document in July 2002, before The University of Nottingham began its negotiation with EMIS about QRESEARCH. The document was submitted to and approved by the Joint Computing Group. It has been included in the QRESEARCH protocol as there is no other similar statement of principles in the field.

12.1 Background

Researchers have long shown an interest in using routine clinical data from general practice. Such access has traditionally been episodic, discussed fully with practices and with ethical committee approval. The world is moving on, with two major innovations.

The first is electronic access to GP databases allowing "background" subliminal downloading of analysis or reports of patient data. This should never occur without practice consent, but such consent can be given and acted on over substantial periods of time. Today's informed consent can be tomorrow's breach of confidentiality.

The second is the requirements and expectations of the National Health Service for use of GP databases for health needs assessment (and therefore commissioning), clinical governance (including clinical audit), patient safety (such as significant event auditing) and contact monitoring. All these are obviously virtuous processes but must not be allowed to compromise the principles of confidentiality and legal data handling.

This paper examines the principles that should inform decision-making and system design in this new world.

12.2 Introduction

There are four broad categories of data extraction from GP systems other than for routine clinical care:

- Patient identifiable data
- Pseudo-anonymised data – report style query
- Anonymised data – report style query
- Anonymised data – analyse style query

First, two terms used need to be defined. A "**Trusted Third Party**" is an organisation or institution of reputation, that is independent of (at least one remove from) the Department of Health, the National Health Service or commercial ownership or control, and that uses its reputation as a guarantee of the security and processing of the data. The essence of such a body is that it earns and maintains the confidence and trust of the public, the health professions and stakeholder organisations through

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
        David Stables, EMIS computing

integrity, transparency and equity. There are many academic units, mainly Universities but sometimes in other settings, with the reputation and track record to be Trusted Third Parties.

A "**Professional and Public Body**" could be a single body, or one could be set up for specific projects extracting data from general practice computer systems. Such a body must:

- Represent firstly the interests of patients and secondly the interests of the health professionals and practices
- Include independent lay people
- Include independent representatives of the medical, nursing and other relevant health professions in primary care
- Have full access if requested to the dataset (if patient identifiable, with strong identifiers removed), the extraction and use audit trail and the resulting analyses if necessary to satisfy themselves that the data are being used ethically and properly
- Have full access to agreements concerning the use of the data

12.3 Patient identifiable data

There will always be legitimate reasons for extracting patient identifiable data from a GP system, other than for routine clinical care. Such extraction must:

- Be with the knowledge and informed consent of the guardian of the record (in this case the GP)
- Follow approval from a Research Ethics Committee

And it must satisfy one of these two conditions:

- Be with the informed, written consent of the patient
- Be approved by the Secretary of State

There must be both an audit trail for the data extraction and retention of the research database in order for both patients and health professionals to satisfy themselves, if necessary, that the data have been handled ethically and within the constraints of the prevailing law such as Section 60 of the Health Act 2002.

Provided both the patient (or the Secretary of State) and the practice have given informed consent, the ethics committee has approved and the data are handled according to the strictures of research governance, then the process should gain professional and public approval. However, researchers extracting these data would be well advised to:

- Inform a professional and public body and, if appropriate, see endorsement from that body (see below)
- Only handle the data through a Trusted Third Party

30
Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
        David Stables, EMIS computing

12.4 Pseudo-anonymised data

Pseudo-anonymised databases do not contain any strong patient identifiers and are anonymised in that no patient can be directly identified using the pseudo-anonymised database alone. They do, however, include a unique patient identifier that allows data from the same source or, potentially, other sources to be associated with that file (hence pseudo-anonymised). In theory there are no risks to patient confidentiality, but the robustness of such assurances needs to be assessed.

There are two main uses of pseudo-anonymised databases by researchers, and one by non-researchers:

a) to identify patients to participate in research projects.

General practice clinical databases can be used as a sampling frame to assist in the research process. Patients can either remain anonymous or choose to reveal their identity depending on the type of study and how they respond to the patient information sheet. For example, a researcher may approach a general practice or research network to recruit patients to a study. Once the researcher has ethical approval and a clear research question, the practice staff could run a computer search to generate a list of patients with a particular condition under study. The researcher can generate a sample of patients who meet the study criteria and who have been, for example, stratified to provide a sample with specific characteristics. The researcher can notify the practice of the unique patient identifiers that the practice translates back into patient identity. The GP can then write to the patients on behalf of the researcher, inviting them to participate in a research study.

In this way GP databases are used to identify patients for an invitation to take part in research without the GP revealing the identity of any patients to the researcher. The patient retains control and confidentiality is not breached.

b) use of pseudo-anonymised GP databases for longitudinal research

Researchers can also use pseudo-anonymised patient linked data to undertake research studies without ever needing to contact the patients for information. Since the unique patient identifier is the same in the original data extraction and subsequent ones, cohorts of patients can be followed up over short or long periods of time, if they remain with the same practice. Provided no research intervention is involved, the patients do not need to be contacted and their confidentiality is not compromised. An example of a research study using this paradigm might be a case control study to examine the risk of thrombo-embolism for patients on HRT or on the oral contraceptive pill.

c) Use for health service monitoring and quality assurance

It is possible to postulate circumstances in which a pseudo-anonymised database is required in order to audit clinical cases and promote patient safety. However these are unusual – if not rare – circumstances and most such purposes can be achieved with anonymous databases or analyses. Any non-research use of pseudo-anonymised

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
        David Stables, EMIS computing

databases should follow the same principles as research use and should only occur when there is a clear case for not using anonymised data.

What are the risks of pseudo-anonymised databases and how can they be minimised?

There are risks to health professional or practice confidentiality since such data must, on receipt at the extracting organisation, be identified by the source practice. In principle the main issues should focus on protecting the confidentiality of the health professionals and the practices. The risk comes from mis-interpretation and mis-use of the data, inappropriate or non-consented association of data from multiple databases, and breaches in health professional and/or practice confidentiality.

These databases will be of significant value for researchers, health service managers, the public and the profession. However they must be handled with a high degree of caution. In order to protect confidentiality with such extractions, retention and association with other data the following conditions should apply:

- The guardian of the record (in this case the GP) should have knowledge of the process of data extraction and give informed consent. Such "knowledge and consent" must include the data to be extracted, the data to which it is to be associated and the uses to which the data will be put
- The methodology for extracting and retaining data should be explicit and easily understood by lay and professional people, and should have the written consent of a Research Ethics Committee. This is regardless of whether the prime purpose of the data is research. If the data is to be used for research, then there must be a clear research question, a protocol and specific ethical approval
- There should be an audit trail for the data extraction and retention of the database in order for health professionals to satisfy themselves, if necessary, that the data have been handled ethically
- The data should be processed and retained only by a Trusted Third Party who must never disclose health professional or practice identity to any other person or organisation without the explicit consent of the relevant health professional or the practice
- A Professional and Public Body must have oversight and approval

The Trusted Third Party's overall task would be to:

- Control access to the database in such a way to best protect the confidentiality of patients, GPs and practices while maximising the benefits from the data for research, management and patient safety.

12.5 Anonymised data – report style query

Anonymised data are extracted from GP systems for a variety of purposes including research, clinical audit, clinical governance, health services management and patient

safety. One way in which the data can be extracted is as a "report" in which one record (data line) represents one patient or one event, but which contains no patient identifier. Such a file carries very small but significant risk for inferred identity – only a reality if the practice and/or health professional is identifiable. Therefore even such anonymous data need to be handled responsibly to ensure anonymity and confidentiality at patient, health professional and practice levels.

Therefore, the extraction of such anonymised report style data must only occur:

- With the knowledge and informed consent of the guardian of the record (in this case the GP). Such "knowledge and consent" must include the data to be extracted, and the uses to which the data will be put, who will have access and the duration before file storage and end of use
- If any research use of the dataset has Research Ethics Committee approval
- If there is an audit trail for the data extraction and retention of the database in order for health professionals to satisfy themselves, if necessary, that the data have been handled ethically.

A GP and practice should not consent to such an extraction without understanding and agreeing the protocol for anonymity or identification of the GP and/or the practice. If anonymity is guaranteed, then such a guarantee should be offered by either a Trusted Third Party or an organisation, such as a Primary Care Trust, in a position to ensure that that guarantee is effective.

Those extracting anonymised data for whatever purpose would be well advised to:

- Use a Trusted Third Party to process and retain the data
- Involve a Professional and Public Body in oversight and approval
- Control access to the database in such a way as to best protect the confidentiality of patients, GPs and practices while maximising the benefits from the data for research, management and patient safety
- Ensure that GPs and practices retain the right to have all their data removed from the database without having to give a reason

12.6 Anonymised data – analysis style query

Analysis style queries yield tables of data. Under normal circumstances these are the data required by the National Health Service for monitoring and clinical governance. They may be used for research. There are very few ethical issues for patients, but still significant ones for health professionals and practices.

The same strictures as for report style queries apply to analysis style queries, but the need for professional and public of oversight for non-research extraction may be less. However many practices will wish to see the use of Trusted Third Parties to ensure that their interests are not compromised.

12.7 Conclusions

If these principles were agreed, then there is a requirement for:

- Trusted Third Parties
- Professional and public body or bodies

Professor Mike Pringle
July 2002

34
Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
          David Stables, EMIS computing

REFERENCES

[1] Depression as a risk factor for ischaemic heart disease in men: population cased case-control study. Hippisley-Cox J, Fielding K, Pringle M. British Medical Journal, 1998; 316: 1714-1749

[2] Are spouses of patients with hypertension at increased risk of having hypertension? A population-based case-control study. Hippisley-Cox J, Pringle M. British Journal of General Practice, 1998; 4: 1580-1584

[3] Inequalities in access to coronary angiography and revascularisation: the association of deprivation and location of primary care services. Hippisley-Cox J, Pringle M. British Journal of General Practice, 2000; 50:449-454

[4] Time changes in new cases of ischaemic heart disease. Meal A, Pringle M, Hammersley V. Family Practice, 2000; 17: 394-400

[5] Sex inequalities in ischaemic heart disease in general practice: cross sectional survey. Hippisley-Cox J, Pringle M, Crown N, Meal A, Wynn A. British Medical Journal, 2001; 322: 832-834

[6] General practice workload implications for the national service framework for coronary heart disease: cross sectional survey. Hippisley-Cox J, Pringle M. British Medical Journal, 2001; 323: 269-270

[7] Antidepressants as risk factor for ischaemic heart disease: case controlled study in primary care. Hippisley-Cox J, Pringle M, Hammersley V, Crown N, Wynn A, Coupland C. British Medical Journal, 2001; 323: 666-669

[8] Married couples' risk of the same disease: cross sectional study. Hippisley-Cox J, Coupland C, Pringle M, Crown N, Hammersley V. British Medical Journal, 2002; 325: 636-638

[9] A cross sectional survey of the effectiveness of lipid lowering drugs in lowering serum cholesterol in 17 general practices: how well do they work? Hippisley-Cox J, Cater R, Pringle M, Coupland C. British Medical Journal 2003 (in press)

[10] Primary and secondary prevention and age inequalities – the effect of the first year of the National Service Framework for Coronary Heart Disease. Hippisley-Cox J, Pringle M, Cater R, Coupland C, Meal A. Submitted British Medical Journal 2003

[11] A comparison of research general practices and their patients with other practices – a cross-sectional study in Trent. Hammersley V, Hippisley-Cox J, Wilson A, Pringle M. British Journal of General Practice, 2002; 52: 463-468

[12] Using MIQUEST in general practice. Hammersley V, Meal A, Wright L, Pringle M. Journal of Informatics in Primary Care, 1998; Nov: 3-7

Authors: Julia Hippisley-Cox, Mike Pringle, University of Nottingham
        David Stables, EMIS computing